'Experimental data seem deceptively easy. This book not only shows why and when estimation is involved. It also pushes the frontier, for example by introducing methods to capture patterned heterogeneity, random utility, or two-part decision rules.' – **Christoph Engel**, *Max Planck Institute, Germany*

'A long-awaited, systematic treatise of the econometric modelling of experimental data brilliantly accomplished. A work of art!' – **Anna Conte**, *University of Westminster, UK*

'*Experimetrics* provides an excellent overview of the issues concerning the econometric analysis of experimental data. Numerous STATA codes enrich the book and make the methods very accessible.' – **Charles Bellemare**, *Université Laval, Canada*

'I believe *Experimetrics* should become an indispensable part of every experimentalist's toolkit.' – **David Butler**, *Murdoch University, Australia*

*Experimetrics* is an essential guide to discovering new and more illuminating ways to analyse experimental econometric data.

Peter Moffatt, one of the world's experts in the field, covers a range of techniques: from the familiar, such as treatment testing, to lesser known ones such as finite mixture models and the method of maximum simulated likelihood. The book takes a hands-on approach by explaining STATA commands in detail. In addition, difficult problems inherent in the methodology are addressed, such as the parametric estimation of social preference models, quantal response models, and learning models.

An indispensable book for researchers and advanced students in experimental and behavioural economics who want to come to grips with the field of Experimetrics.

**Peter G. Moffatt** is Professor of Econometrics at the University of East Anglia, UK. He has published many articles in the areas of applied econometrics and experimental economics and is renowned for presenting masterclasses in Experimetrics around the world.

THE COMPANION WEBSITE *WWW.PALGRAVE.COM/MOFFATT* CONTAINS:
- ALL DATA SETS (IN STATA FORMAT) USED AS EXAMPLES IN THE BOOK;
- AN EXECUTABLE STATA "DO-FILE" CONTAINING STATA COMMANDS AND PROGRAMS USED IN EXAMPLES; AND
- AN EXCEL FILE CONTAINING SOME EXCEL CALCULATIONS PRESENTED IN THE TEXT

macmillan education palgrave

ISBN 978-0-230-25023-9
90101
9 780230 250239
www.palgrave.com

macmillan education

EXPERIMETRICS

PETER G. MOFFATT

palgrave

PETER G. MOFFATT

# EXPERIMETRICS

## ECONOMETRICS FOR EXPERIMENTAL ECONOMICS

Praise for *Experimetrics*

'Experimental data seem deceptively easy. This book not only shows why and when estimation is involved. It also pushes the frontier, e.g. by introducing methods to capture patterned heterogeneity, random utility, or two-part decision rules.' **Christoph Engel, Max Planck Institute, Germany**

'A long-awaited, systematic treatise of the econometric modelling of experimental data brilliantly accomplished. A work of art!' **Anna Conte, University of Westminster, UK**

'*Experimetrics* provides an excellent overview of the issues concerning the econometric analysis of experimental data. Numerous STATA codes enrich the book and make the methods very accessible.' **Charles Bellemare, Université Laval, Canada**

'I believe *Experimetrics* should become an indispensable part of every experimentalist's toolkit.' **David Butler, Murdoch University, Australia**

'This text provides guidance and reference for the increasingly large number of experimental economists who want to do some serious econometric analysis of their experimental data. Not only does it explain the concepts cleanly and precisely, it also provides numerous examples of applications. It will become compulsory reading for all experimental economists.' **John Hey, University of York, UK**

'*Experimetrics* is a very well-written book, providing an outstanding in-depth analysis of the econometric methodology into economics experiments. It is structured in a balanced way covering the most topical issues in the experimental literature. Superb writing style and coverage!' **Michalis Drouvelis, University of Birmingham, UK**

# Experimetrics

# Experimetrics

## Econometrics for Experimental Economics

Peter G. Moffatt

macmillan education palgrave

First published 2016 by
PALGRAVE

Palgrave in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of 4 Crinan Street, London N1 9XW.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave is a global imprint of the above companies and is represented throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978–0–230–25022–2 hardback
ISBN 978–0–230–25023–9 paperback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Printed and bound by CPI Group (UK) Ltd, Croydon, CR0 4YY

# Contents

# Acknowledgements

This book is, in large part, the end-product of a sequence of masterclasses in Experimetrics that I have delivered in various locations since 2006. I thank various people for arranging these courses and for providing hospitality during the visits, including Catherine Eckel, Tobias Uske, Anna Conte, Sylvia Arnold, Vittoria Levati, Angela Fletcher, Lionel Page, Uwe Dulleck, David Butler, Steven Schilizzi, Arne Weiss, and Gina Neff. I also thank the participants on these masterclasses for providing the opportunity to test the material on live audiences. I am grateful to the editors at Palgrave, Jaime Marshall, Aléta Bezuidenhout, Kirsty Reade, Jo Endell-Cooper, and Georgia Walters for their persistent encouragement towards completion of the book, to Alex Antidius Arpoudam of Integra Software Services for managing the book's production so professionally, and to the proofreader, Sarah Perkins, for her meticulous reading of every page of the manuscript and for finding typos and infelicities that I would never have found myself. I am grateful to Phil Bacon and Linford Bacon for providing expert assistance in converting the manuscript to LaTeX. I am also grateful to a number of (partially) anonymous reviewers who have provided invaluable, highly expert feedback and recommendations at various stages of the project – the feedback of Ryan Oprea and John Hey has been particularly influential. Finally, I am grateful to Robert Sugden for suggesting to me – all those years ago – that it would be a good thing if more econometricians were interested in the analysis of experimental data.

# Chapter 1
# Introduction and Overview

## 1.1 What is Experimetrics?

*Experimetrics*[1] comprises the body of econometric techniques that are customized to experimental applications. A wide variety of such techniques appear in the experimental economics literature. The aim of this textbook is to assemble this body of techniques, to demonstrate their use in a hands-on style, drawing on as wide a range of examples as possible, and to interpret each set of results in ways that are most useful to experimental economists. The target audience is mainly researchers in experimental economics. It is also conceivable that the book might be of interest to econometricians who are curious to know what sort of techniques are being used by experimental economists.

The experimetric techniques that already appear in the experimental economics literature range from the very basic to the highly sophisticated. At the basic end of this spectrum we see the class of techniques known as *treatment tests*, that is, tests which compare outcomes with and without a treatment, or before and after a treatment. At the sophisticated end of the spectrum, we see highly complex structural models, with a deterministic core corresponding to an underlying behavioural theory, with possibly many structural parameters, and a stochastic specification including possibly many dimensions of variation, both within and between subject. Needless to say, the type of econometric approach that is chosen is often, and justifiably, dictated by the type of experiment that has been conducted, and by the types of research questions being addressed.

In some experiments, the "home-grown" characteristics of the experimental subjects are the focus, and the objectives are usually simply to investigate how individuals make decisions, or interact with each other, in particular settings. These

[1] The word "Experimetrics" was (to the best of my knowledge) coined by Camerer (2003, p. 42). Houser's (2008) entry in the *New Palgrave Dictionary of Economics* on "Experiments and Econometrics" commences with the line "'Experimetrics' refers to formal procedures used in designed investigations of economic hypotheses." Bardsley & Moffatt (2007) are apparently the first authors to have used the word in the title of a published paper.

studies typically rely on relatively simple experimental designs (e.g. choosing between lotteries; splitting a pie), with the ultimate objective of measuring subjects' characteristics, especially preference parameters. It is normal to expect substantial variation in these measured characteristics. Indeed, it is often the precise features of this variation in which we are most interested; for example, the proportion of the population who are "selfish", or the proportion who are EU maximisers. When data are from this type of experiment, it is often seen as appropriate to model the decision-making process using structural estimation methods, for example methods that simultaneously estimate all of the parameters appearing in the individuals' objective function, as well as distributional parameters, some of which capture preference heterogeneity.

In other experiments, the focus is on the functioning of an economic institution, rather than the characteristics of the individual participants within it, and the objective may be to test a particular theory as applying to that institution. In these settings, *induced value methodology* is commonly used. This technique is based on the idea that the appropriate use of a reward medium allows the experimenter to *induce* pre-specified characteristics (e.g. preferences) in the subjects so that their innate characteristics become irrelevant. Having essentially eliminated the impact of subjects' characteristics in this way, it is clearly much easier to apply close scrutiny to the theory under test. In these settings, the experimental designs are relatively complex, since key features of the economic institutions need to be captured in convincing ways. However, the econometric techniques required are often very simple. The level of control is normally such that straightforward treatment testing is often seen to be the natural framework, and the most suitable means of obtaining answers to the research questions of interest.

The following subsections provide brief overviews of the types of econometric techniques that are best suited to each of these two broad areas. As such, this chapter provides a broad overview of the contents of the remainder of the book.

## 1.2 Experimental Design

The topic of experimental design is clearly much more important in experimental economics than in other areas of economics. This is because, in other areas, typically the data generation process is out of the control of the investigator. In experimental economics, the data generation process is very much within the control of the investigator. Hence, design issues such as the choice of sample size, the sampling process, and the process for assignment of subjects to treatments, all take centre stage.

A central concept is randomization. If randomization is correctly applied in the process of selecting subjects for an experiment, then identification of treatment effects, which has always been a central problem in mainstream econometrics, is not a problem. The other side of this coin is that, since the data have not been collected in a natural environment, experimental results do not necessarily carry over to the world outside the lab. We therefore see that the advantages of experimental research



in terms of identification may be seen to be countered by the disadvantage of non-generalisability (see Al-Ubaydli & List, 2013).

There are several different types of design, including completely randomised designs, within-subject designs, crossover designs, and factorial designs. When groups of subjects are playing against each other, a choice also needs to be made between "partner" and "stranger" matching. Each design has both advantages and disadvantages, and the decision of which to use is often a delicate issue.

The choice of sample size is another key design decision. The question to be addressed here is how many subjects are required for the experimenter to be confident of reliable conclusions. More precisely, how many are required in each treatment? A useful framework for addressing these questions is power analysis (Cohen, 2013); that is, using probability theory to find what sample is required to provide a given "power" of the test being conducted.

A particularly interesting problem in experimental design is how to specify binary choice problems (e.g. lottery choice) in such a way as to generate a data set from which subjects' preference parameters may be estimated with maximal precision. A chapter of the text is dedicated to this problem.

## 1.3 The Experimetrics of Theory Testing

It is often claimed the purpose of an experiment is to test an economic theory. From the econometrician's perspective, a natural way to perceive such a test is as an assessment of whether the *predictions* from the theory provide good approximations to actual behaviour (i.e. the behaviour of subjects in the lab). From this perspective, the role of the experimenter is to find regularities in observed behaviour, and then to ask which theories are best able to account for these regularities.

Competitive equilibrium is the central concept in many theories. The objective of an experiment (a market experiment, say) may be simply to observe how close behaviour is to the competitive equilibrium, and we shall refer to this as testing the *fundamental prediction* of the theory. This type of experiment is usually performed using *induced value* methodology; that is, a system in which each subject is exogenously assigned a valuation of the object being traded, so that the complete demand and supply schedules, and therefore the equilibrium, are known by the experimenter. In this setting it is clearly a simple matter to use experimental data to assess how close actual behaviour is to the fundamental (equilibrium) prediction.

An important aspect of the fundamental prediction of the theory is that it often amounts to a "point prediction", that is, it simply tells us that a particular decision variable will, given the known fixed values of the exogenous variables, take a particular value. There are no *free parameters* in the model that generates the fundamental prediction. To see how free parameters enter a model, consider the following example. Start with a benchmark model that is built on the assumption of risk neutrality (RN). Such a model is likely to lead to a "risk-neutral equilibrium" prediction, $y = constant$, where $y$ is the decision variable. This benchmark model has no free parameters. Now consider what happens when we adopt the assumption of expected utility (EU) maximisation in place of the RN assumption. This inevitably results in the appearance of (at least) one free parameter, which will typically be one of the standard measures of risk aversion. The model's prediction will now depend on the value taken by this free parameter. Next, consider what happens if we futher generalise the model to assnme that subjects behave in accordance with prospect theory (Kahneman & Tversky, 1979) instead of EU. This extension results in the further addition of free parameters capturing probability weighting and loss aversion.

Fairly early in the book, we shall be drawing heavily on the examples of experimental auctions and contests. Auction theory and contest theory are both well developed and lead to very clear predictions. The fundamental prediction in these contexts usually takes the form of risk-neutral Nash equilibrium (RNNE) bidding behaviour, which depends on the precise structure of the type of auction or contest under study. With experimental data on subjects' bids, it is usually a simple matter to test whether behaviour is consistent with the RNNE.

In many situations, behaviour of experimental subjects tends to depart in systematic ways from the "fundamental prediction" of theory. In the context of auctions and contests, these departures take the form of systematic "over-bidding" relative to the Nash equilibrium prediction. Hence, if our only objective were to test the fundamental predictions of Nash equilibrium theories, this objective would be straightforwardly met: we would reject the theory. However, there are other levels at which theory may be tested. A typical theory gives rise to a number of "comparative static predictions". These are predictions of the decision variable moving in a particular direction in response to an exogenous change in another variable. For example, in many auction contexts, it is predicted that a rise in the number of bidders in the auction has a negative effect on bids. If two sets of experimental auctions are run, one with four bidders, and the other with six, we might test to see whether bids are lower in the auctions with six bidders. If they are, it would be reasonable to conclude that the experimental data is consistent with this particular comparative static prediction of the theory, even if the theory's fundamental prediction fails.

The test just described, for testing the effect of the number of bidders on bids, is an example of a treatment test. Auctions with the lower number of bidders may be referred to as the "control", and those with the higher number of bidders as the "treatment". There are a large number of possible ways of conducting a treatment test, both in terms of experimental design and in terms of the statistical procedure used to compute the test statistic. The two groups of subjects could be separate, in which case it is a between-sample test. Alternatively, subjects could be subjected to both treatments, in which case it is a within-sample test. The approprate choice of statistical test depends on which of these sampling approaches has been followed, as well as on a number of other design features such as the sample size, and the process by which subjects have been divided into sessions and groups.

There is another important use of treatment tests. In situations in which behaviour is found to depart from the fundamental prediction of the theory, it is of obvious interest to know the reason for this departure; these reasons are sometimes referred to as the "behavioral drivers of out-of-equilibrium play". It might be suggested, for example, that the reason for over-bidding in a contest is a "joy of winning". To test this, a treatment would be designed in which the joy of winning is absent. If bids are lower under this treatment, we may conclude that joy of winning is indeed a cause of over-bidding. Other suggested reasons for over-bidding include other-regarding preferences, risk aversion, and probability distortion.



When a treatment test is performed, the approach can be non-parametric or parametric. Non-parametric tests are sometimes preferred because their validity relies on fewer assumptions. Parametric tests often rely, to some extent, on assumptions such as normality of the underlying data.

Parametric treatment tests may be performed in the context of a linear regression, in which one of the explanatory variables (or the only explanatory variable) is a "treatment dummy". The advantages of this approach are: the coefficient of the treatment dummy is directly interpretable as the treatment effect; the associated t-statistic is the test statistic for the treatment; more than one treatment may be tested at the same time; the effects of other determinants of the outcome may be controlled for; and regression-related routines, such as clustering of standard errors, may be exploited.

## 1.4 Dependence in Experimental Data

Dependence is an issue that is central to Experimetrics. Basic treatment tests (and many other testing and estimation procedures) rely crucially on the assumption of independent observations. There are several reasons why this assumption can be expected to fail when analysing experimental data (for a recent discussion, see Fréchette, 2012). Firstly, if, as is common, experimental subjects are engaging in a sequence of tasks, there is likely to be "clustering" at the subject level, since some subjects will simply be predisposed to higher values of the decision variable than other subjects. This will of course mean that there are positive correlations between the observations for a given subject. There is also likely to be clustering at the level of the "group" in which a subject operates; a subject's behaviour is likely to depend on the behaviour of other members in the same group. It is often suggested that there is also clustering at the session level; for example, it may be expected that behaviour in the afternoon session differs from that in the morning session, purely as a consequence of the time of day. A more subtle reason for clustering at the session level is when "stranger" treatments are used (Andreoni, 1988), where group composition changes between rounds. Subject-level clustering is the "lowest level" of clustering, while session-level clustering is the "highest level".

There are a number of strategies that may be followed to allow for clustering, in order to validate the tests being performed. One is an ultra-conservative approach: to take the average behaviour over independent units (subjects, groups, or sessions, depending on the level of clustering assumed), and then apply the testing procedure to these averages. Provided the level of clustering at which averaging is performed is sufficiently high, the averages will automatically meet the requirements of independence. The most conservative possible approach would be to take averages at the highest level of clustering. The obvious disadvantage of this approach is that the process of averaging severely reduces the size of the sample to which the test may be applied, and therefore reduces the test's power (i.e. the probability of finding a significant treatment effect when a treatment effect indeed exists).

The second possible approach is to run regressions with treatment dummies, and to use cluster-standard errors (that is, standard errors corrected for clustering at the assumed level) in the computation of the treatment test statistic. The third possible approach is to use panel data estimation techniques, such as random effects and fixed effects models. These techniques deal directly with the panel structure of the data. It is possible to go even further, by using the multi-level modelling approach, which extends the random effects framework to situations in which there is more than one level of dependence (subject-level and session-level, for example).

## 1.5 Parametric versus Non-parametric Approaches

A fundamental choice is that between parametric and non-parametric methods. One of the key issues in this choice is the scale of measurement of the outcome variable (nominal, ordinal, or cardinal). Many parametric tests rely on distributional assumptions which can only hold if the variables under analysis are measured on a cardinal scale.

Most commonly, the distributional assumption that is required is that of normality in the outcome variable. If the outcome variable is normally distributed, this is highly convenient because it means that, subject to certain other requirements, parametric tests, for example the t-test, can be relied upon. However, data from economic experiments are often clearly non-normal, and this appears to be of particular concern to many experimental economists. Many researchers unquestioningly apply non-parametric tests to their data and explain this choice by the concern that their outcome variable is non-normal. However, this strategy is likely to be costly. When non-parametric tests are applied to cardinal data, the cardinal information in the data is disregarded, since the tests are based solely on the ordinality of the data. This is one reason why non-parametric tests tend to be less powerful (i.e. having a lower probability of detecting an effect when one exists) than their parametric counterparts.

There is a definite sense in which the normality requirement is taken too seriously by experimental economists. Even if the data are non-normal, provided that the sample size is sufficiently large (usually taken to mean more than 30 observations in each treatment), the central limit theorem may be invoked (implying that the *standardized mean* of the data is normally distributed in repeated samples) and parametric tests may be relied upon. Moreover, even if the sample size is insufficient for the central limit theorem to apply, methods are available which enable the valid use of parametric tests. One such method is the bootstrap (Efron & Tibshirani, 1993). This method provides a means of validly conducting parametric tests that respect cardinality, without making any assumptions about the distribution of the data. Hence the bootstrap may be viewed as a means of combining the benefits of parametric and non-parametric tests, whilst avoiding the drawbacks.

Another type of non-parametric method is the non-parametric regression. This is a procedure which, roughly speaking, results in a smooth, flexible curve being fitted through a scatter plot. It is very useful at the initial, exploratory, stages of data analysis, and is used to determine the nature of the relationship between two variables. In particular, it can be used to determine whether the relationship is linear, and if not, whether it is U-shaped, inverted-U-shaped, cubic, and so on. Finding out the nature of the relationship in this way is clearly very useful in deciding on an appropriate specification for the parametric model. The non-parametric regression technique adopted in this book is the locally-weighted regression (Lowess) technique of Cleveland (1979), which is available in STATA.



## 1.6 Structural Experimetrics

One of the principal objectives of this textbook is to encourage the wider use of fully structural models and to present complete explanations of how they can be estimated.

A structural econometric model is a model that combines an explicit economic theory with an appropriate statistical model. See Reiss & Wolak (2007) for a thorough survey of the applications of structural models in the area of industrial organisation. In the context of Experimetrics, structural modelling becomes potentially very useful when the objective of the experiment is the measurement of "home-grown" features of experimental subjects. By this we mean features such as risk-attitude or degree of altruism. Many wonld agree that these features are best modelled from the starting point of a utility function: a von Neumann-Morgenstern utility function if risk attitude is the focus; a "two-good" utility function over own payoff and other's payoff if altruism is the focus.

One obvious advantage of the structural approach is it provides a solution to the dependence problem, similar to that provided by multi-level modelling. Under the structural approach, the hypotheses of interest may be validly tested using data at the level of individual decisions. Subject characteristics may be used as explanatory variables to explain "observed" subject-heterogeneity, while a random effect term may be used to account for unobserved heterogeneity. Group or session level random effects may also be included. The structural approach actually allows alternative strategies for controlling for behaviour within the group, for example the inclusion of mean of group contribution in previous round as an explanatory variable.

The scale of measurement is important in the choice between modelling strategies. Much of mainstream econometric modelling is built on the assumption that the outcome is continuous. In experimental economics, this is frequenty not the case. Often the outcome is binary. Sometimes, it is discrete but with more than two outcomes, in which case, we need to consider whether the outcome is nominal (i.e. categorical) or ordinal. Sometimes the outcome from the theoretical model is a continuously distributed variable, but the nature of the data is such that the observed variable is far from continuous: it may be lower or upper censored, or there may be accumulations of data at particular interior "focal points". Even in a situation in which the outcome is truly continuous, we often need to pay careful attention to its distribution. In a fully parametric structural model, it is important to specify all distributional assumptions correctly in order to achieve consistency of parameter estimates.

In some situations, the structural parameters of interest may be obtained using estimation routines that are readily available in software packages. Such routines include linear regression, panel models, binary data models, censored data models, interval regression, and ordinal models. In other situations, the required routines are not readily available, and purpose-built programs need to be written. Fortunately, software facilitating the development of such programs is readily available. One such tool that is used particularly heavily in this book is the `ml` routine in STATA (Gould et al., 2010).

A considerable advantage of structural models is that they incorporate randomness in behaviour in ways that are natural. Stochastic terms, or "error terms", are something that some experimental economists begrudgingly perceive as a component that needs to be appended to their cherished economic theory, in order for it to be able to explain actual behaviour. Moreover, there is a tendency for them to perceive deviations from expected behaviour as (genuine) "errors" (i.e. mistakes). By encouraging experimental economists to embrace the framework of structural models, we will enable them to move forwards in accepting that randomness is a perennial (and natural) feature of human behaviour. The key message here is that the stochastic term is not an afterthought that needs to be appended; it is an integral feature of the model. To quote Harrison et al. (2015), "in short, one cannot divorce the job of the theorist from the job of the econometrician".

Once the importance of the role of the stochastic specification has been established, the next issue to be addressed is where and how the stochastic elements should be introduced. As we shall see, there are a number of possible approaches. In the context of individual decision making, the most obvious approach is simply to apply an additive error to the equation that represents the theoretical prediction; such an approach is analogous to straightforward regression analysis, although it is sometimes much more complicated than this. An alternative approach is to assume that variation in behaviour is explained by variation in the model's parameters, either between or within subjects, or both. We will refer to this as the "Random Preference" approach. A third possible approach is to introduce what has come to be known as a "tremble term". This is a way of capturing misunderstandings and lapses of concentration. Sometimes, these three stochastic approaches are used in combination.

Perhaps the best illustration of all of these issues is in risky choice modelling (see Loomes et al., 2002; Harrison & Rutström, 2009; Conte et al., 2011; Von Gaudecker et al., 2011). The data set would typically consist of a sequence of binary choices between lotteries, performed by each of the subjects in a sample. Central to the modelling strategy is a von Neumann-Morgenstern utility function which captures the risk attitude (or "preferences") of an individual. Risk attitude clearly varies between individuals so the risk-attitude parameter takes the role of a subject-specific random effect. It is also clear that there is significant within-subject variation, and this is captured either by assuming that preferences vary over time for a given subject (the *random preference approach*), or that the subject makes a computational error each time he or she makes a choice (the *Fechner approach*). The tremble assumption is a useful complement to each of these stochastic approaches since it allows for the occasional occurrence of extremely unlikely choices. An obvious theoretical framework on which to build these stochastic models is EU. Typically, however, EU is found to be over-restrictive, and models such as rank dependent (RD) theory (Quiggin, 1982) and cumulative prospect theory (Tversky & Kahneman, 1992) are found to fit the data better. These models include, in addition to risk aversion parameters, probability weighting parameters and (if the outcomes in the experiment include losses as well as gains) loss aversion parameters. The finite mixture approach is sometimes used to separate subjects into EU and RD "types" (Harrison & Rutström, 2009; Conte et al., 2011). All of these models can also be extended to allow certain parameters to depend on experience. This is useful for capturing the phenomena of, say, computational errors decreasing in magnitude with experience, or subjects moving closer to EU-maximization with experience (Loomes et al., 2002). All of these models may be estimated in a maximum likelihood framework, using maximisation routines that are computationally feasible.



## 1.7 Modelling Subject Heterogeneity

Perhaps the most important of all reasons for developing structural models in the context of experimental data is that they enable us to incorporate between-subject heterogeneity. There are two broad types of heterogeneity. *Discrete heterogeneity* is the situation in which the population of subjects is made up of a finite number of different "types" who respond in categorically different ways to stimuli. One example was mentioned at the end of the last sub-section: some individuals are EU-maximisers; others behave according to rank dependent theory (i.e. they weight probabilities). For another standard example, in the context of a public goods experiment, we might start by assuming that the population of subjects divides neatly into four types: "strategist", "altruist", "reciprocator", and "free-rider" (Bardsley & Moffatt, 2007).

*Continuous heterogeneity* is the situation in which subjects differ from each other in a dimension which is continuously measurable. A natural example is risk attitude. Every individual might be assumed to possess his or her own risk-aversion parameter and it is natural to assume that this varies continuously across the population.

The two types of heterogeneity call for different types of structural econometric models. Discrete heterogeneity calls for the use of finite mixture models, while continuous heterogeneity calls for the use of random-effects or random-parameter models. These two classes of model together play a central role in this book.

Of the "types" mentioned above that motivate the development of finite mixture models, probably the most important type in many situations is the "zero-type", that is, a subject who always contributes zero. In the context of a dictator game, such subjects may be labelled "selfish types", and in public goods games, they are labelled "free-riders". A very useful class of model that accounts for the existence of zero types is the "hurdle" framework. Hurdle models, or "double hurdle" models, contain two equations, the first determining whether a subject is a "zero-type", and the second determining behaviour given that they are not a zero-type. Hurdle models play an important role in this book and, in particular, the hurdle framework is extended to panel data so that it is applicable to the situation typically arising in experiments where multiple decisions are observed for each subject. One reason why the hurdle framework is so useful is that it allows for the possibility that a treatment changes a subject's type, as well as just changing his or her behaviour. For example, there is a lot of current interest in whether the subject's endowment is earned or unearned. It is quite possible that if the endowment is a free gift, they are less likely to be a "zero-type" than if it is earned (which is normally the case in "real life"). If experimental conventions are found to change a subject's *type* from his or her real-life type (in addition to simply changing his or her behaviour), the implications are clearly important for the external validity debate. This is something that may be tested easily within the hurdle framework.

An important point about "subject types" is that we are not at any stage in a position to say with certainty that a particular subject is of a given type. For example, a subject observed contributing zero on every occasion in a public goods game is very likely to be a "free-rider", but we cannot say with certainty that he or she is a free-rider. The best we can do is to compute *posterior type probabilities* for each subject, following estimation of the model. This is done using Bayes' rule. The subject contributing zero every time would presumably have a very high posterior probability of being a free-rider. In a situation of continuous heterogeneity, we can use a similar technique. For example, in a risky-choice model in which we assume continuous variation in risk attitude between subjects, we can use Bayes' rule following estimation to obtain a posterior risk aversion estimate for each subject. This measure is useful in a number of ways.

In a finite mixture model, the likelihood contribution corresponding to a particular subject is a weighted average of probabilities or densities corresponding to the different types, with the type probabilities (or *mixing proportions*) as weights. In the presence of continuous heterogeneity, the estimation problem is somewhat more complicated because the likelihood contribution corresponding to a single subject becomes an integral over the variable(s) representing the heterogeneity. Hence, the procedure for evaluating the likelihood function must incorporate some numerical method for the evaluation of integrals. The method adopted throughout this text is the method of maximum simulated likelihood (MSL, see Train, 2003). This method is based on the principle that an integral can be computed by evaluating the function at each of a set of simulated values of the variable of integration, and then taking the mean of these function values. The simulated variables are not random numbers, but instead *Halton draws*, which result in greater efficiency in the evaluation of the integral.

## 1.8 Experimetrics of Other-regarding Preferences

As behavioural economics has taken hold, the economics profession has moved away from rigid assumptions of self-interested utility-maximisation and started to



introduce concepts such as "other-regarding preferences" and "inequity aversion". Nevertheless, it is often recognised that whatever the considerations that are guiding behaviour, they can somehow be incorporated into the utility function that the individual is assumed to be maximising. For example, the utility function may contain a component that represents self-interest, and resembles the traditional utility function, and a second component that represents how much importance the individual attaches to the welfare of those around him or her. Optimisation of such a function results in the "other-regarding" behaviour that many experiments have set out to investigate. An obvious way of doing this in the setting of, for example, a dictator game is to assume a utility function with two arguments, "own payoff" and "other's payoff", specify the utility function parametrically, and estimate the parameters using econometric techniques. This sort of estimation has been performed by Andreoni & Miller (2002) and others. Within the context of these models, the question of whether a treatment, such as whether the initial endowment is earned or unearned, has an impact on behaviour can be addressed by investigating the way in which it impacts on the parameters of the utility function, rather than simply by considering whether it impacts on the outcome variable. Jakiela (2013) carries out treatment tests of this type.

The chapter in this book that covers other-regarding preferences takes this type of utility function as the starting point and considers a variety of estimation methods. Specifically, an extension to the model is developed that incorporates zero observations on other's payoff (i.e. selfishness) in a theoretically consistent way, by treating them as corner solutions to the dictator's constrained optimisation problem. Another extension is a finite mixture model, similar to that of Cappelen et al. (2007), that assumes different subjects have different *fairness ideals*, and their behaviour is determined by this together with their degrees of selfishness. Finally, a completely different estimation strategy is introduced, that is suitable when the data consists of choices between different allocations (Engelmann & Strobel, 2004). The appropriate model is the conditional logit model. This model is found to be particularly useful for estimating the parameters of the well known Fehr & Schmidt (1999) utility function, which separately captures aversion to advantageous and disadvantageous inequality.

## 1.9 Experimetrics of Bounded Rationality

In interactive games, the fundamental prediction usually takes the form of a Nash equilibrium. The Nash equilibrium is based on the assumptions that agents hold correct beliefs about others' actions, and that agents best respond to these correct beliefs. As always, observed behaviour departs from the theoretical prediction and we need to consider ways of modelling such departures. The models used for this purpose are often thought of as models of bounded rationality. For a recent survey, see Crawford et al. (2013).

One such approach is the quantal response equilibrium (QRE) model (McKelvey & Palfrey, 1995), which assumes that each player's behaviour follows a distribution which is a "noisy" best response to other players' noisy behaviour.

Note that the best-response assumption is being relaxed, since decisions are "noisy". However, the correctness-of-beliefs assumption is satisfied, since players have correct beliefs about others' noisy behaviour.

A different approach is the level-k model (Nagel, 1995), which assumes that players have different (finite) levels of reasoning, with each player believing that all other players have a level of reasoning one below their own. Clearly, this assumption implies that, unlike in QRE, players have *incorrect* beliefs about others' behaviour. Closely related to the level-k model is the cognitive hierarchy model (Camerer et al., 2003) which assumes, perhaps more reasonably, that players believe that other players do not all have the same level of reasoning, but are instead distributed over levels of reasoning below their own.

The most obvious approach to operationalising the level-k and cognitive hierarchy models is to assume a zero-mean error around the deterministic best response at each level of reasoning, which amounts to a relaxation of the best-response assumption. This is the approach followed by Bosch-Domènech et al. (2010) and Runco (2013), and is also the approach followed in the relevant chapter of this hook.

## 1.10 Experimetrics of Learning

If agents are found not to be operating exactly at the equilibrium, an obvious question arising is: is there a learning process by which they converge towards that equilibrium? This question may be addressed in settings in which each subject performs a sequence of tasks. The simplest way of addressing the question is to use the task number as an explanatory variable in a model for the decision variable. The effect of task number is then used to judge whether, and how quickly, behaviour is moving towards equilibrium as subjects gain experience.

Using the task number as an explanatory variable is all that is required in some situations. For example, in individual decision making, it is sometimes found that the parameters representing deviations from EU change with experience in a way that implies convergence towards EU (Loomes et al., 2002). It is also found that stochastic terms change with experience. In particular, the tremble probability is found to decay all the way towards zero in the course of an experiment, implying, reassuringly, that misunderstandings and complete losses of concentration are transitory phenomena. In public goods games in which a sequence of games are played, contributions are found to diminish with experience, and simply using the task number as an explanatory variable in the contribution equation is often found to be the best way of capturing this effect.

In those contexts, learning is (usually) only about the task, not about the behaviour of other players, nor even about the outcomes of previous tasks. A fairly standard feature of both of those settings is that, for reasons inherent to the experimental designs, subjects do not obtain feedback in terms of the outcomes of previous rounds. Hence, in those situations, the modelling of a learning process simply involves allowing particular parameters to depend in some way on the task number.



The situation in experimental games with feedback is very different. In each round, subjects observe both the chosen strategy of the opponent and the outcome in terms of their and others' pay-offs. Hence they learn directly about the behaviour of others, and also about the type of strategies that are most profitable for themselves. The process of learning about other players is complicated by the fact that other players' behaviour changes as they, too, gain experience. A comprehensive model of learning should therefore incorporate the effects of a player's own past pay-offs and also the effects of past choices by other players. In econometric terms the modelling strategy moves from *static* to *dynamic* modelling since we now need to capture explicitly the relationship between current behaviour and past behaviour and outcomes.

A number of such models are considered in this book. *Directional learning* theory, first proposed by Selten & Stoecker (1986), is a simple form of dynamic learning model, in which subjects are assumed to adjust their behaviour in each period in response to the outcome of the previous period. *Reinforcement learning* (Erev & Roth, 1998) is based on the idea that a player's propensity to choose a strategy is a positive function of the *pay-offs* received as a result of choosing that strategy in previous periods. *Belief learning* (Cheung & Friedman, 1997) is based on the idea that players adjust their strategies in response to payoffs that *would* have been received under each choice.

The reinforcement learning model has been used primarily by psychologists, while the belief learning model has been used primarily by decision and game theorists. A model which nests these two models is the *experience weighted attraction* (EWA) model developed by Camerer & Ho (1999). This model is useful because it forms a framework for testing which of the two models better fits the data. However, the model contains a large number of parameters, and might be seen as over-parameterised.

## 1.11 What is in this Book?

The main objective of this book is to show the reader clearly and methodically how to carry out a wide range of tasks in Experimetrics. What this book does *not* contain is any detail in econometric theory, for example the derivation of the properties of estimators and tests. For these topics, the reader is referred to mainstream econometrics textbooks, such as Wooldridge (2012) or Greene (2008).

Another objective that has *not* been set for this book is to provide a comprehensive survey of the literature on each topic covered. In the main part of each chapter, the only citations will be to studies that are directly relevant to the techniques being covered. At the end of each chapter, there is a section containing a limited number of suggestions for further reading for the benefit of readers wishing to go further with particular topics.

Tasks are all demonstrated in STATA version 12 (StataCorp, 2011). All data sets used in the text are available online (www.palgrave.com/moffatt), and are listed in Appendix A. Some of the data sets used for demonstration are real and have already been used in published work. Other data sets are simulated; please note that all simulated data sets contain the suffix **sim** in the name – it is important that if these data sets are used by readers, they are used only for practising techniques. Simulated data is clearly useful in situations in which suitable real data sets cannot be found. In fact, one of the chapters of the book is devoted to explaining how to simulate data sets with the required structure and features.

Mauy tasks can be carried out with existing STATA commands. Most of the STATA commands used in the book are listed in Appendix B. In certain situations, the required STATA commands do not exist, but user-written programs are available online which meet the requirement. These programs are found using the STATA `findit` command, from where they may be easily installed. For other tasks, programming in STATA is required. Basic STATA skills, such as how to create and run a do-file, are assumed, but the programs themselves will be explained in detail. In some situations MATA is used. MATA is a matrix programming language built into STATA. Some advanced tasks are only possible by including MATA commands within the STATA code.

In a small number of situations, Excel output is used, and some Excel files, also (www.palgrave.com/moffatt) are referred to. This is for particular types of problem that require flexibility in computation, for example in terms of outputs changing automatically in response to changes in inputs.

Exercises are included at the end of some chapters. The reason for the apparent lack of uniformity here is simply that some topics lend themselves more readily to exercises than others.

Chapter 2 considers the statistical aspects of experimental design that are most relevant to experimental economics. Most importantly, it provides a primer in "power analysis", the procedure used to select an adequate sample size for a treatment test. It also describes various types of design such as factorial, block, and within-subject designs, as well as explaining the difference between one-shot, partners, and strangers designs. Finally, it describes methods for administering multiple tasks per subject, namely the random lottery incentive scheme and the strategy method. Chapter 2 also introduces four very well-known experiments – the *ultimatum game*, the *dictator game*, the *trust game*, and the *public goods game* – which appear as applications many times throughout the book.

Chapter 3 covers treatment testing. This chapter contains many hands-on examples, some of which involve real data sets. It covers parametric and non-parametric tests, highlighting the strengths and weaknesses of the two approaches, and the circumstances under which each should be used. It also introduces the bootstrap, a method which makes a parametric test valid in a situation in which the underlying distributional assumptions do not hold. The chapter also covers tests comparing complete distributions and within-tests.

Chapter 4 considers treatment testing in the context of regression analysis. The applications are to data from auction experiments and contest experiments. A distinction is made between tests of the fundamental prediction of the theory, tests of comparative static predictions, and tests for the causes of out-of-equilibrium behaviour. The chapter includes methods for dealing with dependence, including clustering and the block bootstrap, and progresses to illustrations of panel data estimation and multi-level modelling. Finally, an example of a meta-analysis applied to contest experiments is provided.



Chapter 5 presents a very different application of regression analysis, here to the analysis of decision times. This is an area that has rapidly become popular in recent years, partly because decision time is a useful measure of the effort expended by the subject (see for example Moffatt, 2005*b*). Panel data estimators and tests are once again illustrated in this context.

Chapters 6–7 are mainly concerned with modelling approaches that are appropriate when outcomes are discrete variables, for example binary data models, censored data models, interval regression, and ordinal models. In nearly all of these situations, built-in STATA commands are available for the task. In addition, in Chapter 6, the ML routine (Gould et al., 2010) in STATA is introduced and applied to simple maximum likelihood problems.

Chapter 8 introduces finite mixture models, with further use of the ML routine. The final example uses real data from a public goods experiment and presents the first application of the ML routine to a panel data problem.

Chapter 9 is devoted to simulation of experimental data. This includes sections on how to simulate panel data, dynamic panel data, and binary panel data. As mentioned, simulated data is useful in situations in which real data is unavailable. Simulated data is also useful for the testing of programs, and for investigating the properties of estimators and tests. The latter is done using the Monte-Carlo technique. The use of Monte-Carlo is demonstrated in the chapter, including an application to the problem of evaluating the performance of a test statistic developed in Chapter 7.

Chapter 10 introduces the method of maximum simulated likelihood (MSL) which is adopted as the standard modelling framework for dealing with continuous heterogeneity. This chapter includes examples of MSL being applied to simulated data sets. Chapter 11 contains the first application of MSL to the estimation of the panel hurdle model, a model which allows for the presence of a "zero-type" in panel data settings. The panel hurdle model is applied to real data sets from previously published work. One strength of the hurdle framework that is emphasised is that it allows a treatment to change a subject's type in addition to altering his or her behaviour. Chapter 12 covers theoretical issues relating to choice under risk, preparing the ground for Chapter 13 which covers the econometric modelling of risky choice models. This is the second application of the MSL method.

Chapter 14 is concerned with optimal design of binary choice experiments. This chapter draws on the well-developed theory of optimal design from the statistics literature, and applies it to a particular type of experiment in economics.

Chapter 15 is concerned with the estimation of social preference models. The focus here is the estimation of the parameters of a utility function whose arguments are "own payoff" and "other's payoff". Real data from a dictator game are used for illustration. The parameters of the utility function are estimated using various different approaches, including: a model with binding non-negativity constraints to explain zero observations; a mixture model allowing individuals to differ in the nature of their fairness ideal; and a discrete choice model that uses data on subjects' choices between allocations.

Chapters 16–18 are concerned with the econometric modelling of data from experimental games. Chapter 16 explains the quantal response equilibrium (QRE) model, and illustrates its estimation using real data from the pursue-evade game. The QRE model is also applied to the contest data from Chapter 4. Chapter 17 develops the level-k and cognitive hierarchy models, explaining how they are estimated using simulated data from a guessing game. Chapter 18 covers a number of models of learning: directional learning (DL); reinforcement learning (RL); and belief learning (BL). It ends with an explanation of the experience weighted attraction (EWA) model, which is a heavily parameterised model nesting both RL and BL.

An important feature of the book is the close linkages between chapters. Because this is a subject area in which themes tend to emerge in different contexts, there is a good deal of cross-referencing between chapters. For example, the data set used in Chapter 13, on the econometric modelling of choice under risk, is simulated using the techniques described in Chapter 9, and estimation is conducted in the framework of MSL which is explained in detail in Chapter 10. Also, one of the by-products from the estimation of the risky-choice model in Chapter 13 is a measure of "closeness-to-indifference" which plays a very important role in the model of decision times estimated in Chapter 5.

Finally, there are three appendixes. Appendix A provides a list of the data sets and other files that are referred to in the book, and are available online at www.palgrave.com/moffatt. Appendix B provides a list of most of the STATA commands used in the book, with a brief explanation of each. Appendix C contains a table defining the 50 choice problems used in the (simulated) risky choice experiment analysed in Chapters 5 and 13.

# Chapter 2
# Statistical Aspects of Experimental Design in Experimental Economics

## 2.1 Introduction

This chapter is concerned with aspects of experimental design in experimental economics. In experimental economics, unlike in most other areas of economics, the data generating process is governed by the analyst. Because of this, assumptions required for the identification of a treatment effect are much less stringent than in other areas. In an experimental setting, the only major assumption required for identification is appropriate randomisation (with appropriate sample sizes). As pointed out by List et al. (2011), randomisation plays the role that an instrumental variable would play in a situation of naturally occurring data.

Issues of experimental design are relatively simple if the framework is one of treatment testing. One central design issue is the required sample size. A useful framework for choosing a sample size is power analysis. The principal objective is to find the sample size that is required to attain a pre-set level of power for a given treatment test. This chapter includes a primer on power analysis.

Other design issues include matching methods: partners versus stranger matching. This is closely related to the issue of clustering. We will also consider designs such as the Random Lottery Incentive (RLI) system and the Strategy Method (SM).

## 2.2 The Average Treatment Effect

This section provides a formal framework for the analysis of treatment effects.

It is usually recognised that every individual has his or her own treatment effect. Also, it is usually assumed that individual treatment effects vary randomly around an average. We are principally interested in estimating the *average treatment effect* (ATE).

Consider the effect of a particular treatment on a particular outcome variable $Y$. Let $T$ be a binary variable representing treatment status: $T = 1$ for treatment; $T = 0$ for control. Let $Y_i(T)$ be the outcome for subject $i$ given treatment status $T$.

We assume the following simple model for the outcome variable:

$$Y_i(T) = \alpha + \beta' X_i + \bar{\tau} T + \tau_i T + \epsilon_i \tag{2.1}$$

$X_i$ is a vector of observed individual characteristics (such as gender) which are thought to affect the outcome; $\bar{\tau}$ is the average treatment effect (ATE); $\tau_i$ is the subject-specific treatment effect, where $E(\tau_i) = 0$; $\epsilon_i$ is an independent and identically distributed (i.i.d.) random error term. The ATE may then be defined as:

$$\bar{\tau} = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)] \tag{2.2}$$

A problem that is central to treatment testing is that we cannot, in general, observe the two quantities appearing on the right hand side of (2.2). For a given $i$, we can only observe the quantities $E[Y_i(1)|T = 1]$ and $E[Y_i(0)|T = 0]$; that is, it is only possible to observe average behaviour under the treatment for those whom we choose to treat, while it is only possible to observe average behaviour without treatment for those whom we choose not to treat. If the propensity to receive treatment is correlated with observed or unobserved characteristics of the subject, the estimate ($\hat{\tau}$) of the ATE will be biased, since:

$$\hat{\tau} = E[Y_i(1)|T = 1] - E[Y_i(0)|T = 0] \neq E[Y_i(1)] - E[Y_i(0)] = \bar{\tau} \tag{2.3}$$

Randomisation is used to ensure that assignment to treatment is independent of other sources of variation, so that $E[Y_i(1)|T = 1] = E[Y_i(1)]$ and $E[Y_i(0)|T = 0] = E[Y_i(0)]$, giving equality in (2.3), implying that the estimated treatment effect is unbiased for the ATE.

A key assumption implicit in this framework is that the distribution of subject-specific treatment effect ($\tau_i$ in (2.1)) is "well-behaved", that is, has a bell-shaped and symmetric distribution arround the ATE. In certain situations, it is reasonable to expect this assumption to fail. For example, it may be that within the population of subjects, half respond to the treatment with a treatment effect of $+1.0$, while the remaining half are irresponsive to the treatment, and therefore have a treatment effect of zero. The ATE in this situation would obviously be $+0.5$ but this is a misleading measure of the effect of the treatment, since it is not close to the actual treatment effect of any individual subject. The best way to deal with such discreteness in the distribution of treatment effects is to apply the mixture modelling framework (McLachlan & Peel, 2000), in which it is assumed that there is more than one "subject type", with types differing from each other in the way in which they respond to the treatment, and with the proportion of each type in the population, the "mixing proportions", being estimated as additional parameters. The mixture modelling approach is one that is used many times in later chapters of this book.

## 2.3 Randomisation Techniques

As implied in the last section, the importance of randomisation is that it gives rise to a situation in which identification is not a problem. Here, we provide a discussion



of popular randomization techniques. For more detail the reader should consult List et al. (2011).

### 2.3.1 *Completely randomised designs*

The simplest of experimental designs is the completely randomised design. A random sample is drawn from the entire subject pool, and treatments are probabilistically assigned to subjects, independently of any observed or unobserved characteristics. The advantage of this procedure is that, by definition, it minimises the risk that treatment is correlated with subject characteristics. The disadvantages are that the sample sizes in each treatment are random, and the variance of the outcome may be large. Both of these factors tend to reduce the ability of the analyst to draw statistical inferences from the experimental data.

### 2.3.2 *Factorial designs*

An obvious solution to the problem of random sample sizes arising with the completely randomised design is to assign pre-determined numbers of subjects to each treatment, or to each combination of treatments. However, it is important that subjects are not assigned to treatments in the order in which they arrive, since time of arrival is likely to be correlated with subject characteristics. On arrival subjects should be given a random number determining treatment assignment, and recruitment should cease when all of the pre-determined targets are met.

Consider the following example of a giving experiment (e.g. a dictator game). Let us assume that there are two treatments: high stakes versus low stakes (with low stakes treated as "control"); and communication (between the two players) versus no communication (with no communication treated as "control"). Let us assume that we design the experiment with the numbers appearing in the following table; that is, with each *combination* of treatments being applied to 30 subjects, and a total of 120 subjects. Each cell in the table represents a "trial".

| | Low stakes | High stakes |
|---|---|---|
| No communication | 30 | 30 |
| Communication | 30 | 30 |

This is known as a "full-factorial design", because all possible treatment combinations are being covered. It might also be referred to as a "2 × 2 design". This sort of design is useful if, in addition to the two average treatment effects, we are interested in the "interaction" between the two treatments. For example, we might hypothesise that communication is less important when the stakes are higher, perhaps because financial incentives "crowd out" intrinsic motivation. In order to test such an interaction effect between the two treatments, the full-factorial design would be necessary.

However, if we are not interested in such an interaction effect, and only interested in the "main effects", that is, the effects of the treatments themselves, then the following design would suffice:

| | Low stakes | High stakes |
|---|---|---|
| No communication | 30 | 30 |
| Communication | 30 | 0 |

Here, only three trials are used, and the total number of subjects is only 90. This is known as a "fractional factorial design". If there were a larger number of treatments, the difference between full-factorial and fractional-factorial would become more substantial. To be precise, if there are $m$ different treatments, the full-factorial design would require $2^m$ trials, while a fractional-factorial design which identifies all $m$ main effects would only require $m+1$ trials. Clearly large savings in the sample size are possible in situations in which only main effects are of interest.

### 2.3.3 Block designs

If the subject pool is heterogeneous in certain observable dimensions, it may be advantageous to apply a block design. Experimental units are divided into blocks in accordance with the observable characteristics (contained in the vector $X_i$ in (2.1)). Then randomisation is performed within, but not between, blocks. The variable on which blocking is applied is known as the blocking factor. Typically, a blocking factor is a source of variability that is not of primary interest to the experimenter.

One obvious choice of blocking factor is gender. Gender may be an important source of variability in the outcome, and by blocking on it, this source of variability is controlled for, leading to greater accuracy in the estimation of the treatment effect(s) of central interest.

### 2.3.4 Within-subject designs

A within-subject design (or repeated measures design) can be thought of as a special case of a block design in which the experimenter blocks on a single subject, and the subject experiences more than one treatment. The considerable advantage of within-subject designs is that the impact of the subject-specific effect ($\alpha_i = \alpha + \epsilon_i$ from (2.1)) is essentially eliminated, and this has the potential to improve greatly the precision of the treatment effect estimate.

Formally, if $\hat{\tau}_{bs}$ is the between-sample estimate and $\hat{\tau}_{ws}$ is the within-sample estimate obtained using the same total number of observations, $N$, then (List et al., 2011):

$$V(\hat{\tau}_{ws}) = V(\hat{\tau}_{bs}) - \frac{2}{N} V(\alpha_i) \tag{2.4}$$



The interpretation of (2.4) is that if all subjects are identical, so that $V(\alpha_i) = 0$, there is no benefit from using a within-sample design, but if subjects differ greatly, so that $V(\alpha_i)$ is large, the benefits of the within-subject design are considerable.

A disadvantage of the within-subject design is the possibility of "order effects"; that is, the behaviour of subjects depending on the order in which the treatments are experienced.

### 2.3.5 Crossover designs

The problem of order effects may be addressed by varying the order of treatments between subjects. For example, if there are two treatments A and B, half of the subjects may be assigned to the sequence AB, and the other half to the sequence BA. Differences between these two groups would confirm the existence of an order effect, which would then need to be controlled for in treatment tests.

### 2.3.6 ABA designs

"ABA" refers to a design which starts with a baseline period in which no treatment is given (A), followed by a period in which the treatment is introduced (B), and then a period in which the treatment is removed so the behaviour can be observed a second time (A). This makes it possible to measure behaviour before treatment, during treatment, and once treatment is removed.

## 2.4 How Many Subjects? A Primer in Power Analysis

In deciding how many subjects to recruit, it is often useful to make use of power analysis (Cohen, 2013). Power analysis is the name given to the formal process for determining the sample size for a research study. The essence of the process is that it determines the sample size necessary for the test to achieve a specified *power*. The power of a test is the probability of detecting a "true" effect when a true effect actually exists.

The power analysis used in this section is based on the assumption that the outcome is a continuous. In Chapter 14, we tackle the perhaps more demanding problem of optimal design in a situation in which the outcome is binary.

### 2.4.1 The case of one sample

Suppose that we are interested in the continously distributed outcome measure $Y$ whose population mean is $\mu$. Suppose further that we are interested in testing the null hypothesis $\mu = \mu_0$ against the alternative hypothesis $\mu = \mu_1$, where $\mu_1 > \mu_0$.[1] We plan to collect a sample of size $n$ for this purpose, and we need to decide what $n$ should be. Before we do this, we need to set two quantities. The first is the test size, $\alpha$, which is the probability of rejecting the null hypothesis when it is true (or the probability of type I error). The second is the probability of failing to reject the null hypothesis when it is false (or the probability of type II error). This second probability is conventionally labelled $\beta$. Note that the probability of rejecting the null hypothesis when it is false is $1 - \beta$ and this is the *power* of the test. We shall denote power by $\pi$.

It has become standard to set $\alpha$ to 0.05, unless there are compelling reasons to do otherwise. Although there are no formal standards for power, many researchers assess the power of their tests using $\pi = 0.80$ as a standard for adequacy. The corresponding value of $\beta$ is 0.2. These conventions imply a four-to-one trade off between the probability of type II error and the probability of type I error.

Having decided on these values of $\alpha$ and $\beta$, we proceed to apply power analysis. The test that will be performed is the one-sample t-test, which is based on the following test statistic:

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \tag{2.5}$$

where $\bar{y}$ and $s$ are respectively the mean and standard deviation of the sample which is of size $n$. Under the null hypothesis, $t$ defined in (2.5) has a $t(n-1)$ distribution. Hence, the rejection rule, given our chosen value of $\alpha$, is $t > t_{n-1,\alpha}$.

Based on the anticipation that the value of $n$ eventually chosen will be reasonably large, the normal approximation may be used and the rejection rule becomes $t > z_\alpha$. This simplifies the analysis considerably.

The power of the test is given by:

$$\begin{aligned}
P(t > z_\alpha | \mu = \mu_1) &= P\left(\frac{\bar{y} - \mu_0}{s/\sqrt{n}} > z_\alpha \mid \mu = \mu_1\right) \\
&= P\left(\bar{y} > \mu_0 + \frac{z_{0.05}}{s/\sqrt{n}} \mid \mu = \mu_1\right) \\
&= P\left(\frac{\bar{y} - \mu_1}{s/\sqrt{n}} > \frac{\mu_0 + z_\alpha s/\sqrt{n} - \mu_1}{s/\sqrt{n}} \mid \mu = \mu_1\right) \\
&= \Phi\left(\frac{\mu_1 - \mu_0 - z_\alpha s/\sqrt{n}}{s/\sqrt{n}}\right)
\end{aligned}$$

[1] Alternative hypotheses nearly always involve inequalities, for example, $\mu > \mu_0$ or $\mu \neq \mu_0$. However, in the context of power analysis, it is necessary for both the null and the alternative hypotheses to be equalities, in order for the problem of finding the desired sample size to be properly defined. The value under the alternative is assumed to derive either from prior beliefs, from a previous study, or from a pilot study.



If the desired power of the test is $1 - \beta$, we then have:

$$\frac{\mu_1 - \mu_0 - z_\alpha s/\sqrt{n}}{s/\sqrt{n}} = z_\beta \tag{2.6}$$

Rearranging (2.6) we obtain:

$$n = \frac{s^2(z_\alpha + z_\beta)^2}{(\mu_1 - \mu_0)^2}$$

Recalling that our chosen values of $\alpha$ and $\beta$ are 0.05 and 0.20 respectively, we have $z_\alpha = 1.645$ and $z_\beta = 0.84$. Hence we may write the formula for the required sample size as:

$$n = \frac{6.17 s^2}{(\mu_1 - \mu_0)^2} \tag{2.7}$$

For an example of the use of the formula (2.7), suppose that we are testing the null $\mu = 10$ against the alternative $\mu = 12$, and we happen to know that the standard deviation of the data is 5. Then we apply formula (2.7) to obtain:

$$n = \frac{6.17 \times 5^2}{(12 - 10)^2} = 38.6$$

Clearly $n$ needs to be an integer, and in order to ensure that the power requirement is met (i.e. that the power is *at least* 0.8), we should round up rather than down. The required sample size in this example is therefore 39.

The STATA command `sampsi` can be used to perform this calculation directly. This is an example of one of STATA's "immediate" commands. An immediate command (which always ends with the letter i) is a command that obtains results not from the data stored in memory but from numbers typed as arguments. To conduct the analysis for the above example using the `sampsi` command, the required syntax is:

```
. sampsi 10 12 , sd(5) onesam oneside p(0.8)
```

The main arguments are the values under the null and alternative (10 and 12). The options are as follows: "sd(5)" indicates that the known standard deviation is 5; "onesam" indicates that a one-sample test is required; "oneside" indicates that a one-sided test is required; "p(0.8)" indicates that the required power is 0.8. The output from this command is as follows. Note that the required sample size is 39, in agreement with the calculation performed above.

```
Estimated sample size for one-sample comparison of mean
  to hypothesized value
Test Ho: m =   10, where m is the mean in the population
Assumptions:
         alpha =   0.0500  (one-sided)
         power =   0.8000
```

```
alternative m =        12
           sd =         5
Estimated required sample size:
            n =        39
```

### 2.4.2 Choosing the sample size in a treatment test

We now consider the slightly more complicated situation that is more usual in experimental economics, in which there are two samples, a control and a treatment, and the objective of the study is to discover whether there is a significant difference in the outcome between the two samples. Again power analysis can be used to determine the sample size that is required to meet this objective.

Let $\mu_1$ and $\mu_2$ be the population means of the control group and the treatment group respectively. The null hypothesis of interest is $\mu_2 - \mu_1 = 0$ (i.e. the treatment has no effect), and the alternative is $\mu_2 - \mu_1 = d$ (i.e. the treatment has an effect of magnitude $d$). $d$ is known as the "effect size" and it is necessary to specify its value at the outset in order for the problem of finding the required sample size to be properly defined. The chosen value of $d$ is assumed to be derived either from prior beliefs, from a previous study, or from a pilot study.

The testing procedure that is required to test the null hypothesis $\mu_2 - \mu_1 = 0$ is the independent samples t-test. If the two sample sizes are $n_1$ and $n_2$, the sample means are $\bar{y}_1$ and $\bar{y}_2$, and the sample standard deviations are $s_1$ and $s_2$, the independent samples t-test statistic is given by:

$$t = \frac{\bar{y}_2 - \bar{y}_1}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s_p$ is the "pooled" sample standard deviation and is given by:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

The distribution of $t$ under the null hypothesis is $t_{n_1+n_2-2}$. Again, matters are simplified using the normal approximation. We will therefore use the critical value $z_\alpha$.

In this two-sample situation, we clearly need to find two required sample sizes, $n_1$ and $n_2$ say, one for each sample. However, we start by constraining the two sample sizes to be equal, that is, $n_1 = n_2 = n$. The test statistic becomes:

$$t = \frac{\bar{y}_2 - \bar{y}_1}{s_p\sqrt{\frac{2}{n}}}$$



The power of the test is given by:

$$\begin{aligned}
P(t > z_{0.05} | \mu_2 - \mu_1 = d) &= P\left( \frac{\bar{y}_2 - \bar{y}_1}{s_p\sqrt{\frac{2}{n}}} > z_\alpha \middle| \mu_2 - \mu_1 = d \right) \\
&= P\left( \bar{y}_2 - \bar{y}_1 > z_\alpha s_p\sqrt{\frac{2}{n}} \middle| \mu_2 - \mu_1 = d \right) \\
&= P\left( \frac{\bar{y}_2 - \bar{y}_1 - d}{s_p\sqrt{\frac{2}{n}}} > \frac{z_\alpha s_p\sqrt{\frac{2}{n}} - d}{s_p\sqrt{\frac{2}{n}}} \middle| \mu_2 - \mu_1 = d \right) \\
&= \Phi\left( \frac{d - z_\alpha s_p\sqrt{\frac{2}{n}}}{s_p\sqrt{\frac{2}{n}}} \right)
\end{aligned}$$

If the desired power of the test is $1 - \beta$, we then have:

$$\frac{d - z_\alpha s_p\sqrt{\frac{2}{n}}}{s_p\sqrt{\frac{2}{n}}} = z_\beta \tag{2.8}$$

Rearranging (2.8) we obtain:

$$n = \frac{2s_p^2(z_\alpha + z_\beta)^2}{d^2}$$

Once again applying our chosen values of $\alpha$ and $\beta$, we have $z_\alpha = 1.645$ and $z_\beta = 0.84$, and we may write the formula for the required sample size as:

$$n = \frac{12.35 s_p^2}{d^2} \tag{2.9}$$

For an example of the use of formula (2.9), suppose that we are testing the effect size $d = 2$, and we know that the standard deviations of populations 1 and 2 are 4.0 and 5.84 respectively. Given that the two sample sizes are constrained to be equal, the pooled standard deviation is 5.0. Then we apply formula (2.9) to obtain:

$$n = \frac{12.35 \times 25}{4} = 77.2$$

Rounding up, we arrive at the required sample size (in each treatment) of 78.

The STATA syntax for the test just performed is:

```
. sampsi 10 12 , sd1(4.0) sd2(5.84) oneside p(0.8)
```

The main arguments are the values of $\mu_1$ and $\mu_2$. We could use any values here, provided their difference is 2 (the effect size). The options are the two standard deviations, and the request for a one-sided test. The output from this command is shown below. Note that the required sample size is in agreement with the calculation performed above.

```
Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (one-sided)
         power =   0.8000
            m1 =       10
            m2 =       12
           sd1 =        4
           sd2 =     5.84
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =       78
            n2 =       78
```

## 2.4.3 Treatments with unequal costs

In the above analysis, there was a constraint that the sample sizes were equal across the two treatments. One reason for relaxing this constraint would be a difference in the sampling costs between treatment and control. Suppose the focus of the experiment is on the effect of incentives, and there is a low incentive treatment and a high incentive treatment. It is logical that costs are expected to be higher in the latter treatment.

In this situation, in order to attain a desired level of power while respecting a cost constraint, the sample sizes should be set so that the ratio of the sample sizes is proportional to the square root of the cost ratio. Specifically, if the cost-per-subject in the two treatments are $c_1$ and $c_2$, it should be the case that:

$$\frac{n_2}{n_1} \propto \sqrt{\frac{c_1}{c_2}} \tag{2.10}$$

For example, let us suppose the cost per subject in the high-incentive treatment is four times as high as that in the low-incentive treatment. Applying (2.10), we obtain that there should be twice as many subjects in the low-incentive treatment, as in the high-incentive treatment.

This requires one more option (r(.)) in the sampsi command. With this option, the ratio $n_2/n_1$ is specified. If the low-incentive treatment is treatment 1, we here require $r = 0.5$. The command and output are shown below. Note that, as required, the desired low-incentive sample is (almost exactly) twice as large (131) as the desired high-incentive sample (66).



```
. sampsi 10 12 , sd1(4) sd2(5.84) oneside p(0.8) r(0.5)

Estimated sample size for two-sample comparison of means Test Ho: m1
= m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (one-sided)
         power =   0.8000
            m1 =       10
            m2 =       12
           sd1 =        4
           sd2 =     5.84
         n2/n1 =     0.50

Estimated required sample sizes:

            n1 =      131
            n2 =       66
```

Of course, if the experimenter really is working subject to a cost constraint, and the desired sample sizes turn out to be too large for the budget, it becomes necessary to relax the power requirement. For example, reducing the power from 0.80 to 0.60, we obtain considerably lower required sample sizes, which may be within the budget constraint:

```
. sampsi 10 12 , sd1(4) sd2(5.84) oneside p(0.6) r(0.5)

Estimated required sample sizes:

            n1 =       76
            n2 =       38
```

## 2.4.4 Sample size in cluster designs

Consider the following situation. You teach a module on which there are 300 students. The students are divided into 30 "seminar groups", each of size 10, which meet weekly. You wish to conduct an experiment and you wish to select the sample of subjects from the students taking the module. Let us suppose that you have used power analysis along the lines of the previous sub-sections, and you have determined that 60 subjects are required for your experiment. It would be administratively convenient randomly to select 6 of the 30 seminar groups, and to perform the experiment on the 60 students in the selected groups. This would be a "cluster design", in which the seminar groups are the "clusters".

Unfortunately, set against the convenience of the cluster design, there is a complication. If, as we might expect, outcomes are correlated within groups, a larger sample is required to obtain the desired level of power. To formalise this, we assume the following model, in which $u_j$ is a group-specific error term for group $j$:

$$Y_{ij}(T) = \alpha + \bar{\tau}T + u_j + \epsilon_{ij}$$

To attain the desired power, the sample sizes determined on the assumption of independent observations need to be inflated by the following factor (see List et al. (2011)):

$$1 + (c-1)\rho$$

where $c$ is the size of each cluster (10 in the example), and $\rho$ is the "coefficient of intracluster correlation", defined as:

$$\rho = \frac{var(u_j)}{var(u_j) + var(\epsilon_{ij})} \tag{2.11}$$

To understand equation (2.11), first imagine that there are no differences between groups, so that $var(u_j) = 0$ and therefore $\rho = 0$. The inflation factor (2.4.4) is then 1, meaning that the required sample size is the same as before. Next, imagine that there are differences between groups, and furthermore that all members of a group behave identically to each other, so that $var(\epsilon_{ij}) = 0$, and therefore $\rho = 1$. Then, again using (2.4.4), the sample size would need to rise by a factor $c$ (group size). This is because, in this extreme situation, sampling more than one subject from within a group is worthless, and the sample-size requirement simply becomes a requirement on the number of clusters.

In practice, the value of $\rho$ that we expect is a small positive number, representing modest intergroup differences. For example, if $\rho = 0.05$ in the example, the sample size would need to rise by a factor 1.45, that is, from 60 to 87. This would mean that we would need to increase the number of sampled seminar groups from 6 to 9.

## 2.5 Four Very Popular Experiments

Four of the most popular experimental settings are the ultimatum game, the dictator game, the trust game, and the public goods game. These, or combinations of them, appear many times as applications throughout the text. For this reason, we describe them in this section.

### *2.5.1 The ultimatum game*

The *ultimatum game*, introduced by Güth et al. (1982) is described as follows. Two players ("proposer" and "responder") bargain over the division of some fixed money amount ($100 say). The proposer moves first by offering some split of the pie (e.g. "I get $65; you get $35"). The responder then has the choice between two actions: they either "accept", in which case the proposed division is implemented; or they "reject", in which case both players receive zero. There is a unique subgame-perfect equilibrium in which the proposer offers a split that gives the responder the smallest allowable payoff (perhaps $1), and the responder accepts. There is extensive experimental evidence that behaviour deviates from the subgame-perfect equilibrium in predictable ways: proposals are often very generous, and sometimes specify an even split; responders frequently reject "low" offers. Camerer (2003) reports the overall findings that proposers offer an average of 40% of the money, and "low" offers of around 20% are rejected about half the time.



### *2.5.2 The dictator game*

The *dictator game* is a simplified version of the ultimatum game. In the dictator game, the proposer again determines an allocation of some fixed money amount. However, the responder simply receives the remainder left by the proposer. The responder's role is entirely passive; he has no strategic input towards the outcome of the game. It is sometimes said that the responder has no "power of veto" in the dictator game. Observed behaviour in this game allows a very straightforward test of the *homo economicus* model of individual behaviour: if individuals were only concerned with their own economic well being, proposers would allocate the entire money amount to themselves and would give nothing to the responder. There is much experimental evidence to reject this model: a significant proportion of dictators are observed giving positive amounts to the responder. On average, dictators give about 20% of the endowment to the responder (Camerer, 2003).

### *2.5.3 The trust game*

In the *trust game* (Berg et al., 1995), the two players are "sender" and "recipient". The sender has an endowment, and he or she has the opportunity to send some or all of it to the recipient. The experimenter then (typically) triples the amount sent. After the recipient receives the transfer (i.e. three times the amount sent), he or she may return money back to the sender. The amount sent by the sender is a natural measure of "trust", while the amount returned by the recipient is (after controlling for the amount sent) a natural measure of "trustworthiness".

The trust game is related to the dictator game in the sense that the recipient is a dicator whose endowment has been given to him or her by the sender.

In the trust game thus decribed, the unique subgame-perfect equilibrium is such that the sender sends zero to the recipient. This is because the sender has no reason to send any positive amount given that they expect the recipient to send zero in return. Once again, experimental evidence departs from this prediction in more than one way. According to the meta-analysis of Johnson & Mislin (2011), senders send on average around 50% of their endowment, signalling a modest degree of trust, while recipients return around 37% of the (tripled) amount received, which is (just) sufficient for the trust to "pay".

### *2.5.4 The public goods game*

A public goods experiment is typically conducted using what is known as the voluntary contributions mechanism (VCM), outlined as follows. Subjects are divided into groups of $n$ members. Each group member has an endowment of $E$ tokens, that they must divide between a private account and a public account. Each token that a subject allocates to her own private account earns one point for her (and nothing for anyone else); in contrast, each token that any group member allocates to the public account is multiplied by $m$ and then divided equally among all $n$ group members. Hence, for each token the group member allocates to the public account, *every* group member earns $m/n$ units. This ratio is called the marginal per capita return (MPCR). It is usually the case that $n > m > 1$, so that the MPCR, though positive, is strictly less than unity.

The game has a unique Nash equilibrium consisting of zero contributions by every subject. This is obvious when considering that while the subject earns one unit when allocating a token to her private account, she earns an amount $m/n < 1$ when allocating the same token to the public account. Hence, regardless of the allocations of other group members, each subject maximises her own pay-off by allocating all of her endowment to her private account.

It is also obvious that this Nash equilibrium is socially inefficient, as it yields $E$ units per group member when, had all members contributed their full endowment to the public account, each would have received $mE > E$ units. Note that if the group consisted of only two players, the situation would be very similar to the well-known prisoner's dilemma.

Experiments using the VCM procedure described above have been surveyed by Ledyard (1995). One overall finding is that the average subject contributes around 40% of his or her endowment to the public account. When analysing data from such games, our primary interest will be in identifying the motivations behind such contributions. One motivation to which we shall pay particularly close attention is the reciprocity motive: subjects' contributions depending positively on previous contributions by others. Another phenomenon of interest will be the widely observed decay in the level of contributions as the game is repeated.

## 2.6 Other Aspects of Design

### 2.6.1 The random lottery incentive (RLI) mechanism

The random lottery incentive (RLI) mechanism is an elaborate form of within-subject design. Each subject completes a number of distinct decision tasks. The tasks might be valuations of lotteries, choices between pairs of lotteries, or choices of strategy in a game. Each task typically has a well-defined reward structure where the payoff is a function of the choice made, and also of the moves of nature (in games of chance), or of the moves of other players (in games of strategy). At the beginning of the experiment, the subject is made aware that when the sequence of tasks is completed, *one* of the tasks will be selected at random, and the payoff to the subject will be the outcome of the selected task.



Since only one task will be for real, the RLI is likely to encourage subjects to think about each task as if it were for real, and as if it were the only task faced. If subjects do think in this way, then the mechanism will have the desirable consequence of eliminating wealth effects that could arise if payoffs from tasks influence subsequent decisions.

The RLI may also be administered as a form of "crossover" design if the order of tasks is varied between subjects. This is very useful for identifying the role of experience in decision making.

### 2.6.2 The strategy method

The strategy method, introduced by Selten (1967), is a procedure for eliciting subjects' responses in games. The ultimatum game provides a useful context in which to describe this method. In standard implementations of the ultimatum game, the data generated consists of *observed decisions*; that is, the proposer's offer and the responder's response are recorded. This is known as the "direct decision approach". Notice that a limitation of this approach is that it only reveals the responder's decision and no more: we observe their response to the proposer's actual offer, but we do not observe how they would have responded to different offers that could have been made.

The basic principle of the strategy method is to ask each player to reveal their entire *strategy*. This requires the responder to give a conditional response to each offer that could be made, before knowing what offer has actually been made. Once the strategy has been elicited, the game can be played out by implementing the strategy.

The obvious advantage of the strategy method is that the information obtained is richer since it contains information about how agents would behave at information sets that might not arise in the actual course of play. This is particularly relevant in the case of the ultimatum game because the strategy method allows us to observe how agents respond to very low offers, despite low offers being very rare.

### 2.6.3 "One-shot", "partners", and "strangers" designs

Theories in their purest form tend to assume that the game is played only once, and also that players are fully rational with correct beliefs, and do not require experience of the game. If we are really only interested in subjects' behaviour in a single play of the game, we simply ask them to play one time against opponents randomly selected from the pool of subjects present in the current session. This design may be labelled the "one-shot" design.

However, in game theory experiments, it is more usual for subjects to play repeatedly over a sequence of rounds, for the obvious reason that this generates more data, and also because it provides an opportunity for subjects to gain experience. The reason why we are keen for subjects to gain experience is that we are often more interested in the behaviour of experienced subjects than that of inexperienced subjects. We are also interested in the process by which the subjects gain experience, that is, the *learning* process.

The question which then arises is how exactly the repetition should be administered. There are two popular types of protocol. In *partners* designs (sometimes called *fixed rematching designs*), the same group of subjects play together in every round. In this situation, there are two possible reasons why behaviour might change between rounds: learning (i.e. subjects require experience in order to gain an understanding of the incentives of the game); and strategic considerations (i.e. a subject's behaviour changes as she updates her beliefs about other players in the group). The other type of protocol is the *strangers* design (sometimes called the *random rematching design*). Here, the groups who play together are randomly reselected (from the pool of subjects present in the current session) each round. As suggested by Andreoni (1988), the purpose of the strangers design is to separate the effect of learning from that of strategic considerations. If behaviour changes between rounds under a strangers design, it is reasonable to attribute the change to learning alone, since strategic considerations are absent.

There are different forms of strangers design, according to whether restrictions are imposed on the random process that reselects groups. The term *perfect strangers* is often used for a design that restricts the random process in such a way that the probability of any two players playing each other twice is zero. However, this leaves scope for the possibility that player $i$ anticipates that a player ($j$) that she is playing with in round $t$ will meet another player ($k$) in round $t+1$, and that she ($i$) might meet $k$ in round $t+2$. Hence it is conceivable that subjects make choices on the basis that these choices might indirectly influence the behaviour of subjects they will meet in future rounds. Arguably, the term *perfect strangers* should be reserved for designs in which, in each round, each subject $i$ plays subjects who have not previously met $i$, or any of those whom $i$ has met, or any of those whom they have met, and so on. Such a design would completely eliminate the possibility of player $i$'s choices in round $t$ influencing the choices of players with whom $i$ plays in future rounds. However, such a design is considerably more complex and costly to implement than unconstrained random rematching.

## 2.7 Summary and Further Reading

A key concept covered in this chapter is randomisation. This, and other issues central to experimental design in economics, is covered in detail by List et al. (2011) and Green & Tusicisny (2012).

The point was made early in the chapter that randomisation is important because it gives rise to a situation in which identification is not a problem. However, it should be recognised that a problem that remains is generalisability (also known as external validity). According to Al-Ubaydli & List (2013), there is a tradeoff between identification and generalisability. Some popular randomisation techniques



have been discussed. For more detail on these the reader should consult List et al. (2011).

Power analysis has also been explained in some detail in this chapter. For further detail, the reader is referred to Cohen (2013).

For more information on the "popular experiments" described in Section 2.5, and also on the various aspects of experimental design covered in Section 2.6, the reader is referred to the relevant sections of Bardsley et al. (2009).

## Exercises

1. Burnham (2003) considers a "photograph" treatment in a dictator game. Given the effect sizes reported, find the optimal sample sizes.
2. What assumptions are required for the random lottery incentive system to be incentive compatible?

# Chapter 3
# Treatment Testing

## 3.1 Introduction

The last chapter was concerned with issues of experimental design in the context of treatment testing. The most important design feature was the choice of sample size in treatment testing. This chapter is concerned with methods for implementing treatment tests, having implemented the design and collected the data.

There are two broad approaches to treatment testing. The first is the "between-subject" approach, in which the sample is divided into two groups, the treatment group, to whom the treatment is applied, and the control group, to whom no treatment is applied. An outcome measure is recorded for each subject. The other approach is the "within-subject" approach. This is to obtain the outcome measure from each member of the sample, both without and with the treatment. Whichever of these two approaches is followed, the central question that is always being addressed is whether the treatment influences the outcome, and if so, in which direction.

Having decided between the "within-subject" and the "between-subject" approaches, there are then many different ways in which the test can be performed. Another broad division is between parametric and non-parametric tests. Both types of test have advantages and disadvantages.

The key factor in the choice between non-parametric and parametric tests is the scale of measurement of the data. There is a large literature on this point (see, for one example, Harwell & Gatti, 2001). There are essentially three scales of measurement: nominal, ordinal, and cardinal. Parametric tests, to an extent, rely on distributional assumptions which can only hold if the variables in question are measured on a cardinal scale.

Even if measurement is on a cardinal scale, some experimental economists seem uncomfortable with the use of parametric tests, because they worry that the distributional assumptions may not be met. There are two important responses to this concern. Firstly, provided that the number of observations in each treatment is sufficiently large, it is possible to appeal to the central limit theorem (CLT) which, under certain conditions, implies that the (standardised) mean of a sample follows a normal distribution even when the sample is drawn from a distribution that is not normal (see e.g. Berenson et al., 1988). Secondly, even in a situation in which the CLT cannot be relied upon (e.g. low sample sizes), a method is available for ensuring that inferences made on the basis of parametric tests are valid regardless of the distribution of the data. This method is the bootstrap.

As stressed in Chapter 1, an important issue in Experimetrics is dependence, mainly with regard to the multiple observations per subject that are typically analysed. In this chapter, however, we sidestep such problems, by restricting attention to situations in which only one observation is available per subject (or two observations per subject in a within-subject test). The problem of treatment testing in the presence of dependence will be covered in the next chapter.

## 3.2 The Mechanics of Treatment Testing

Siegel & Castellan (1988) provide a useful summary of the mechanics of treatment testing. Here, we shall be somewhat brief.

A treatment test always has a null hypothesis and an alternative hypothesis. The null hypothesis is generally the hypothesis that there is no effect. The alternative hypothesis is that there is an effect. If the alternative hypothesis specifies the direction of the effect, it is a one-sided alternative and we conduct a one-tailed test. Otherwise it is a two-sided alternative and we conduct a two-tailed test. One-sided alternatives are usually proposed when the researcher has a prior belief about the direction of the effect, the prior belief perhaps coming from economic theory. The first stage of the application of the test is to compute the test statistic which is a function of the data values. Then the test statistic is compared to the null distribution (i.e. the distribution that the statistic would in theory follow if the null hypothesis were true). If the test statistic falls in the rejection region, the null hypothesis is rejected in favour of the alternative. If the test statistic falls elsewhere, the null hypothesis is not rejected, and it may be concluded that the test result is consistent with the null hypothesis. The rejection region is determined by whether the test is two-tailed or one-tailed, and by the chosen "size" of the test. The "size", usually denoted as $\alpha$, is the probability of rejecting the null hypothesis when it is true, and this is normally set to 0.05. The point at which the rejection region starts is referred to as the critical value of the test.

The p-value of the test is the probability of obtaining a test statistic that is more extreme than the one obtained. The p-value is useful because it allows a conclusion to be drawn without comparing a test statistic to a critical value (i.e. it avoids the need to consult statistical tables). The p-value represents the strength of evidence in favour of the alternative (i.e. evidence of an effect). The words used to represent "strength of evidence" are a matter of individual taste. Popular terminology is: if $p<0.10$, there is *mild* evidence of an effect; if $p<0.05$, there is *evidence*; if $p<0.01$, there is *strong* evidence; if $p<0.001$, there is *overwhelming* evidence.

Note that, in addition to considering whether there is an effect, and the strength of evidence of the effect, none of this is any use without also reporting the *direction* of the effect. As mentioned, a prior belief about the direction of an effect leads to a one-tailed test. For a one-tailed test (assuming the test statistic has the expected sign) the p-value is half of the p-value for the corresponding two-tailed test. Hence



one-tailed tests are more likely to find evidence of an effect. This is the value of prior beliefs in the form of economic theories.

## 3.3 Testing with Discrete Outcomes

### 3.3.1 The binomial test

We will commence with what is perhaps the simplest of all tests.

Consider the following choice problem, where the two circles represent lotteries, and the areas within them represent probabilities of the stated outcomes. The left-hand lottery is the "safe" lottery and it pays \$5 with certainty. The right-hand lottery is the "risky lottery" and represents a 50:50 gamble involving the outcomes \$0 and \$10, as shown in Figure 3.1. Clearly, by choosing between these lotteries, a subject is conveying information about his or her attitude to risk. In particular, note that, since the two lotteries have the same expected value (\$5) (and assuming subjects obey expected utility theory), a risk-averse subject would choose S, a risk-seeking subject would choose R, and a risk-neutral subject would be indifferent between S and R. Accordingly, if all individuals were risk-neutral, we might expect 50% to choose the safe lottery, and 50% to choose risky. If, as is commonly found, individuals are predominantly risk averse, we would expect more than 50% to choose the safe lottery.

On this basis, the choices between the two lotteries of a sample of subjects may be used to conduct a test of the hypothesis of risk neutrality. The file **lottery_choice_sim** contains information on a sample of 30 subjects. The data set has 30 rows, one per subject, and one of the variables is y, consisting of ones and zeros representing choice of S and R respectively. The following command reveals that 21 of the 30 subjects chose S:

```
. tab y

          y |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |          9       30.00       30.00
          1 |         21       70.00      100.00
------------+-----------------------------------
      Total |         30      100.00
```

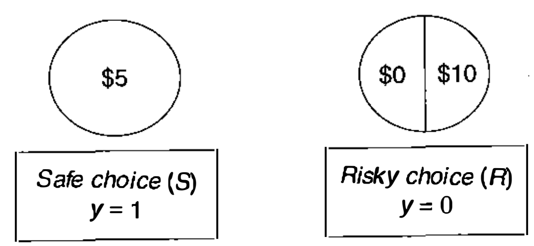



Figure 3.1: A typical lottery choice problem

If the null hypothesis (risk neutrality) is true, the probability of any subject choosing S is 0.5. Therefore, if the null hypothesis is true, the probability of 21 or more of the 30 subjects choosing S is the following sum of binomial probabilities:

$$P(N_s \geq 21) = \sum_{n=21}^{n=30} \frac{30!}{n!(30-n)!}(0.5)^{30} = \underline{\underline{0.0213}} \qquad (3.1)$$

Because this probability is smaller than 0.05, we may conclude that there is evidence that the null hypothesis is false, that is, that subjects are not risk-neutral (and, more precisely, that they are risk-averse).

The test that has just been performed is the binomial test, and the probability that has been computed is the p-value of the test. This test is used when the outcome is binary, and when the null hypothesis can be expressed simply in terms of the probability of one of the two outcomes.

The binomial test can be performed easily in STATA using the `bitest` command. Applying it to the present data set, we obtain the results:

```
. bitest y==0.5

    Variable |        N   Observed k   Expected k   Assumed p   Observed p
-------------+------------------------------------------------------------
           y |       30         21           15       0.50000      0.70000

  Pr(k >= 21)           = 0.021387  (one-sided test)
  Pr(k <= 21)           = 0.991938  (one-sided test)
  Pr(k <= 9 or k >= 21) = 0.042774  (two-sided test)
```

Three different p-values are provided. The first one is the same as the one computed above. Note that the test performed above is a one-sided test because the alternative hypothesis is that agents are risk averse, and hence we have a prior expectation that the proportion of safe choices will be greater than 0.5. If we had no such prior belief, we would use the two-sided test. Note that the p-value for the two-sided test is exactly two times that of the one-sided test. Note also that it is less than 0.05, meaning that even in the absence of the prior belief, we still have evidence in the sample that agents are not risk-neutral.

### 3.3.2 *Fisher's exact test*

Continuing with the same example, let us now assume that information on the subjects' gender (represented by the variable "male"; 1 if male, 0 if female) is also available. A cross-tabulation of gender and choice is obtained using the "tabulate" command:

```
. tab y male, col

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+
```



```
           |         male
         y |         0          1 |     Total
-----------+----------------------+----------
         0 |         1          8 |         9
           |      8.33      44.44 |     30.00
-----------+----------------------+----------
         1 |        11         10 |        21
           |     91.67      55.56 |     70.00
-----------+----------------------+----------
     Total |        12         18 |        30
           |    100.00     100.00 |    100.00
```

We see that, of the 30 subjects, 12 are female and 18 are male. We also see that of the 12 female subjects, 11 (91.67%) chose S, while of the 18 males, 10 (55.56%) chose S. Note that these (column) percentages are shown as a result of using the "col" option. The difference in percentage choosing S appears to indicate that females are more risk averse than males. This difference is what needs to be tested for statistical significance.

For this purpose, we may use Fisher's exact test. This test asks what is the probability of obtaining the combination of numbers in the tabulation, or a more extreme combination, for the given row totals and column totals. To consider how such a probability might be computed, let us consider the following tabulation:

| | Male=0 | Male=1 | Total |
|---|---|---|---|
| $Y=0$ | A | B | A+B |
| $Y=1$ | C | D | C+D |
| Total: | A+C | B+D | A+B+C+D |

The probability of obtaining this combination of numbers A, B, C, and D, for the given row and column totals, is given by:

$$P = \frac{\binom{A+C}{A}\binom{B+D}{B}}{\binom{A+B+C+D}{A+B}} = \frac{(A+B)!\,(C+D)!\,(A+C)!\,(B+D)!}{(A+B+C+D)!A!B!C!D!} \qquad (3.2)$$

Applying (3.2) to the numbers in the cross-tabulation, we have:

$$P = \frac{9!21!12!18!}{30!1!8!11!10!} = 0.0367$$

Next, we need to ask what is the probability of obtaining a "more extreme" combination than the one above (i.e. one for which females appear to be even more risk-averse), for the given row and column totals. There is only one such combination:

| | Male=0 | Male=1 | Total |
|---|---|---|---|
| $Y=0$ | 0 | 9 | 9 |
| $Y=1$ | 12 | 9 | 21 |
| Total: | 12 | 18 | 30 |

Applying (3.2) to this combination, we obtain a probability of 0.0034. The probability that is required (i.e. the p-value for Fisher's exact test) is therefore:

$$0.0367 + 0.0034 = 0.0401$$

The test is performed in STATA using the "exact" option with the tabulate command:

```
. tab y male, col exact

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |         male
         y |         0          1 |     Total
-----------+----------------------+----------
         0 |         1          8 |         9
           |      8.33      44.44 |     30.00
-----------+----------------------+----------
         1 |        11         10 |        21
           |     91.67      55.56 |     70.00
-----------+----------------------+----------
     Total |        12         18 |        30
           |    100.00     100.00 |    100.00

           Fisher's exact =                 0.049
   1-sided Fisher's exact =                 0.040
```

Two p-values are given. The "1-sided Fisher's exact" p-value is 0.040 and this agrees with the p-value computed above. Because it is less than 0.05, and because it is a one-sided test, we conclude that there is evidence that females are more risk-averse than males.

The other p-value is a two-tailed p-value. This is the p-value that we would use if we were testing for a gender effect in the absence of any prior belief over the direction of the effect. Computing this two-tailed p-value is slightly awkward. We need to add to the one-tailed p-value the probability of obtaining a more extreme outcome in the "other" direction. To determine whether an outcome is more extreme, we use the difference between the proportion of females choosing S and the proportion of males choosing S. For the data, this difference is $0.9167 - 0.5556 = 0.3611$.

We now need to consider various other possible outcomes. Consider:

| | Male=0 | Male=1 | Total |
|---|---|---|---|
| $Y=0$ | 9 | 0 | 9 |
| $Y=1$ | 3 | 18 | 21 |
| Total: | 12 | 18 | 30 |

For this outcome, the difference in proportions is $0.25 - 1 = -0.7500$. This is clearly more extreme than (and in the opposite direction to) 0.3611. The outcome

| | Male=0 | Male=1 | Total |
|---|---|---|---|
| $Y=0$ | 8 | 1 | 9 |
| $Y=1$ | 4 | 17 | 21 |
| Total: | 12 | 18 | 30 |



has a difference in proportions of $0.3333 - 0.9444 = -0.6111$. Again this is more extreme. The outcome

| | Male=0 | Male=1 | Total |
|---|---|---|---|
| $Y=0$ | 7 | 2 | 9 |
| $Y=1$ | 5 | 16 | 21 |
| Total: | 12 | 18 | 30 |

has a difference in proportions of $0.4166 - 0.8888 = -0.4722$. Yet again this is more extreme.

When the probability function (3.2) is applied to each of these three outcomes, we obtain (respectively) 0.00001, 0.00062, and 0.00847. The p-value for the two-sided test is therefore:

$$0.040 + 0.00001 + 0.00062 + 0.00847 = 0.0491.$$

This number is in agreement with the "Fishers exact" p-value given in the STATA output above. The fact that it is less than 0.05 indicates that there is evidence of a gender effect even in the absence of a prior belief over the direction of the effect.

### 3.3.3 *The chi-squared test*

The gender effect may alternatively be tested using the chi-squared test. This requires the `chi2` option with the `tab` command:

```
. tab y male, col chi2

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |         male
         y |         0          1 |     Total
-----------+----------------------+----------
         0 |         1          8 |         9
           |      8.33      44.44 |     30.00
-----------+----------------------+----------
         1 |        11         10 |        21
           |     91.67      55.56 |     70.00
-----------+----------------------+----------
     Total |        12         18 |        30
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =   4.4709   Pr = 0.034
```

We see that the p-value of this test is 0.034, slightly lower than that of Fisher's exact test, and provides further evidence of a gender difference in risk attitude.

Let us consider how the chi-squared statistic is computed. We need to ask what we would *expect* the numbers in the table above to be if the null hypothesis were true (i.e. if there really were no effect of gender). The answer is:

*Expected* frequencies:

```
           |          male
         y |         0          1 |     Total
-----------+----------------------+----------
         0 |       3.6        5.4 |         9
           |     30.00      30.00 |     30.00
-----------+----------------------+----------
         1 |       8.4       12.6 |        21
           |     70.00      70.00 |     70.00
-----------+----------------------+----------
     Total |        12         18 |        30
           |    100.00     100.00 |    100.00
```

We label the numbers in the first table O (for observed) and those in the second E (for expected), and we compute the following sum over the four cells in the table:

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(1-3.6)^2}{3.6} + \frac{(11-8.4)^2}{8.4} + \frac{(8-5.4)^2}{5.4} + \frac{(10-12.6)^2}{12.6} = 4.4709$$

The test statistic has a $\chi^2$ distribution under the null hypothesis of no effect. What are the degrees of freedom? Here we need to ask how many of the four entries in the table are "free". Notice that if you fix the two entries in one of the rows, the two entries in the other row are determined (by the column totals); also, if you fix one of the entries in a row, the other is determined (by the row total). This means that there is only one free number in the table, and this is the degrees of freedom for the test. Generally, if the cross-tabulation has $m$ rows and $n$ columns, the degrees of freedom for the chi-squared test is $(m-1)(n-1)$.

Recall the critical values of the $\chi^2$ distribution:

| Degrees of freedom | 5% point of $\chi^2$ |
|---|---|
| 1 | 3.84 |
| 2 | 5.99 |
| 3 | 7.82 |
| 4 | 9.49 |
| 5 | 11.07 |
| 6 | 12.59 |

Here, we reject $H_0$ if $\chi^2 > 3.84$, which it is. As previously noted, this tells us that we have evidence of a difference between genders.

The p-value is the area to the right of the test statistic under the $\chi^2(1)$ distribution, i.e. p-value $= P(\chi^2(1) > 4.4709)$. The p-value of 0.034 being somewhat less than 0.05 is consistent with the test statistic being somewhat to the right of the 0.05 critical value. Note that this p-value may be computed exactly using =CHIDIST(4.4709,1) in Excel. This gives the same answer as STATA, of 0.034.



There is another way of carrying out the chi-squared test in STATA. Let us imagine that the complete data set is not available, and all that is available is the numbers in the cross tabulation. To carry out the chi-squared test with only this information, we would use the "immediate" STATA command `tabi`. Recall that an "immediate" command (which always ends with the letter i) is a command that obtains results not from the data stored in memory but from numbers typed as arguments. One such command seen previously in Chapter 2 is the `sampsi` command for finding the required sample size for a particular test. To reproduce the result of the last test using the command tabi, we use:

```
. tabi 1 8 \ 11 10, chi

           |          col
       row |         1          2 |     Total
-----------+----------------------+----------
         1 |         1          8 |         9
         2 |        11         10 |        21
-----------+----------------------+----------
     Total |        12         18 |        30

          Pearson chi2(1) =   4.4709   Pr = 0.034
```

### 3.3.4 The chi-squared test on a real data set

Various tests, starting with the chi-squared test, will be demonstrated in the context of the "hold-up" problem, using a data set from the experiment conducted by Ellingsen & Johannesson (2004). This data set is available in the STATA file **holdup**.

We shall start by describing the "holdup" experiment. There are two players: "seller" and "buyer". The game consists of three stages:

1. The seller is given 60 units, and has the opportunity to invest it in order to create the greater amount of 100 units. Note that this is a dichotomous choice: either he invests 60, or he does not invest at all.
2. If the seller has invested, the buyer proposes how to split the 100 units between the two players.
3. The seller chooses whether to accept the buyer's proposal, realising the proposed split, or to reject it. If the seller rejects, both parties receive zero, leaving the seller with a net loss of 60.

First of all, note that the holdup experiment is a combination of the trust game described in Section 2.5.3 and the ultimatum game described in Section 2.5.1. To be precise, the first two stages of the holdup experiment constitute a trust game, although it should be termed a "binary trust game" since the first mover's decision is either to invest or not to invest and is therefore binary. Stages 2 and 3 of the holdup experiment amount precisely to the ultimatum game.

Given conventional assumptions of self-interest, the unique subgame perfect equilibrium outcome is that the seller chooses not to invest. Why? At the last stage of the game, the seller should accept any proposal that gives him more than 0 units.

Hence there is no reason for the buyer to offer any more than 1 unit. Hence the seller expects to lose 59 units if he invests. So it is irrational for the seller to invest.

As usual, experimental findings cast serious doubt on the theory: about one-third of sellers choose to invest, and they often benefit from doing so.

Ellingsen & Johannesson (2004) focus on the effect of *communication* between buyer and seller. To this end, they consider three treatments:

Treatment 1 (T1): No communication except the actions themselves.

Treatment 2 (T2): Buyer can send a message to seller, before seller makes an investment decision (presumably, a message of the form "I am a principled person; if you reveal your trust in me by investing, I promise to reward you").

Treatment 3 (T3): Seller can send a message to buyer, along with the investment decision. (Presumably, the wording would be along the lines of "I am investing because I trust you, even though I do not know who you are. However, I also want you to know that I am not stupid; if I fail to benefit as a result of trusting you, I will make sure you receive zero".)

Clearly, the purpose of Treatment 2 is to assess the impact of *promises*, while Treatment 3 is to assess the impact of *threats*. Of course, neither promises nor threats alter the theoretical prediction of no investment. The issue is whether they have an impact on the actual decisions of either sellers or buyers.

*Research Question 1: What impact does communication have on the seller's decision of whether to invest?*

Here, we simply look at the proportion of investors for each treatment. The best way to look at these is in a cross-tabulation:

```
. tab invest treatment, col chi2

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

    sellers |
 investment |
  decision: |
       1 if |
   invest; 0 |            treatment
     if not |         1          2          3 |     Total
-----------+---------------------------------+----------
         0 |        26         14         12 |        52
           |     65.00      46.67      36.36 |     50.49
-----------+---------------------------------+----------
         1 |        14         16         21 |        51
           |     35.00      53.33      63.64 |     49.51
-----------+---------------------------------+----------
     Total |        40         30         33 |       103
           |    100.00     100.00     100.00 |    100.00

          Pearson chi2(2) =   6.1788   Pr = 0.046
```

Note that there are 103 pairs of subjects in the experiment. The overall proportion of investors is very close to 50%. However, there appear to be major differences between treatments: the proportion of investors is highest in Treatment 3 (64%) and lowest in Treatment 1 (35%). This tells us that communication does have a favourable impact on the investment decision, especially communication in the form of the seller himself being in a position to make a threat (although we shall return later to the difference between the two types of communication).



To assess whether this difference between treatments is statistically significant, we conduct a chi-squared test, by including the `chi2` option in the STATA command above. We see that the p-value of this test is 0.046, indicating that there is evidence (although not strong evidence) that communication has an impact on the investment decision. An explanation of how the chi-squared test statistic is computed was provided in Section 3.3.3 above.

Note that the degrees of freedom for this test is 2. This is because the number of rows ($m$) is 2 while the number of columns ($n$) is 3, and the degrees of freedom is given by $(m-1)(n-1)$. As explained in Section 3.3.3, this essentially means that only two of the numbers in the table are "free"; the other four can always be deduced, with knowledge of the row and column totals.

Here, we reject $H_0$ if $\chi^2 > 5.99$, which it is. As previously noted, this tells us that we have evidence of a difference between treatments. The p-value of 0.046, being slightly less than 0.05, is consistent with the test statistic being slightly to the right of the 0.05 critical value.

We have established that communication does have an effect on the seller's decision to invest. However, we still need to establish whether one type of communication is more effective than the other.

*Research Question 2: Do the two different types of communication have differing effects on the seller's decision of whether to invest?*

This requires a chi-squared test again, but using only Treatments 2 and 3, i.e. using only two of the three columns of the contingency table:

```
. tab invest treatment if treatment!=1, chi2 col

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

    sellers |
 investment |
  decision: |
       1 if |
   invest; 0 |      treatment
     if not |         2          3 |     Total
-----------+----------------------+----------
         0 |        14         12 |        26
           |     46.67      36.36 |     41.27
-----------+----------------------+----------
         1 |        16         21 |        37
           |     53.33      63.64 |     58.73
-----------+----------------------+----------
     Total |        30         33 |        63
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =   0.6882   Pr = 0.407
```

Note that the STATA command contains `if treatment!=1` which means "if treatment is not equal to 1", which results in the test being applied to a comparison of treatments 2 and 3.

While the percentage investing in T3 (63.64%) is somewhat higher than the percentage in T2 (53.33%), the chi-squared test reveals that there is no evidence of a difference between these. We therefore have no evidence that the two forms of communication differ in effectiveness with respect to the seller's decision.

## 3.4 Testing for Normality

In this section, we will continue to use the holdup data as an example, but we shall turn to the buyer's decision, given that the seller has invested. Unlike the seller's decision which is dichotomous, the buyer's decision is represented by an amount of money – the amount which he or she offers to return to the seller. The particular hypothesis in which we are interested in this section is normality of the data. Knowledge of whether the data is distributed normally is very useful in deciding between the different tests introduced in subsequent sections.

Since 51 of the sellers invested, we have 51 observations on the buyer's decision. The frequency distribution of these 51 offers is shown in Figure 3.2. The STATA syntax for obtaining this histogram is:

```
hist offer, disc freq normal xline(60) xlabel(0(10)100)
```

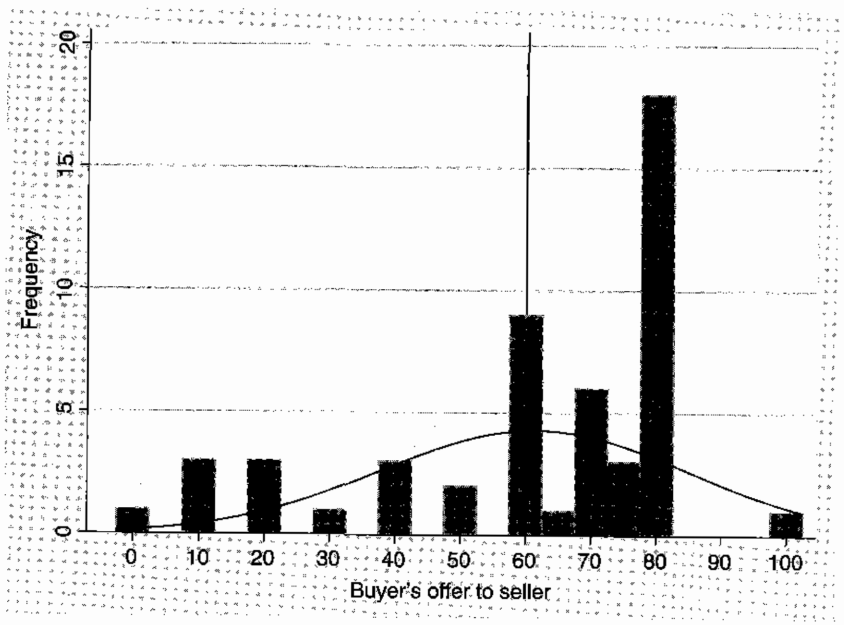



Figure 3.2: Frequency histogram of buyers' offers



This command has a number of options. The `disc` option ensures that there is a different bar for each discrete value of the data; `freq` causes frequency to be measured on the vertical axis, rather than density; `normal` causes a normal density to be superimposed on the histogram, with the same mean and standard deviation as the data; `xline(60)` causes a vertical line to appear at the offer of 60 (i.e. returning exactly the amount invested); `xlabel(0(10)100)` results in the x-axis covering the range 0 to 100 with tick-marks positioned every 10 units.

The modal buyer's offer is 80, indicating the tendency of the buyer to return the investment of 60 to the seller, in addition to splitting the profit of 40 evenly. However, some buyers offer amounts considerably below this, resulting in net losses for the seller.

The distribution does not appear to correspond closely to the normal curve. This is preliminary evidence that the data does not follow a normal distribution. A formal test of normality is the "skewness-kurtosis" test, obtained using the `sktest` command in STATA. Applying this test to the complete set of offer data, we obtain:

```
. sktest offer

                    Skewness/Kurtosis tests for Normality
                                                       ------- joint ------
    Variable |    Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+-----------------------------------------------------------------
       offer |     51      0.0021         0.4595          8.60         0.0136
```

The output actually contains three different test results, in the form of p-values. Pr(Skewness) is the p-value for the test of the hypothesis that skewness[1] equals zero (i.e. that the distribution is symmetric). The p-value of 0.0021 implies that symmetry is strongly rejected by the data. Pr(Kurtosis) is the p-value for the hypothesis of "normal kurtosis".[2] The third p-value represents the result of a joint test of skewness and kurtosis.

It is also useful to test for normality separately by treatment. We see that normality is rejected in Treatments 1 and 2, strongly in the latter.

```
. sktest offer if treatment==1

                    Skewness/Kurtosis tests for Normality
                                                       ------- joint ------
    Variable |    Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+-----------------------------------------------------------------
       offer |     14      0.5120         0.0088          6.52         0.0383

. sktest offer if treatment==2

                    Skewness/Kurtosis tests for Normality
                                                       ------- joint ------
    Variable |    Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+-----------------------------------------------------------------
```

[1] *Skewness* is measured by the third central moment of the distribution. Skewness is zero for a symmetric distribution. If skewness is positive, it is said that the distribution is "positively skewed" or "right-skewed", and the distribution is characterised by a long right-tail. Negative skewness (or left-skewness) is characterised by a long left-tail.

[2] *Kurtosis* is a measure of the fourth central moment of a distribution. For a standardised normal distribution, kurtosis is 3. If kurtosis is larger than 3, the distribution is said to be *leptokurtic* (fat-tailed); if less than 3, *platykurtic*.

```
       offer |       16       0.0003          0.0012          16.65          0.0002

. sktest offer if treatment==3

                   Skewness/Kurtosis tests for Normality
                                                         ------- joint ------
    Variable |     Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)    Prob>chi2
-------------+----------------------------------------------------------------
       offer |      21       0.2886          0.6833          1.42        0.4918
```

Another test for normality is the Shapiro-Wilk test, conducted using the swilk command in STATA. When this test is applied to the complete sample, we reach the same conclusion as before: normality is strongly rejected.

```
. swilk offer

                   Shapiro-Wilk W test for normal data

    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
       offer |     51    0.87616      5.916     3.795    0.00007
```

## 3.5 Treatment Testing

### *3.5.1 Parametric tests of treatment effects*

In this section, we apply treatment tests to the buyer's decision in the holdup data. Because this decision is represented by a continuous variable (the amount which the buyer offers to return to the seller), a different set of tests are required from those introduced in the context of discrete outcomes in Section 3.3. Also, in Section 3.4, normality tests were applied to the buyer's offer, and evidence was found that the variable did not follow a normal distribution. This is important in choosing between the tests introduced in this section.

Since 51 of the sellers invested, we have 51 observations on the buyer's decision. The distribution of these 51 offers is shown in Figure 3.2. As noted in the previous section, the modal buyer's offer is 80, indicating the tendency of the buyer to return the investment of 60 to the seller, in addition to splitting the profit of 40 evenly. However, some buyers offer amounts considerably below this, resulting in net losses for the seller.

Again we are interested in any differences between treatments, and it is natural to compare "average" offers between the three treatments. The following table shows the mean offer by treatment:

```
. table treatment, contents(n offer mean offer)

----------------------------------
treatment |   N(offer)  mean(offer)
----------+-----------------------
        1 |         14     48.5714
        2 |         16          70
        3 |         21     63.3333
----------------------------------
```



Firstly, note that the cases in which no investment was made have been excluded from this table. This is because, when no investment is made, the offer is coded as a *missing value* (. in STATA). Please recognise the importance of this. A common mistake is to code missing observations as zeros – this would be very misleading here, since it would impose a severe downward bias on the mean offer.

Secondly, note that there appear to be differences between treatments, with communication tending to increase offers, this time with promises (T2) being more effective than threats (T3).

Again we need to consider whether these differences are statistically significant. There are three comparisons to be made:

*Research Question 3: What is the effect of seller communication (threats) on the buyers offer? (T3 vs T1)*

*Research Question 4: What is the effect of buyer communication (promises) on the buyers offer? (T2 vs T1)*

*Research Question 5: Do the two forms of communication differ in their impact on buyers offers? (T3 vs T2)*

There are a number of possible ways to address these questions.

#### *3.5.1.1 The independent samples t-test (or two-sample t-test)*

We are comparing two samples. Let us assume that the first sample comes from a population with mean $\mu_1$ and standard deviation $\sigma_1$, while the second sample comes from a population with mean $\mu_2$ and standard deviation $\sigma_2$.

We wish to use the information in the two samples to test the null hypothesis $H_0$: $\mu_1 = \mu_2$ against the alternative $H_1$: $\mu_1 \neq \mu_2$. The information in the two samples is in the form of sample sizes $n_1$ and $n_2$, sample means $\bar{x}_1$ and $\bar{x}_2$, and sample standard deviations $s_1$ and $s_2$.

The independent samples t-test statistic is based straightforwardly on the difference between the two sample means:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{3.3}$$

where $s_p$ (the pooled standard deviation) is just a weighted average of the two individual sample standard deviations:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{3.4}$$

The pooled standard deviation ($s_p$) is used when it is assumed that the two populations have the same variance, i.e. that $\sigma_1 = \sigma_2$. If there are reasons for not assuming this, there is also an "unequal variances" version of the test.

Given that certain assumptions are met, the t-statistic presented in (3.3) has a $t(n_1 + n_2 - 2)$ distribution under the null hypothesis.

Let us apply the test to Research Question 3 (seller communication):

```
. ttest offer if treatment!=2, by(treatment)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |      14    48.57143    8.619371    32.25073    29.95041    67.19245
       3 |      21    63.33333    4.230464    19.38642    54.50874    72.15793
---------+--------------------------------------------------------------------
combined |      35    57.42857    4.383753    25.93463    48.51971    66.33743
---------+--------------------------------------------------------------------
    diff |           -14.7619    8.711774               -32.48614    2.962333
------------------------------------------------------------------------------
    diff = mean(1) - mean(3)                                      t =  -1.6945
Ho: diff = 0                                     degrees of freedom =       33

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0498         Pr(|T| > |t|) = 0.0996          Pr(T > t) = 0.9502
```

The test statistic is $-1.6945$, which is compared with the t(33) distribution. This is done for us. Various p-values are shown in the final row of the results. The "two-tailed p-value" is seen to be 0.0996. This represents mild evidence of a difference between the two treatments (1 and 3).

In this situation, we have a prior belief about the direction of the effect: we expect seller communication to have a positive effect on buyer offer. For this reason we only reject the null if the statistic is in the lower tail of the distribution, and we divide the p-value by 2, giving 0.0498. This is the p-value shown on the left. Note that having a prior belief allows us to interpret the evidence as being stronger, and we are able to upgrade the evidence from "mild evidence" to simply "evidence".

As noted above, there is a version of the test that can be used if the two population variances are not assumed to be equal. This just requires the `unequal` option:

```
. ttest offer if treatment!=2, by(treatment) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |      14    48.57143    8.619371    32.25073    29.95041    67.19245
       3 |      21    63.33333    4.230464    19.38642    54.50874    72.15793
---------+--------------------------------------------------------------------
combined |      35    57.42857    4.383753    25.93463    48.51971    66.33743
---------+--------------------------------------------------------------------
    diff |           -14.7619    9.601583               -34.83782    5.314009
------------------------------------------------------------------------------
    diff = mean(1) - mean(3)                                      t =  -1.5374
Ho: diff = 0                     Satterthwaite's degrees of freedom =    19.29

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0702         Pr(|T| > |t|) = 0.1404          Pr(T > t) = 0.9298
```

Note that the evidence is downgraded to "mild" as a consequence of not assuming equal variances. This is a common experience in hypothesis testing; the less that we can assume before conducting a test, the weaker the evidence is likely to be.



If you really wish to know whether you can assume equal variances, conduct a variance-ratio test, as follows:

```
. sdtest offer if treatment!=2, by(treatment)

Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |      14    48.57143    8.619371    32.25073    29.95041    67.19245
       3 |      21    63.33333    4.230464    19.38642    54.50874    72.15793
---------+--------------------------------------------------------------------
combined |      35    57.42857    4.383753    25.93463    48.51971    66.33743
------------------------------------------------------------------------------
    ratio = sd(1) / sd(3)                                         f =   2.7675
Ho: ratio = 1                                    degrees of freedom =   13, 20

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.9801         2*Pr(F > f) = 0.0398           Pr(F > f) = 0.0199
```

This unfortunately tells us that there is evidence of a difference in variances between treatments 1 and 3.

The two-sample t-test is applied to the other two research questions. Results from all three (ignoring the problem of unequal variances) are shown below:

| | | Two-tailed p-value (equal variances) |
|---|---|---|
| Q3 | No comm. (T1) vs seller comm. (T3) | 0.0996 |
| Q4 | No comm. (T1) vs buyer comm. (T2) | 0.0250 |
| Q5 | Seller comm. (T3) vs buyer comm. (T2) | 0.2673 |

Two-tailed p-values from two-sample t-tests for treatment effects on buyer's offer.

As remarked above, the two-sample t-test relies on quite strong assumptions about the data. Most importantly, unless the two samples happen to be "large", it is required that the two populations are normally distributed. In Section 3.4 we found strong evidence that the buyer's offer is not normally distributed. This result, together with the fact that the numbers of observations in each treatment are considerably lower than the 30 required for the central limit theorem to apply, leads us to doubt the validity of the tests carried out in this sub-section.

### 3.5.2 *Non-parametric tests of treatment effects: the Mann-Whitney test*

A test for comparing two samples which does not rely on any distributional assumptions (such as normality of the data) is the Mann-Whitney U test. Because no such assumptions are made, it is classed as a non-parametric test.

To carry out the test, all of the observations from both samples are ranked by their value, with the highest rank being assigned to the largest value, and with ranks averaged in the event of a tie. Then the sum of ranks are found for each sample, and compared. The test is based on this comparison.

The test is carried out in STATA using the `ranksum` command. The following compares T1 and T3, and is therefore a test of Research Question 3.

```
ranksum offer if treatment!=2, by(treatment)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

    treatment |      obs    rank sum    expected
-------------+---------------------------------
            1 |       14       220.5         252
            3 |       21       409.5         378
-------------+---------------------------------
     combined |       35         630         630

unadjusted variance      882.00
adjustment for ties      -26.56
                      ----------
adjusted variance        855.44

Ho: offer(treatm~t==1) = offer(treatm~t==3)
             z =  -1.077
    Prob > |z| =   0.2815
```

The p-value of 0.2815 indicates that there is no evidence of a difference in offers between T1 and T3.

The Mann-Whitney test may be applied to all of Research Questions 3–5, and the results, in terms of p-values, are presented (and compared with the corresponding results from t-tests) in the following table:

| | | Two-sample t-test | Mann-Whitney test |
|---|---|---|---|
| Q3 | T1 vs T3 | 0.0996 | 0.2815 |
| Q4 | T1 vs T2 | 0.0250 | 0.0886 |
| Q4 | T3 vs T2 | 0.2673 | 0.1765 |

Two-tailed p-values from two-sample t-tests and Mann-Whitney tests

Using the Mann-Whitney test, we do not find any significant differences, with the exception of Q4, for which we find only mild evidence of a difference. Intuitively, we expect non-parametric tests such as this one to indicate weaker evidence of an effect than the corresponding parametric test, simply because fewer assumptions are being made. Comparisons of p-values seen in this table with those in the previous table are broadly consistent with this expected pattern.

### 3.5.3 The bootstrap

We have introduced the Mann-Whitney test as a non-parametric analogue to the two-sample t-test, indicating that the former may be preferred in situations in which one doubts the assumption of normality of the data. However, one drawback of non-parametric tests of this type is that they are based solely on the *ordinality* of the data, and hence they completely disregard the (possibly) rich *cardinal* information in the data.

The "bootstrap" technique (Efron & Tibshirani, 1993) provides a means of conducting a parametric test such as the two-sample t-test (which definitely respects cardinality), without making any assumptions about the distribution of the data. The technique was applied by Ellingsen & Johannesson (2004) to their holdup data. We will attempt to reproduce their results below.

The bootstrap procedure consists of the following five steps:

1. Apply the parametric test on the data set, obtaining a test statistic, $\hat{t}$.
2. Generate a healthy number, B, of "bootstrap samples". These are samples of the same size as the original sample. They are also drawn from the original sample, but the key point is that they are drawn *with replacement*. For each bootstrap sample, compute the test statistic, $\hat{t}^*_j$, $j = 1, \ldots, B$.
3. Compute the standard deviation $s_B$ of the bootstrap test statistics $\hat{t}^*_j$, j=1,. . ., $B$.
4. Obtain the new test-statistic $z_B = \hat{t}/s_B$.
5. Compare $z_B$ against the standard normal distribution in order to find the "bootstrap p-value".

According to MacKinnon (2002), the number of bootstrap samples, $B$, should be chosen so that $\alpha(B+1)$ is a whole number, where $\alpha$ is the chosen test size. Since $\alpha$ is usually set to 0.01, 0.05, or 0.10, this requirement essentially means that $B$ should be either 99 or 999 or 9999, etc. This recommendation is followed here.

The following command applies the bootstrap two-sample t-test to Research Question 3 (T3 vs. T1):

```
bootstrap t=r(t), rep(999) nodrop : ///
ttest offer if treatment!=2, by(treatment)

(running ttest on estimation sample)

Bootstrap replications (999)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
..................................................   550
..................................................   600
..................................................   650
..................................................   700
..................................................   750
..................................................   800
..................................................   850
..................................................   900
..................................................   950
.................................................

Bootstrap results                               Number of obs      =       103
                                                Replications       =       999

      command:  ttest offer if treatment!=2, by(treatment)
            t:  r(t)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -1.694477    1.144873    -1.48   0.139    -3.938386    .5494314
------------------------------------------------------------------------------
```

Since the `bootstrap` command is too long to fit onto one line, we have spread it over two lines with the continuation marker "///". When the command is run, a sequence of dots appears on the screen, one for each bootstrap sample, so that the user is aware of the progress of the procedure.

Another useful thing to do when learning this technique is to add the `saving` option to the `bootstrap` command. In the next command, we use this option, along with the higher number of bootstrap samples, 9,999. Since we do not want 9,999 dots to appear on the screen, we also use the option `nodots`.

```
bootstrap t=r(t), nodots rep(9999) nodrop saving("hello.dta", replace) : ///
ttest offer if treatment!=2, by(treatment)

Bootstrap results                          Number of obs     =        103
                                           Replications      =       9999

      command:  ttest offer if treatment!=2, by(treatment)
            t:  r(t)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |  -1.694477     1.1553    -1.47   0.142    -3.958824    .5698692
------------------------------------------------------------------------------
```

When the `saving` option is used, the bootstrap test statistics $\hat{t}_j^*$, j=1,. . .,9999 are stored in the new data set "hello". If we read the contents of this file, we can then investigate the distribution of the 9,999 bootstrap test statistics:

```
. summ t

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           t |      9999   -1.718972      1.1553  -8.077973   2.979723
```

Note in particular that the standard deviation of these 9,999 numbers is the "bootstrap standard error" that appears in the results table from the bootstrap command shown above. The distribution of bootstrap test statistics is also shown as a histogram in Figure 3.3. We see that the distribution is bell-shaped and symmetric, with the centre very close to the "actual" t-statistic for the test, which was $-1.6945$.

We can now add the bootstrap column to the table of test results for Research Questions 3–5. This is done in the following table. These results are similar to those appearing in Table 2 of Ellingsen & Johannesson (2004), although they are not identical and this is an expected consequence of the randomness implicit in the bootstrap procedure. We agree with Ellingsen & Johannesson (2004) that the comparison of T1 and T2 (buyer communication) is the only one that shows a significant effect.

| | | Two-sample t-test | Mann-Whitney test | Bootstrap |
|---|---|---|---|---|
| Q3 | T1 vs T3 | 0.0996 | 0.2815 | 0.140 |
| Q4 | T1 vs T2 | 0.0250 | 0.0886 | 0.045 |
| Q5 | T3 vs T2 | 0.2673 | 0.1765 | 0.289 |

Two-tailed p-values from: two-sample t-tests; Mann-Whitney tests; bootstrap tests



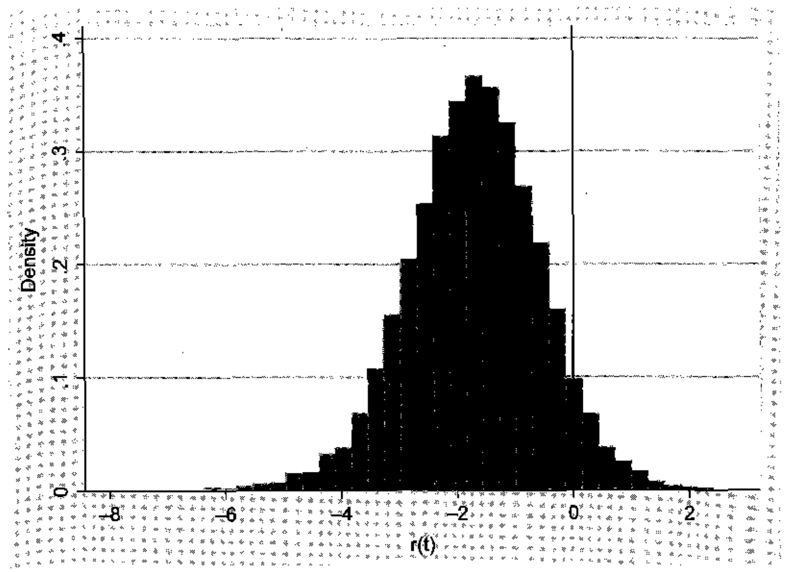



Figure 3.3: The distribution of the 9,999 bootstrap test statistics

### 3.5.4 Tests comparing entire distributions

In this section, we consider data from the ultimatum game and the dictator game. The reader is referred back to Section 2.5 for an explanation of these games, a discussion of their theoretical predictions, and a summary of empirical evidence relating to them.

Forsythe et al. (1994) set out to test whether fairness alone can explain proposers' willingness to give to responders in these settings. They do this by subjecting one group to an ultimatum game, and the other group to a dictator game. The basic idea is that, if giving is the same in both, then fairness is the only explanation for giving. If giving is greater in the ultimatum game, then other factors must be influencing the decision (e.g. fear of the offer being rejected).

The tests that they favour are based on a comparison of the entire distributions of proposals under the two treatments, rather than a comparison of a particular characteristic of the distribution such as mean or variance. This is because (Forsythe et al., 1994, p. 351):

> conventional theory predicts that proposals will be concentrated at a single point. . . Since theory does not predict a distribution of proposals, it provides no guidance about which functionals of the distribution should be tested. Invariance of the entire distribution has the appealing property of implying that all functionals are invariant.

Tests that make comparisons of entire distributions include: the Kolmogorov-Smirnov test; the Epps-Singleton test; the Cramer-von Mises test; the Anderson-Darling test. We will demonstrate the first two of these on the data of Forsythe et al. (1994). The data is contained in the file **forsythe**.

The Kolmogorov-Smirnov test is implemented by the command `ksmirnov` in STATA. Results are as follows:

```
. ksmirnov y, by(dic_ult)

Two-sample Kolmogorov-Smirnov test for equality of distribution functions

 Smaller group        D        P-value  Corrected
 ------------------------------------------------
 1:                 0.3516    0.000
 2:                -0.0110    0.989
 Combined K-S:      0.3516    0.000       0.000

Note: ties exist in combined dataset;
      there are 11 unique values out of 182 observations.
```

In order to understand how the Kolmogorov-Smirnoff test statistic is computed, a very useful graph is a "cdfplot" (this is a user-written STATA command that needs to be installed; start by typing `findit cdfplot`). Here, we use the command as follows:

```
cdfplot y, by(dic_ult)
```

and the result is shown in Figure 3.4. The higher of the two lines in the plot is the cumulative distribution function (cdf) of giving in the dictator game; the lower is the same for the ultimatum game. That the dictator game cdf is higher is consistent with dictator game offers typically being lower than ultimatum game offers.

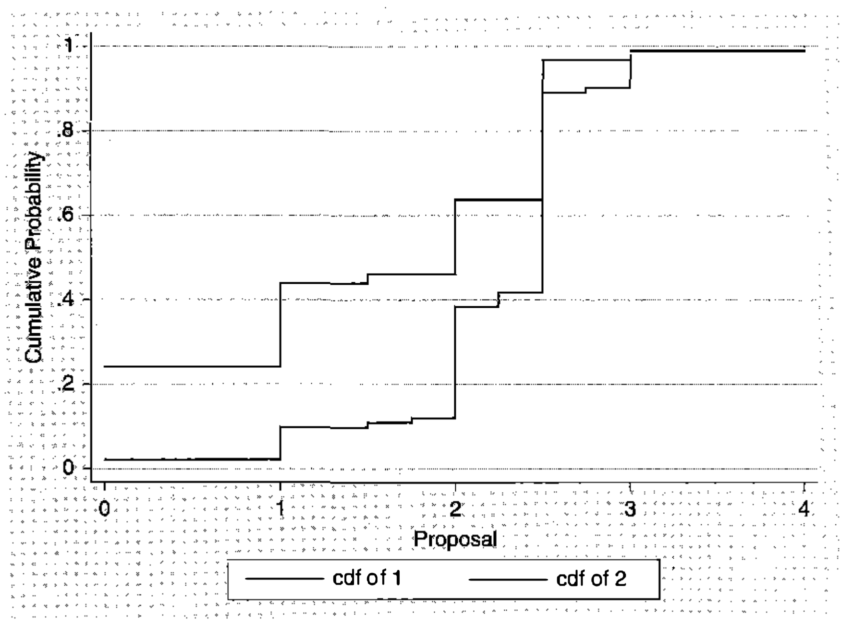



Figure 3.4: cdfs of giving in the ultimatum game (lower line) and dictator game (higher line)



The Kolmogorov-Smirnov test statistic is computed as the largest vertical distance between the two cdfs, and we see from the STATA output, and also from the plot, that this maximal distance (occurring for proposals between 1.5 and 1.75) is 0.3516. This difference is then compared to a null distribution in order to obtain the p-value for the test, which is 0.000. This p-value indicates that there is overwhelming evidence of a difference between the two distributions.

The Epps-Singleton test (Epps & Singleton, 1986) does not compare the two distributions directly, but instead compares the empirical characteristic functions. This test is believed to perform similarly to the Kolmogorov-Smirnov test in terms of power, and has the added advantage of being applicable when the outcome has a discrete distribution (e.g. if the outcome is the number of questions answered correctly in a quiz). The test is implemented in STATA using the user-written command `escftest` (Georg, 2009). This is another command that needs to be installed; start by typing `findit ecftest`.

Here is the output from applying the Epps-Singleton test to the comparison of dictator and ultimatum giving.

```
. escftest y, group(dic_ult)

Epps-Singleton Two-Sample Empirical Characteristic Function test

Sample sizes: dic_ult = 1            91
              dic_ult = 2            91
              total                 182
t1                                0.400
t2                                0.800

Critical value for W2 at 10%      7.779
                          5%      9.488
                          1%     13.277
Test statistic W2                35.624

Ho: distributions are identical
P-value                         0.00000
```

We have applied both the Kolmogorov-Smirnov test and the Epps-Singleton test to the problem of comparing the distributions of giving in the ultimatum and dicatator games. Both tests result in a p-value of 0.000, amounting to overwhelming evidence of a difference between the two distributions. The interpretation of this result is that fairness is not the only consideration that enters the decision of how much to give.

## 3.6 Testing for Gender Effects

In this section, we show how treatment tests can be used to test for a gender effect. This is done by simply treating gender as the "treatment", although it is clearly not a typical treatment since it is not assigned by the experimenter. This will be done in the context of the ultimatum game, which was explained in Section 2.5. As in the last section, we are interested in the proposer's decision.

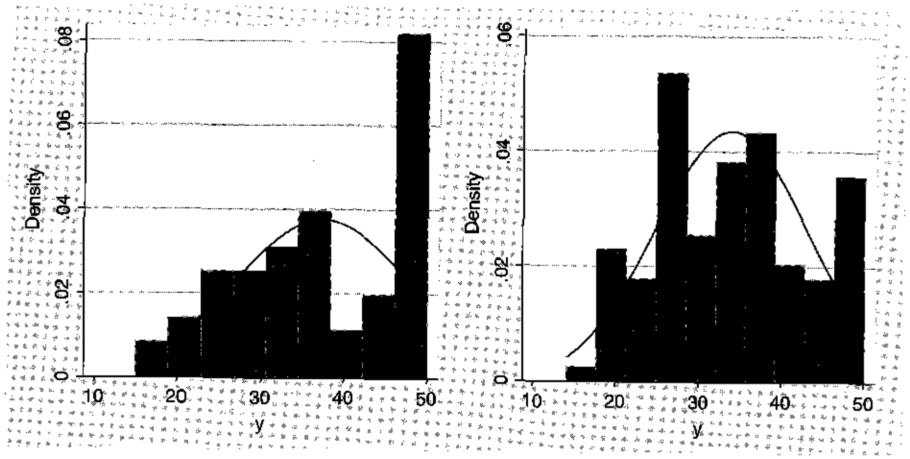



Figure 3.5: Proposers offers: females (left); male(right)

The file **ug_sim** contains (simulated) data from 200 subjects who participated in an ultimatum game, in which the size of the pie is 100 units. Each subject plays twice, once as proposer, and once as responder, with a different opponent each time. The variables are:

| | |
|---|---|
| i: | Proposer ID |
| j: | Responder ID |
| male_i: | 1 if proposer is male; 0 otherwise |
| male_j: | 1 if responder is male; 0 otherwise |
| y: | Proposer's offer |
| d: | Responder's decision: 1 if accept; 0 if reject |

A large amount of research has been done on gender effects in the ultimatum game. A good paper to start with is Eckel & Grossman (2001). Here, we will look for gender differences in proposers' offers. The distributions of proposers' offers are shown separately by gender in Figure 3.5. As expected, offers are distributed between zero and 50 (one half of the pie), and there is an accumulation of offers at 50 for both genders. Normal densities are superimposed. Neither distribution appears close to the normal.

Next, we test formally for normality in the distribution of proposer offers, for each gender separately, using one of the tests for normality introduced in Section 3.4.

```
. sktest y if male_i==0

                   Skewness/Kurtosis tests for Normality
                                                    ------- joint ------
    Variable |  Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)   Prob>chi2
-------------+-------------------------------------------------------------
           y |   91      0.3391         0.0000        21.61        0.0000
```



```
. sktest y if male_i==1

                   Skewness/Kurtosis tests for Normality
                                                    ------- joint ------
    Variable |  Obs   Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)   Prob>chi2
-------------+-------------------------------------------------------------
           y |   109     0.3899         0.0135         6.42        0.0403
```

Normality is rejected for both genders. For females, rejection is particularly strong, and this is a consequence of females having a larger accumulation (than males) of observations at the equitable allocation of 50 (see Figure 3.5).

The next thing we might wish to do is to test for equal variances between the two samples. The results of this test are as follows:

```
. sdtest y, by(male_i)

Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      91    37.37363    1.115618    10.64231    35.15726       39.59
       1 |     109    33.86239    .8779076    9.165624    32.12222    35.60255
---------+--------------------------------------------------------------------
combined |     200       35.46    .7067101     9.99439     34.0664     36.8536
------------------------------------------------------------------------------
    ratio = sd(0) / sd(1)                                         f =   1.3482
Ho: ratio = 1                                  degrees of freedom =  90, 108

    Ha: ratio < 1               Ha: ratio != 1                Ha: ratio > 1
  Pr(F < f) = 0.9314         2*Pr(F > f) = 0.1372           Pr(F > f) = 0.0686
```

This test shows mild evidence that the two variances are different. We might wish to allow for these difference in variance in the next test, the independent samples t-test for a gender difference in proposer offers:

```
. ttest y, by(male_i)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      91    37.37363    1.115618    10.64231    35.15726       39.59
       1 |     109    33.86239    .8779076    9.165624    32.12222    35.60255
---------+--------------------------------------------------------------------
combined |     200       35.46    .7067101     9.99439     34.0664     36.8536
---------+--------------------------------------------------------------------
    diff |            3.511241    1.400706                .7490252    6.273457
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   2.5068
Ho: diff = 0                                     degrees of freedom =      198

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9935         Pr(|T| > |t|) = 0.0130          Pr(T > t) = 0.0065

. ttest y, by(male_i) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      91    37.37363    1.115618    10.64231    35.15726       39.59
       1 |     109    33.86239    .8779076    9.165624    32.12222    35.60255
```

```
---------+--------------------------------------------------------------------
combined |      200       35.46    .7067101     9.99439     34.0664     36.8536
---------+--------------------------------------------------------------------
    diff |             3.511241    1.419621                .7098767    6.312605
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   2.4734
Ho: diff = 0                     Satterthwaite's degrees of freedom =  178.831

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9928         Pr(|T| > |t|) = 0.0143          Pr(T > t) = 0.0072
```

Whether or not we assume equal variances, there is strong evidence of a gender difference, with females offering more than males on average. The evidence is slightly stronger (indicated by the slightly smaller p-value) when equal variances are assumed.

Despite the rejection of normality, this is a situation in which the result of the independent samples t-test may be relied upon, by virtue of the large sample sizes. However, for good measure, we shall also conduct a non-parametric test for a gender effect.

```
. ranksum y, by(male_i)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      male_i |      obs    rank sum    expected
-------------+---------------------------------
           0 |       91     10122.5      9145.5
           1 |      109      9977.5     10954.5
-------------+---------------------------------
    combined |      200       20100       20100

unadjusted variance   166143.25
adjustment for ties    -1147.79
                     ----------
adjusted variance     164995.46

Ho: y(male_i==0) = y(male_i==1)
             z =   2.405
    Prob > |z| =   0.0162
```

Again we see evidence of a gender difference, although the effect is not strong this time, since the p-value is greater than 0.01. A general conclusion from this sequence of tests is that the more that can be assumed about the data (i.e. the more "parametric" the test) the stronger the test result tends to be (in terms of the closeness to zero of the p-value).

We could, of course, apply other tests to this problem, such as the tests of equality of entire distributions introduced in Section 3.5.4. We would expect to reach a similar conclusion. These tests are left to the reader.

## 3.7 Within-subject Tests

Within-subject tests are used to test the effect of a treatment in a situation in which each subject is observed both before and after the treatment. This is in contrast to the between-subject tests that have been considered up until now in this chapter, in which one group of subjects are exposed to a treatment, while a different group is not exposed. From a theoretical point of view, within-subject tests are preferred to between-subject tests; they have more statistical power. However, there are various reasons why within-subject tests are not favoured by experimental economists. The issue of "order effects" is much discussed (see for example Harrison et al., 2005; Holt & Laury, 2002); an order effect is present if the result of the test depends on the order in which the control and the treatment are administered. More generally, there are concerns that the experience of one treatment impacts on behaviour in the treatment that follows.

There are however some instances in experimental economics in which within-subject tests are the most natural approach.

The test that is most appropriate to many within-subject settings arising in experimental economics is the McNemar change test (see Siegel & Castellan, 1988). This is because the two decisions are usually binary, and we are simply interested in the subjects who "switch" from one choice to the other, and in the direction in which they are switching. The Conlisk (1989) test is an alternative test that is applicable in this setting. If the outcome observed on the two occasions has a continuous distribution, it is appropriate to use the paired-comparison t-test, or, if a non-parametric test is preferred, the Wilcoxon signed ranks test.

### 3.7.1 The Allais paradox

The Allais paradox (Allais, 1953) is perhaps the most well-known contradiction of expected utility (EU) theory. It is normally tested using within-subject tests of the type introduced above.

The paradox is demonstrated by addressing a sequence of two (usually hypothetical) questions to a sample of subjects. The first question asks which they would prefer out of the lotteries A and A* below. The second question asks them to choose between B and B*.

| | |
|---|---|
| Lottery A: | Certainty of $1 million |
| Lottery A*: | 0.01 chance of nothing |
| | 0.89 chance of $1 million |
| | 0.10 chance of $5 million |
| Lottery B: | 0.89 chance of nothing |
| | 0.11 chance of $1 million |
| Lottery B*: | 0.90 chance of nothing |
| | 0.10 chance of $5 million |

If a subject chooses A in the first question, and B in the second, we shall label their sequence of answers as "AB". There are clearly four different ways in which a subject can answer the two questions: AB, A*B*, AB*, A*B. Of these four possibilities, AB and A*B* are consistent with EU; AB* and A*B both indicate a violation of EU.

In practice, a significant number of subjects do violate EU by choosing either AB* or A*B. However, what is of particular interest is the pattern, known as "Allais behaviour", of AB* violations being much more frequent than A*B violations.

In order to develop tests for the presence of Allais behaviour, we will use the notation $n(.)$ to represent the number of subjects who answer with a particular sequence, for example, $n(AB^*)$ is the number of subjects who answer AB*.

The McNemar change test is conducted as follows. The null hypothesis is that AB* and A*B are equally likely. That is, we expect $n(AB^*)$ and $n(A^*B)$ to be approximately equal. To test this null, we apply the chi-squared test introduced in Section 3.3.3 to these two groups. The test statistic is:

$$\chi^2 = \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i} = \frac{\left[n(AB^*) - \frac{n(AB^*)+n(A^*B)}{2}\right]^2}{\frac{n(AB^*)+n(A^*B)}{2}} + \frac{\left[n(A^*B) - \frac{n(AB^*)+n(A^*B)}{2}\right]^2}{\frac{n(AB^*)+n(A^*B)}{2}} \tag{3.5}$$

Expanding and simplifying (3.5), we obtain:

$$\chi^2 = \frac{\left[n(AB^*) - n(A^*B)\right]^2}{n(AB^*) + n(A^*B)} \tag{3.6}$$

The distribution of (3.6) is $\chi^2(1)$ under the null hypothesis of no Allais behaviour.[3]

Conlisk (1989) presented the two choice problems to 236 subjects. The numbers providing each response combination are given in the following table:

| | B | $B^*$ |
|---|---|---|
| A | 18 | 103 |
| $A^*$ | 16 | 99 |

Results from Conlisk's (1989) Allais experiment. Source: Conlisk (1989) Table 1.

It is the off-diagonal entries in the table on which we focus, and we see that the number of subjects answering $AB^*$ is considerably greater than the number answering $A^*B$. To test the difference statistically, we apply the McNemar test (3.6):

$$\chi^2 = \frac{[103 - 16]^2}{103 + 16} = 63.6$$

[3] In Section 3.3.3 it was explained that the null-distribution of the chi-squared test based on a $m \times n$ cross-tabulation is $\chi^2$ with $(m-1)(n-1)$ degrees of freedom. An exception arises when either $m = 1$ or $n = 1$, that is, when the cross-tabulation consists of only one row or one column. If $m = 1, df = n - 1$ and if $n = 1, df = m - 1$. In the present context, the cross-tabulation is essentially $2 \times 1$, and hence the degrees of freedom is one.



Since the null distribution is $\chi^2(1)$, any value of this test statistic greater than 3.84 would constitute evidence of Allais behaviour, and any value greater than 6.63 constitutes strong evidence. The statistic of 63.6 therefore represents strong evidence of Allais behaviour.

One further point about the McNemar test is that when the numbers appearing in the formula are small, the approximation by the $\chi^2(1)$ distribution may become poor, since a continuous distribution is being used to approximate a discrete distribution. To deal with this problem, a "continuity correction" may be applied (see Yates, 1934). The formula for the test statistic including the continuity correction is given by:

$$\chi^2 = \frac{\left[|n(AB^*) - n(A^*B)| - 1\right]^2}{n(AB^*) + n(A^*B)} \tag{3.7}$$

In the present case, the test statistic changes from 63.6 to 62.15 when the continuity correction (3.7) is applied.

Conlisk (1989) suggested an alternative test statistic for detecting Allais behaviour, and this has come to be known as the "Conlisk test". The statistic is given by the following formula:

$$Z = \frac{\sqrt{N-1}\left(S - \frac{1}{2}\right)}{\sqrt{\frac{1}{4V} - \left(S - \frac{1}{2}\right)^2}} \tag{3.8}$$

where $N$ is the total number of subjects, $V$ is the proportion of subjects who violate EV by giving $AB^*$ or $A^*B$ answers, that is:

$$V = \frac{n(AB^*) + n(A^*B)}{N} \tag{3.9}$$

and S is the proportion of violators who answer $AB^*$ rather than $A^*B$, that is:

$$S = \frac{n(AB^*)}{n(AB^*) + n(A^*B)} \tag{3.10}$$

The test statistic (3.8) has a standard normal distribution under the null hypothesis of no Allais behaviour. A value in the upper tail of the standard normal distribution provides evidence that the proportion $S$ is significantly greater than one half, that is, evidence of Allais behaviour.

Applying the Conlisk test to the data in the table, we obtain $V = 0.504$, $S = 0.866$, and

$$Z = \frac{\sqrt{236-1}\left(0.866 - \frac{1}{2}\right)}{\sqrt{\frac{1}{4\times 0.504} - \left(0.866 - \frac{1}{2}\right)^2}} = \underline{9.32}$$

This test statistic is certainly in the upper tail of the standard normal distribution, again providing evidence of Allais behaviour in this sample.

### 3.7.2 Preference reversals

"Preference reversal" (PR) is a term normally used to refer to the phenomenon of subjects choosing the safer of two lotteries (the "p-bet") when asked to choose between them, but to contradict this choice by placing a higher valuation on the riskier lottery (the "$-bet") when asked to value them (that is, to provide their certainty equivalent) separately. The phenomenon was apparently discovered by Lichtenstein & Slovic (1971), and later introduced to the economics literature by Grether & Plott (1979).

For a particular example, let us look to the study of Tversky et al. (1990). The very first pair of lotteries they consider (in their Table 1: study 1, set 1, triple 1) is:

p-bet: 0.97 chance of $4; 0.03 chance of $0
$-bet: 0.31 chance of $16; 0.69 chance of $0

The number of subjects they presented with this pair of lotteries was 179. From the information in their Table 2, we may deduce the following.

| | Value p higher | Value $ higher |
|---|---|---|
| Choose p | 43 | 106 |
| Choose $ | 4 | 26 |

Results from Tversky et al. (1990), study 1, set 1, triple 1.

If a subject chooses p and values $ more highly, they are said to be making a "standard reversal". If they choose $ and value p more highly, they are said to be making a "non-standard reversal". If the number of subjects making standard reversals is significantly greater than the number making non-standard reversals, we may conclude there is evidence of the PR phenomenon.

This is clearly a situation in which "within-subject" tests are essential. It is obvious that for a PR to be observable, it is necessary for the same subject to be "observed" twice: once making the choice, and once reporting their two valuations. The tests which are appropriate are the same as those used in Section 3.7.1 for testing Allais behaviour.

Applying the McNemar test to the numbers in the above table, we obtain:

$$\chi^2 = \frac{[106-4]^2}{106+4} = 94.58$$

and, being $\chi^2(1)$ under the null, this amounts to strong evidence of the PR phenomenon.

Applying the Conlisk test, we obtain $V = 0.614$, $S = 0.963$, and

$$Z = \frac{\sqrt{179-1}\left(0.963-\frac{1}{2}\right)}{\sqrt{\frac{1}{4\times 0.614} - \left(0.963-\frac{1}{2}\right)^2}} = \underline{\underline{14.07}}$$



Being standard normal under the null, this positive test statistic amounts to strong evidence of the PR phenomenon.

### 3.7.3 Continuous outcome

All of the within-subject tests considered so far have been in the context of binary outcomes. We now turn to the situation in which subjects are again observed both without and with a treatment, but the outcome is a continuous variable.

One such situation arises when investigating the impact of a "take treatment" in a dictator game experiment. A "take game" is a dictator game in which dictators are allowed to take money away from the recipient, that is, to "give" less than zero. Bardsley (2008) and List (2007) find that dicator game giving is lower when this opportunity to take is introduced. Both of those studies use between-subject tests. However, an obvious alternative approach is a within-subject design in which subjects play two dictator games in succession: the first a "give only" game; the second a "give or take" game. The amount given may then be compared between the two treatments in order to test the effect of the "take treatment". To our knowledge, Chlaß & Moffatt (2012) are the only authors to have adopted the within-subject approach in this particular setting.

Here, we assume the following design. Subjects each play two dictator games with a pie of size 10 units. In the first game, they are asked how much, if any, of the pie they would like to give to the recipient. In the second game, they are again asked how much they would like to give to the recipient, but their opportunity set is extended such that they are allowed to "give" a negative amount up to 10 units; that is, they are allowed to *take* up to 10 units from the recipient. After they have made both decisions, one of the two games is selected by a random device, and the payoffs are implemented in accordance with the decision made in the selected game.

The file **give_take_sim** contains simulated data from 50 subjects. The variables are:

i: subject id
y1: giving in give-only game
y2: giving in the give-or-take game ($y2 < 0$ if amount is taken from recipient)

It is useful to start by plotting the two giving variables against each other. We use the following scatter command:

```
scatter y2 y1, msize(1)  jitter(2) yline(0) xlabel(0(1)5) ylabel(-5(1)5)
```

The scatterplot is shown in Figure 3.6. The `jitter` option is useful in this situation because it applies small random perturbations to the position of each point in the scatter, making it possible to see in which locations there are large accumulations of points. We see that a significant number of subjects lie on the 45 degree line, implying that the amount they give is the same in both treatments. However, some subjects lie below the 45 degree line, implying that when there is an option to take,

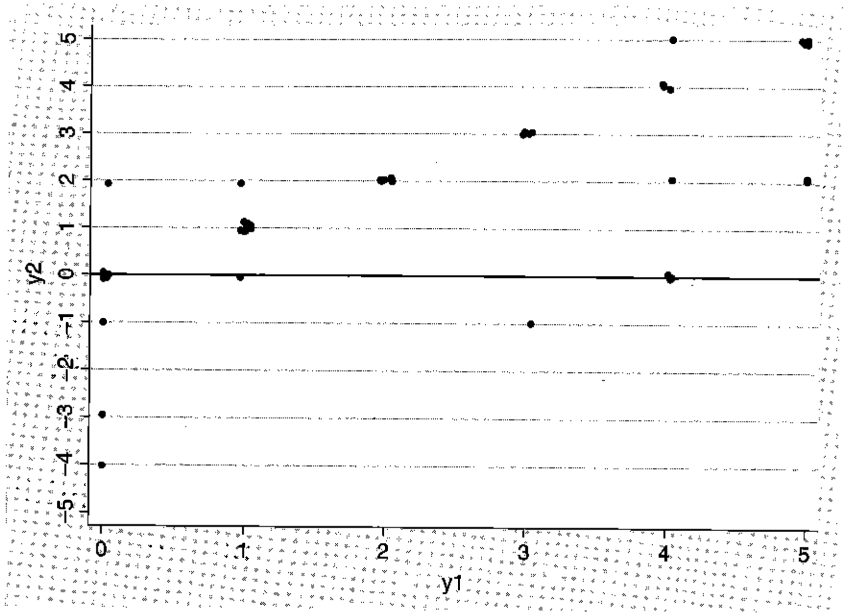



Figure 3.6: Giving in a give-or-take game against giving in a give-only game

they give less, and sometimes "give" negative amounts. Very few subjects are above the 45 degree line.

Readers may be confused that giving is being described as a "continuous" outcome while it is evident from Figure 3.6 that giving in fact takes only a small number of discrete values. The important point here is that the variable "giving" is, in theoretical terms, a continuous variable. The fact that the observed variable is discrete is simply a consequence of the manner in which the variable has been measured – namely, inducing subjects to select a whole number. This is, of course, a feature of all measurement systems: all continuous variables must be measured at some level of rounding. It is possible to deal with rounding econometrically, by estimating the interval regression model, which will be covered in Chapter 6. However, it can be verified that in situations like the current one, applying such a model yields results very similar to those obtained by treating the outcome variable as a continuous variable, as done here.

To test formally for a treatment effect, we may, as usual, choose between a parametric and a non-parametric test. The parametric test is the paired comparisons t-test. This test computes the difference in giving between treatments for each observation, and then applies the t-test to test whether these differences have mean zero. The results are:

```
. ttest y2=y1

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      y2 |      50        1.56    .2971429    2.101117    .9628691    2.157131
      y1 |      50        2.18    .2613778     1.84822    1.654742    2.705258
```



```
---------+--------------------------------------------------------------------
    diff |      50        -.62    .2038206     1.44123   -1.029593   -.2104071
------------------------------------------------------------------------------
     mean(diff) = mean(y2 - y1)                                  t =  -3.0419
 Ho: mean(diff) = 0                             degrees of freedom =       49

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0019         Pr(|T| > |t|) = 0.0038          Pr(T > t) = 0.9981
```

We see that giving in the give-or-take game is lower on average, and the average difference between treatments is −0.62. Moreover, there is strong evidence that giving is lower in the give-or-take treatment, since the one-tailed(<) p-value is 0.0019.

The non-parametric test appropriate in this situation is the Wilcoxon signed ranks test (see Siegel & Castellan, 1988). As with the parametric test, this test is based on the differences in giving between treatments for each observation. The absolute differences are ranked from lowest to highest, so that the largest difference gets the highest value. Then these ranks are summed separately for the positive differences and the negative differences. If the take-treatment has no effect, these two rank sums should be roughly equal. The test is therefore based on a comparison of these two numbers. The test is performed using the `signrank` command in STATA, as below.

```
. signrank y2=y1

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        3       116.5         340
    negative |       13       563.5         340
        zero |       34         595         595
-------------+---------------------------------
         all |       50        1275        1275

unadjusted variance    10731.25
adjustment for ties       -7.50
adjustment for zeros   -3421.25
                      ----------
adjusted variance       7302.50

Ho: y2 = y1
             z =  -2.615
    Prob > |z| =   0.0089
```

The rank sum for the negative differences is clearly a higher number, at 563.5. The test gives a (two-tailed) p-value of 0.0089 which represents strong evidence that giving is different under the take treatment. The one-tailed p-value is 0.0045. This p-value is, as expected, larger than the 0.0019 obtained above from the corresponding parametric test, indicating that the evidence from this non-parametric test is less strong.

Actually, the point must be made that the Wilcoxon signed ranks test is not completely distribution-free. It relies on the assumption that the distribution of paired differences is symmetric around the median. A test which avoids this assumption is the paired-sample sign test. This test simply compares the number of positive differences to the number of negative differences, and asks if this difference is significantly different from one half according to a binomial distribution. This test can also be peformed in STATA:

```
. signtest y2=y1

Sign test

        sign |    observed    expected
-------------+------------------------
    positive |           3           8
    negative |          13           8
        zero |          34          34
-------------+------------------------
         all |          50          50

One-sided tests:
  Ho: median of y2 - y1 = 0 vs.
  Ha: median of y2 - y1 > 0
      Pr(#positive >= 3) =
         Binomial(n = 16, x >= 3, p = 0.5) =  0.9979

  Ho: median of y2 - y1 = 0 vs.
  Ha: median of y2 - y1 < 0
      Pr(#negative >= 13) =
         Binomial(n = 16, x >= 13, p = 0.5) =  0.0106

Two-sided test:
  Ho: median of y2 - y1 = 0 vs.
  Ha: median of y2 - y1 != 0
      Pr(#positive >= 13 or #negative >= 13) =
         min(1, 2*Binomial(n = 16, x >= 13, p = 0.5)) =  0.0213
```

The relevant p-value is the second one, 0.0106. Again there is evidence that giving is lower under the take treatment. However, being the result of the most non-parametric of tests, the evidence is weaker still, and since the p-value is greater than 0.01, the evidence can no longer be classified as strong.

In the discussion of within-subject designs in Chapter 2, the problem of order effects was raised. The problem may be relevant to the current situation. It might be expected that a subject's experience of the give-only game somehow influences their behaviour in the subsequent give-or-take game, and hence the treatment effect is confounded by the order of the two treatments. As mentioned in Chapter 2, a way of addressing such a concern is to use a crossover design; that is, a design in which half of the subjects are presented with the same two treatments but in the reverse order. Any order effect could then be controlled for in the process of treatment testing. This is in fact the approach taken by Chlaß & Moffatt (2012).

## 3.8 Summary and Further Reading

This chapter has attempted to cover a wide variety of treatment tests, with examples. Readers wishing to learn more about any of these tests are directed to Siegel & Castellan (1988). Camerer (2003, ch. 2) surveys a large number of treatment tests applied to behaviour in ultimatum games, dictator games and trust games.

An important decision is the choice between non-parametric and parametric approaches, and this decision is often guided by the scale of measurement of the



data (nominal, ordinal, or cardinal). There is a large literature on this point, including Harwell & Gatti (2001).

One particular application of treatment testing that has been covered is the testing of gender effects. Croson & Gneezy (2009) provide a thorough review of the literature on the testing of gender effects in economic experiments.

The within-subject testing approach was applied to a number of situations including the preference reversal (PR) phenomenon. The PR phenomenon was first introduced to the economics literature by Grether & Plott (1979). A very useful survey of research on the phenomenon was later provided by Seidl (2002).

Another technique covered in this chapter is the bootstrap. Readers seeking more detail on this technique are referred to Efron & Tibshirani (1993) and MacKinnon (2002).

## Exercises

1. Burnham (2003) considers the binary decision to give nothing using the binomial test. Reproduce the results appearing in his Table 4.
2. Eckel & Grossman (1998) present a large number of tests of the effect of gender on dictator game giving. Their data set is presented within the article. Reproduce as many of their results as you can.
3. Branas-Garza (2007) investigate framing effects in dictator game giving. Dictators in the treatment group (2) had an additional line at the end of the instructions, which, translated from Spanish into English, means roughly "Note that your recipient relies on you". Dictators in the control group (1) did not see such a line. The experiment was performed both in a classroom (C) and in a lab (L). The distributions of contributions for each group are shown in the table:

| Donation | C1 | C2 | L1 | L2 |
|---|---|---|---|---|
| 0 | 11 | 2 | 5 | 1 |
| 1 | 1 | 0 | 2 | 2 |
| 2 | 4 | 3 | 6 | 3 |
| 3 | 2 | 7 | 6 | 5 |
| 4 | 1 | 5 | 5 | 10 |
| 5 | 1 | 3 | 3 | 5 |
| Total | 20 | 20 | 27 | 26 |

   (a) Create a data set with 93 rows (1 row for each subject), and create the following three variables:
   Setting: 1 if classroom; 2 if lab
   Treatment: 1 if additional sentence not seen; 1 if seen
   X: contribution

   (b) Reproduce their results.

# Chapter 4
# Theory Testing, Regression, and Dependence

## 4.1 Introduction

This chapter is mainly concerned with the role of experimental data in theory testing. As stressed at the start of Chapter 1, the focus of many economic experiments is on the functioning of a particular economic institution, and the objective is to test a particular economic theory in the context of that institution. The two institutions used as examples in this chapter are auctions and contests. For both of these institutions, the economic theory is very well developed, and leads to very clear "fundamental predictions", often in the form of the "risk-neutral Nash equilibrium prediction". This fundamental prediction provides a natural starting point for the range of tests demonstrated in this chapter.

In the settings considered in this chapter, and indeed in many experimental settings, behaviour of experimental subjects tends to depart in systematic ways from the "fundamental prediction" of theory. In the examples seen later, these departures take the form of systematic "over-bidding" relative to the Nash equilibrium prediction in auctions and contests. Hence, if our only objective were to test the fundamental predictions of Nash equilibrium theories, this objective would be straightforwardly met: the theory would be rejected. However, there are other levels at which theory may be tested. A typical theory gives rise to a number of "comparative static predictions". An example that is prominent in this chapter is the effect of changing the number of bidders in an auction. If the theory predicts that an increase in the number of bidders causes a decrease in bids, this is a comparative static prediction that can easily be tested using experimental data, simply by comparing the levels of bids between treatments with different numbers of bidders.

The principal objective of the chapter is to demonstrate how such comparative static predictions of the theory, and other types of treatment effect, may be tested within the framework of a linear regression model. The basic idea is that a treatment test can be performed as a test of significance of a "treatment dummy" in a linear regression model whose dependent variable is the outcome variable for the test. In fact, it can be shown that when a regression is performed with only a single dummy variable and an intercept, the t-test for the significance of the effect of the dummy is equivalent to the independent samples t-test, covered in Chapter 3, for testing the difference in means between the treatment and control groups.

The use of regression analysis for the purposes of treatment testing has a number of major advantages. Firstly, it is possible to test the effects of more than one treatment simultaneously, and, if necessary, interactions between them. Secondly, it is possible to control for the effects of other variables (e.g. subject characteristics) that might affect the outcome. Thirdly, it is possible to adjust for *dependence* between observations. Dependence normally takes the form of clustering, either at the level of the individual subject, or at the level of the group of subjects, or at the level of the experimental session. When a straightforward linear regression is being performed, an adjustment may easily be made to the standard errors to make them "cluster-robust", hence validating the treatment tests. Of course, a superior approach is fully to respect the "panel" structure of the data by using panel estimators, that are fully efficient. If there is more than one level of clustering (e.g. subject and session), a fully efficient estimator is one that follows the multi-level modelling approach. All of these methods are covered in this chapter.

There are several reasons for the choice of experimental auctions as the principal context in which the treatment tests are demonstrated. Firstly, as mentioned above, auction theory is very well developed (see Krishna, 2010) and gives rise to many clear predictions which may be tested econometrically. The central prediction is the risk-neutral Nash equilibrium (RNNE) bid function. Secondly, while it is possible to test these predictions using data on real auctions (see for example Laffont et al., 1995), the econometrics of experimental auctions is much simpler. This is because experimental auctions are conducted using the *induced value* methodology, in which, in any given round of the experiment, each subject is given a "private value" or "signal" by the experimenter, and makes their bidding decision on this basis. In other words, the private values are fully known by the investigator. This is unlike real auction data, in which private values are clearly unobserved, and estimation of their distribution presents a major obstacle to estimation and testing. With experimental data, because the private values are known, adherence or otherwise to the RNNE can be tested directly using a one-sample test, and certain comparative static predictions of auction theory can be tested using two-sample treatment tests. Also, the "bid function", that is, the equation showing bid as a function of private value, can be estimated using linear regression analysis, since both the dependent and explanatory variables are fully observed. This regression can of course be made to include treatment dummies so that the comparative static predictions of interest may be tested. Thirdly, the regression framework also allows the investigation of other determinants of the bid, such as experience and accumulated balance. Fourthly, experimental auction data has a very clear panel structure, with clustering at the level of bidder, and at the level of the experimental session, and so provides the ideal setting in which to demonstrate the panel data techniques described in the last paragraph.

When data from experimental auctions are analysed, there is an almost universal tendency for bidding to be more "aggressive" (i.e. higher) than predicted by RNNE. There are a number of explanations for this phenomenon, the dominant one being failure to adjust for the fact that the highest private signal is likely to be too high, this failure being known as the "winner's curse". Other explanations for over-bidding include: joy of winning the auction; use of simple heuristics; regret; confusion; experimenter demand effect; and house money effect. Some of these explanations are considered in applications of the testing techniques developed later.



The other context used for illustration is that of the contest experiment. Contest experiments are similar to auction experiments in some ways: it is straighforward to compute the RNNE, and there is a tendency to over-bid relative to the RNNE. However, the treatment tests we apply to contest data have a different sort of objective. With auctions, we are mainly interested in testing comparative static predictions *of the theory*. With contests, the emphasis shifts to an investigation of the "drivers of out-of-equilibrium play" – that is, which features *of the experimental design* lead to behaviour that is closer to, or further away from, the RNNE.

We also use the context of contest experiments to demonstrate the technique of meta-analysis, which provides yet another means of testing the predictions of theory and of identifying the drivers of out-of-equilibrium play.

Section 4.2 provides a minimal overview of auction theory, intended simply to introduce the concepts required for an understanding of the various theoretical predictions that are tested later in the chapter. Simulated experimental auction data is used in all of the sections that follow. Section 4.3 considers some basic tests of the fundamental predictions of auction theory. Section 4.4 considers tests of campara-tive static predictions, both as standard treatment tests and as tests within regression models. It is also explained how dependence can be accommodated. Section 4.5 extends the model to a multiple regression context, allowing for the effects of the level of uncertainty, the accumulation of bidding experience, and the accumulation of cash balances. Section 4.6 introduces panel data estimators and applies them to the models of Section 4.5. Section 4.7 demonstrates how multilevel modelling can be used to accommodate both subject-level and session-level dependence. Section 4.8 introduces contest experiments, and demonstrates some tests in this context. Section 4.9 considers a meta-analysis of a sample of published results from contest research. Section 4.10 summarises the chapter.

## 4.2 Experimental Auctions

### *4.2.1 Overview of auction theory*

There are two types of auction: common-value auctions and private-value auctions. In a common-value auction, each bidder submits a bid for a particular object which has the same value to all bidders, but this value is unknown. Since the value is unknown, a bidder can make a loss if they win the auction and then find that the price they are paying is higher than the true value of the object.

In a private-value auction, each bidder has a different value for the object, and this value is known to the bidder. All the bidder needs to do is make a bid lower than their own private value. That way, they make a profit, if they win the auction.

However, bidding too far below the private value clearly reduces the probability of winning.

The winner of the auction is the one who submits the highest bid. But what price does this bidder pay? It might seem obvious that they should pay the price that they bid. If they do, they are playing according to the rules of a "first-price" auction. However, some auctions, labelled "second price" auctions, are designed so that the winning bidder pays whatever price was bid by the *second highest* bidder. One important reason why we are interested in second-price auctions is that they are strategically equivalent to English (increasing) auctions. This is the type of auction with which we are most familiar, with an auctioneer starting with a low price and gradually increasing the price until only one bidder remains. Obviously the highest bidder will stop bidding immediately after the second highest bidder drops out.

A first-price (sealed-bid) auction is strategically equivalent to the less familiar "Dutch auction": an auctioneer starts with a high price and gradually lowers the price until it is accepted by the highest bidder.

One of the most remarkable results in auction theory is the *revenue-equivalence theorem*: with risk-neutral bidders, the expected price paid under both auctions (first-price and second-price) is the same.

Here, we are interested in the "bid function". That is the function $b(x)$ that tells us what the bid should be if the private signal (or private value) is x. Given any auction type, auction theory may be applied to predict the bid function. The bid function derived from the theory is usually the RNNE bid function.

One RNNE bid function in which we are particularly interested is that arising in a second-price common-value auction. As shown by Kagel et al. (1995), this bid function is:

$$b(x) = x - \frac{\epsilon(N-2)}{N} \tag{4.1}$$

where $N$ is the number of bidders, and $\epsilon$ is a measure of the level of uncertainty implicit in the private signals. The latter will be explained fully in due course.

Naturally, bids are predicted to be lower than the private signal. According to (4.1), the amount by which they are lower than the private signal clearly depends positively on the level of uncertainty ($\epsilon$) and positively on the number of bidders ($N$). These are comparative static predictions that will be tested in later sections by applying a variety of different testing techniques to the (simulated) using of data on bids.

The amount by which the bid is lower than the private signal is known as the "bid factor". We will label the bid factor "y". In the context of a second-price common-value auction, the (RNNE) bid factor is, from (4.1):

$$y = x - b(x) = \frac{\epsilon(N-2)}{N} \tag{4.2}$$

A very important feature of bidding behaviour in common-value auctions is "winner's curse". This is the tendency for subjects to fail fully to take into consideration the adverse selection problem: winning the lottery often implies that the private signal was "too high". Failure to adjust fully for this means that observed



bidding is usually somewhat higher than the RNNE prediction (4.1), and hence that the bid factor is somewhat *lower* than the RNNE bid factor (4.2).

### 4.2.2 Carrying out an experimental auction

We will describe the experimental methodology relevant to a second-price, common-value auction.

Subjects are recruited to sessions consisting of a series of auction periods. Because "stranger matching" is desirable, the number of subjects in a session must be greater than the number required for a single bidding group, and, in each round, different bidding groups are formed as random combinations of subjects within the session. For a given bidder-group in a given period, the experimenter generates a "true value" for the imaginary object, $x_0$; this value is not revealed to the bidders. The experimenter then provides each bidder with a "private signal", $x$, drawn from a uniform distribution on $[x_0 - \epsilon, x_0 + \epsilon]$. Bidders know the value of $\epsilon$, which is a measure of bidder uncertainty. But while they know their own private signal, they do not know the private signals of other bidders. Each subject in the bidding group submits a sealed bid for the item.

Each subject is given an initial balance at the start of the experiment. For a given bidder-group in a given period, the winner of the auction is the bidder with the highest bid. The winning bidder buys the object at a price equal to the bid of the *second highest* bidder. At this point, the true value of the object is revealed. The winning bidder receives a profit which is the true value minus the price paid. It is possible that this profit is negative. Other bidders receive a profit of zero for that round. In each round, each bidder's profit is added to their existing balance. Any subject whose balance falls below zero is declared "bankrupt" and is excluded from further rounds. Bankrupt subjects only receive a "participation fee" at the end of the session. Subjects who survive to the end of the session are paid their end-of-experiment balance as well as their participation fee.

### 4.2.3 The simulated auction data

Data have been simulated from a second-price common-value auction.[1] The simulated data is contained in the file **common_value_sim**. There are a total of 160 subjects, divided into 16 sessions. Each subject experiences 30 auction periods. The true values ($x_0$) have been drawn from a uniform distribution on [25, 975].

[1] The design assumed in the simulation bears strong similarities to the designs of Kagel et al. (1995) and Ham et al. (2005), although both of those are private value auctions – here we consider common value auctions.

Then, with each drawn $x_0$, a set of "private signals", $x$, have been drawn from a uniform distribution on $[x_0 - \epsilon, x_0 + \epsilon]$, with the parameter $\epsilon$ varying between sessions. Between the 16 sessions, values of $N$ and $\epsilon$ vary, as follows:

| Sessions | # subjects | $N$ | $\epsilon$ | RNNE bid function | RNNE bid factor | Observations |
|---|---|---|---|---|---|---|
| 1–4 | 8 | 4 | 12 | $b(x) = x - 6$ | 6 | 871 |
| 5–8 | 12 | 6 | 12 | $b(x) = x - 8$ | 8 | 1250 |
| 9–12 | 8 | 4 | 24 | $b(x) = x - 12$ | 12 | 840 |
| 13–16 | 12 | 6 | 24 | $b(x) = x - 16$ | 16 | 1273 |

For each of Sessions 1–4 and 9–12, there are eight subjects, and in each round, these subjects are divided into two groups of $N = 4$ bidders. In each of Sessions 5–8 and 13–16, 12 subjects are divided into two groups of $N = 6$ bidders.

Also given in the table are the RNNE bid functions and RNNE bid factors, for each group of sessions. Note that, in accordance with (4.1), these are different for each group of sessions, by virtue of the values of $N$ and $\epsilon$ changing between groups of sessions.

The design is a 2 × 2 *full-factorial* design, since it includes all four of the possible combinations of $N$ and $\epsilon$. This is essential for a proper test of the theory, since the theoretical bid function contains an interaction effect between $N$ and $\epsilon$, in addition to the two main effects.

In the simulation, each subject starts with a balance of 14 experimental units. The bids are simulated in such a way as to depend in expected ways on the private signal, the period number, the number of players, the level of uncertainty, and the cumulative balance of the subject. The total numbers of available observations (shown in the final column of the table) varies between treatments, for the following reasons. Treatments with $N = 6$ yield more data than those with $N = 4$. Also, when a subject goes bankrupt, they are excluded from further rounds so no further observations are generated from them. Finally, observations for which the private signals are outside of the range [60, 963] are excluded from estimation, because the RNNE bid function is more complicated at the ends of the distribution.

Figure 4.1 shows a screenshot of the first 31 rows of the data. It is useful to focus on one of these rows. The fifth row contains the following information. In the fifth period of the first session, subject 1 participated in market 1 (remember that in each period the group of subjects are divided into two markets), the true value of the object is 592, and subject 1 receives a private signal of 591. Subject 1 bids 587, so that her bid factor is $591 - 587 = 4$. With this bid, the subject wins the auction (indicated by winner = 1), and pays the second price (spr) which is 584. She earns a profit of $592 - 584 = 8$, which is added to her current balance of 14, making her balance 22 in the following period. Note from other rows that this subject wins the auction in 5 of the 30 rounds, and accumulates a balance of 38 at the end of the 30 rounds.

In order to examine particular auctions, it is more convenient to sort the data using the command `sort session period market i`. The sorted data set is shown in Figure 4.2. To see what happened in period 1 of session 1, we examine the first four rows. We see that subjects 2, 3, 4, and 6 were selected to take part in



| | session | i | period | market | x_0 | x | bid | winner | spr | profit | balance |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 2 | 596 | 589 | 582 | 0 | 602 | 0 | 14 |
| 2 | 1 | 1 | 2 | 2 | 375 | 378 | 370 | 0 | 370 | 0 | 14 |
| 3 | 1 | 1 | 3 | 2 | 556 | 545 | 548 | 0 | 591 | 0 | 14 |
| 4 | 1 | 1 | 4 | 2 | 945 | 950 | 942 | 0 | 942 | 0 | 14 |
| 5 | 1 | 1 | 5 | 1 | 592 | 591 | 587 | 1 | 584 | 8 | 14 |
| 6 | 1 | 1 | 6 | 1 | 387 | 395 | 391 | 1 | 386 | 1 | 22 |
| 7 | 1 | 1 | 7 | 1 | 754 | 744 | 743 | 0 | 752 | 0 | 23 |
| 8 | 1 | 1 | 8 | 1 | 744 | 735 | 734 | 0 | 750 | 0 | 23 |
| 9 | 1 | 1 | 9 | 2 | 661 | 670 | 661 | 0 | 661 | 0 | 23 |
| 10 | 1 | 1 | 10 | 1 | 913 | 903 | 899 | 0 | 905 | 0 | 23 |
| 11 | 1 | 1 | 11 | 2 | 927 | 933 | 928 | 0 | 926 | 0 | 23 |
| 12 | 1 | 1 | 12 | 1 | 370 | 359 | 352 | 0 | 374 | 0 | 23 |
| 13 | 1 | 1 | 13 | 1 | 727 | 719 | 713 | 0 | 716 | 0 | 23 |
| 14 | 1 | 1 | 14 | 2 | 703 | 712 | 709 | 1 | 706 | -5 | 23 |
| 15 | 1 | 1 | 15 | 1 | 311 | 303 | 297 | 0 | 298 | 0 | 18 |
| 16 | 1 | 1 | 16 | 2 | 139 | 142 | 138 | 0 | 138 | 0 | 18 |
| 17 | 1 | 1 | 17 | 2 | 65 | 67 | 58 | 0 | 58 | 0 | 18 |
| 18 | 1 | 1 | 18 | 1 | 438 | 437 | 435 | 0 | 439 | 0 | 18 |
| 19 | 1 | 1 | 19 | 2 | 203 | 194 | 196 | 0 | 196 | 0 | 18 |
| 20 | 1 | 1 | 20 | 2 | 531 | 533 | 525 | 0 | 529 | 0 | 18 |
| 21 | 1 | 1 | 21 | 1 | 328 | 332 | 324 | 0 | 326 | 0 | 18 |
| 22 | 1 | 1 | 22 | 1 | 575 | 586 | 576 | 1 | 565 | 10 | 18 |
| 23 | 1 | 1 | 23 | 2 | 765 | 756 | 746 | 0 | 751 | 0 | 28 |
| 24 | 1 | 1 | 24 | 2 | 519 | 509 | 508 | 0 | 513 | 0 | 28 |
| 25 | 1 | 1 | 25 | 2 | 507 | 500 | 495 | 0 | 505 | 0 | 28 |
| 26 | 1 | 1 | 26 | 2 | 487 | 484 | 482 | 0 | 483 | 0 | 28 |
| 27 | 1 | 1 | 27 | 2 | 709 | 714 | 711 | 0 | 711 | 0 | 28 |
| 28 | 1 | 1 | 28 | 1 | 779 | 787 | 783 | 1 | 769 | 10 | 28 |
| 29 | 1 | 1 | 29 | 2 | 697 | 685 | 676 | 0 | 687 | 0 | 38 |
| 30 | 1 | 1 | 30 | 2 | 185 | 180 | 172 | 0 | 184 | 0 | 38 |
| 31 | 1 | 2 | 1 | 1 | 705 | 717 | 713 | 1 | 698 | 7 | 14 |

Figure 4.1: Common value auction data

| | session | i | period | market | x_0 | x | bid | winner | spr | profit | balance | y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 1 | 1 | 705 | 717 | 713 | 1 | 698 | 7 | 14 | 4 |
| 2 | 1 | 3 | 1 | 1 | 705 | 700 | 696 | 0 | 698 | 0 | 14 | 4 |
| 3 | 1 | 4 | 1 | 1 | 705 | 698 | 697 | 0 | 698 | 0 | 14 | 2 |
| 4 | 1 | 6 | 1 | 1 | 705 | 705 | 698 | 0 | 698 | 0 | 14 | 7 |
| 5 | 1 | 1 | 1 | 2 | 596 | 569 | 582 | 0 | 602 | 0 | 14 | 7 |
| 6 | 1 | 5 | 1 | 2 | 596 | 595 | 562 | 0 | 602 | 0 | 14 | 13 |
| 7 | 1 | 7 | 1 | 2 | 596 | 605 | 602 | 0 | 602 | 0 | 14 | 3 |
| 8 | 1 | 8 | 1 | 2 | 596 | 605 | 603 | 1 | 602 | -6 | 14 | 2 |
| 9 | 1 | 2 | 2 | 1 | 303 | 302 | 290 | 0 | 309 | 0 | 21 | 12 |
| 10 | 1 | 6 | 2 | 1 | 303 | 299 | 298 | 0 | 309 | 0 | 14 | 1 |
| 11 | 1 | 7 | 2 | 1 | 303 | 313 | 309 | 0 | 309 | 0 | 14 | 4 |
| 12 | 1 | 8 | 2 | 1 | 303 | 315 | 312 | 1 | 309 | -6 | 8 | 3 |
| 13 | 1 | 1 | 2 | 2 | 375 | 378 | 370 | 0 | 370 | 0 | 14 | 8 |
| 14 | 1 | 3 | 2 | 2 | 375 | 372 | 368 | 0 | 370 | 0 | 14 | 4 |
| 15 | 1 | 4 | 2 | 2 | 375 | 373 | 372 | 1 | 370 | 5 | 14 | 1 |
| 16 | 1 | 5 | 2 | 2 | 975 | 379 | 362 | 0 | 370 | 0 | 14 | 17 |
| 17 | 1 | 3 | 3 | 1 | 442 | 433 | 428 | 0 | 430 | 0 | 14 | 5 |
| 18 | 1 | 6 | 3 | 1 | 442 | 448 | 441 | 1 | 430 | 12 | 14 | 7 |
| 19 | 1 | 7 | 3 | 1 | 442 | 432 | 430 | 0 | 430 | 0 | 14 | 2 |
| 20 | 1 | 8 | 3 | 1 | 442 | 446 | 441 | 1 | 430 | 12 | 2 | 5 |
| 21 | 1 | 1 | 3 | 2 | 556 | 545 | 548 | 0 | 551 | 0 | 14 | -3 |
| 22 | 1 | 2 | 3 | 2 | 556 | 559 | 551 | 0 | 551 | 0 | 21 | 8 |
| 23 | 1 | 4 | 3 | 2 | 556 | 563 | 560 | 1 | 551 | 5 | 19 | 3 |
| 24 | 1 | 5 | 3 | 2 | 556 | 560 | 547 | 0 | 551 | 0 | 14 | 13 |
| 25 | 1 | 2 | 4 | 1 | 322 | 333 | 317 | 0 | 320 | 0 | 21 | 16 |
| 26 | 1 | 4 | 4 | 1 | 322 | 321 | 320 | 0 | 320 | 0 | 24 | 1 |
| 27 | 1 | 7 | 4 | 1 | 322 | 331 | 926 | 1 | 320 | 2 | 14 | 5 |
| 28 | 1 | 8 | 4 | 1 | 322 | 316 | 313 | 0 | 320 | 0 | 14 | 3 |
| 29 | 1 | 1 | 4 | 2 | 945 | 950 | 942 | 0 | 942 | 0 | 14 | 8 |
| 30 | 1 | 3 | 4 | 2 | 945 | 845 | 940 | 0 | 942 | 0 | 14 | 5 |
| 31 | 1 | 5 | 4 | 2 | 949 | 938 | 924 | 0 | 942 | 0 | 14 | 14 |

Figure 4.2: Common value auction data: re-sorted using "sort session period market i"

"market 1". The true value was 705, and the auction was won by subject 2 with a bid of 713. This subject pays a price 698, the bid of subject 6, this being the second highest bid. The profit earned by subject 2 is $705 - 698 = 7$.

## 4.3 Tests of Auction Theory

### *4.3.1 A test of RNNE in a second-price common value auction*

In this section we consider direct tests of the fundamental predictions of the theory. These tests are one-sample tests applied to the bid factor. An important point is that when conducting these tests, we are making the assumption of independence of observations, an assumption which, as we will see many times later, is hard to justify. Hence the results obtained in this section might not be taken too seriously. On several occasions later in the chapter, the fundamental prediction will be again tested, but within the context of regression. There, appropriate adjustments are made for dependence in the data simply because it is straightforward to make such adjustments in the regression framework. Those regression-based tests will provide more reliable conclusions regarding the data's closeness to theoretical predictions.

Here, we consider the two benchmark models used by Kagel et al. (1995): a naïve bidding model, and the RNNE. The first model represents an extreme form of naïvety. The second represents an extreme form of rationality. It is not anticipated that observed behaviour will correspond closely to either of the two models.

Let us focus on Sessions 1–4, in which the number of bidders ($N$) is fixed at 4, and the uncertainty parameter ($\epsilon$) is fixed at 12. If we insert these values into (4.1) we observe that the RNNE prediction in these sessions is simply:

$$b(x) = x - 6 \tag{4.3}$$

Another way of stating this prediction is that the bid factor ($y$) equals 6. In Figure 4.3, we show a histogram of the bid factor for sessions 1–4, with a vertical line at the RNNE prediction of 6. The histogram shows that there is a good deal of variation in the bid factor, but it is fairly clear that the main part of the distribution is to the left of the RNNE prediction of 6.

An obvious way to conduct a formal statistical test of the theory is to test the null hypothesis that the population mean of $y$, $\mu$ say, is equal to 6.0, against the alternative hypothesis that it equals some value less than 6.0 (since values less than 6.0 arise as a result of "winner's curse"). If there are $n$ observations in the data set, we would first find the mean and standard deviation of the bid factor, $\bar{y}$ and s, and we would compute the one-sample t-test statistic:

$$t = \frac{\bar{y} - 6.0}{s/\sqrt{n}} \tag{4.4}$$



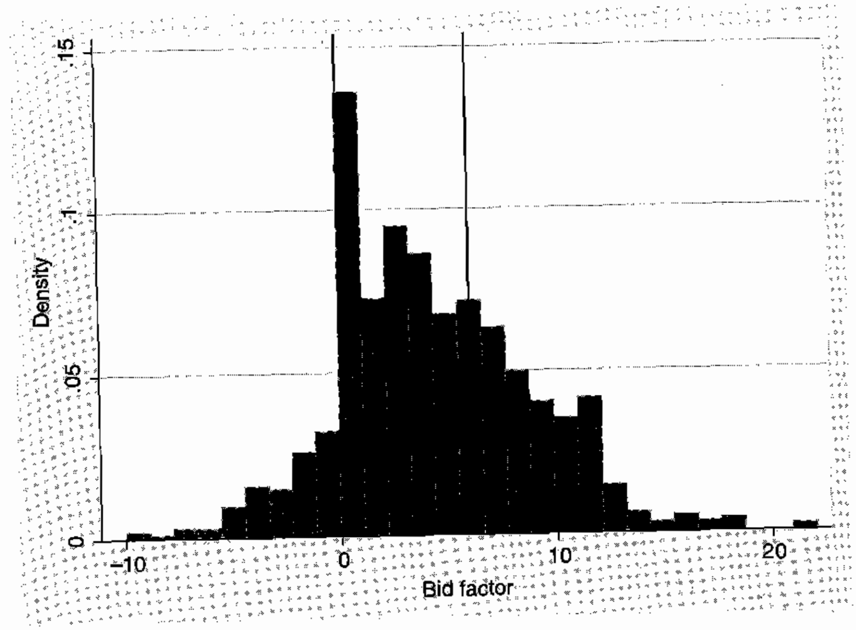



Figure 4.3: A histogram of the bid factor for sessions 1–4. Vertical lines drawn at RNNE prediction (6) and naïve prediction (0)

It is easier to do this test in STATA:

```
. * USE ONLY FIRST 4
SESSIONS (TO FIX N=4 AND EPSILON=6)

. keep if session<5
(1349 observations deleted)

. * test of RNNE

. ttest y=6

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       y |     871    4.337543       .1535      4.5302    4.036269    4.638817
------------------------------------------------------------------------------
    mean = mean(y)                                              t = -10.8303
Ho: mean = 6                                     degrees of freedom =      870

    Ha: mean < 6                 Ha: mean != 6                 Ha: mean > 6
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

We see that the mean bid factor is 4.34, but this is significantly below 6.0, as is clear from the p-value of 0.0000. We therefore strongly reject the RNNE prediction, and we conclude that "winner's curse" is taking hold (since bids are, on average, higher than predicted by RNNE).

The naive bidding model is one in which bidders simply bid their own private signal, implying that the bid factor is zero. This is clearly a very extreme form of winner's curse. Another vertical line in Figure 4.3 represents this naïve prediction, and it is clear that the majority of the distribution is to the right of this prediction. To conduct a formal test of this naïve model we test the null hypothesis that the bid factor has a mean of zero:

```
. * test of naive behaviour

. ttest y=0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       y |     871    4.337543       .1535      4.5302    4.036269    4.638817
------------------------------------------------------------------------------
    mean = mean(y)                                                t =  28.2576
Ho: mean = 0                                     degrees of freedom =      870

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

This model is also strongly rejected (p=0.0000). Although we have found evidence that bidders are falling prey to winner's curse, they clearly understand that sensible bids are somewhat below their private signals.

## 4.4 Tests of Comparative Static Predictions

Two very simple models were tested in the last section: the RNNE and the naïve bidding model. Notwithstanding the failure to take account of dependence in the data, both models were decisively rejected. The truth is clearly somewhere in between these two extremes. However, there are many other predictions of the theory that can be tested.

In this section, we consider tests of one of the comparative static predictions of the RNNE theory. Consider the effect on bidding of the number of bidders, $N$. Do we expect bidding behaviour to change if $N$ is changed and all other features of the auction remain the same? Once again appealing to (4.1), we see that the answer is yes. Under RNNE theory, we expect an increase in $N$ to have a negative effect on bids, and therefore a positive effect on the bid factor.

First, we conduct these tests with the assumption of independence between observations. In Section 4.4.3, we shall start to consider methods for allowing for dependence.

### *4.4.1 Standard treatment tests*

The comparative static prediction relating to the number of bidders ($N$) may be tested using standard treatment tests. $N$ takes two different values in the experiment: 4 and 6. The data contains a dummy variable named "N6" taking the value 1 when



$N = 6$ and 0 when $N = 4$. This dummy variable is used as the separation variable for the test. Think of N6 as a "treatment dummy": $N6 = 1$ for the "treatment group" and $N6 = 0$ for the "control group".

Note that for these tests we are using only Sessions 1–8. This is because the other design parameter ($\epsilon$) is fixed over these sessions, as is required if we are to focus on the effect of changes in $N$.

```
. keep if session<9
(2247 observations deleted)

. ranksum y, by(N6)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

          N6 |      obs    rank sum    expected
-------------+---------------------------------
           0 |      871   843078.5      924131
           1 |     1250  1407302.5     1326250
-------------+---------------------------------
    combined |     2121    2250381     2250381

unadjusted variance   1.925e+08
adjustment for ties  -984989.31
                     ----------
adjusted variance     1.915e+08

Ho: y(N6==0) = y(N6==1)
             z =  -5.856
    Prob > |z| =   0.0000

. ttest y, by(N6)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     871    4.337543       .1535      4.5302    4.036269    4.638817
       1 |    1250       5.348    .1163788    4.114613     5.11968     5.57632
---------+--------------------------------------------------------------------
combined |    2121     4.93305    .0937551    4.317827    4.749189    5.116912
---------+--------------------------------------------------------------------
    diff |           -1.010457    .1893543               -1.381797   -.6391172
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -5.3363
Ho: diff = 0                                     degrees of freedom =     2119

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Both the Mann-Whitney test (`ranksum` in STATA) and the independent samples t-test (`ttest` in STATA) result in very strong rejections of the null hypothesis that $N$ has no effect on bids. Also, we must not forget to check the direction of the effect. The second table tells us that when $N = 4$, the mean bid factor is 4.34, while when $N = 6$, the mean bid factor is higher at 5.35. This implies that bids are lower when $N$ increases, and this is what is predicted by the theory. We may therefore conclude that the results of these treatment tests are consistent with the predictions of RNNE theory.

### 4.4.2 Treatment testing using a regression

Treatment tests may also be performed using a regression. In fact, regression on only a dummy variable is equivalent to an independent samples t-test. To verify this, we perform a regression of the bid factor on the dummy N6.

```
. regress y N6

      Source |       SS       df       MS              Number of obs =    2121
-------------+------------------------------           F(  1,  2119) =   28.48
       Model |  524.110821     1  524.110821           Prob > F      =  0.0000
    Residual |  39000.3823  2119  18.4050884           R-squared     =  0.0133
-------------+------------------------------           Adj R-squared =  0.0128
       Total |  39524.4932  2120  18.6436289           Root MSE      =  4.2901

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          N6 |   1.010457   .1893543     5.34   0.000     .6391172    1.381797
       _cons |   4.337543    .145365    29.84   0.000      4.05247    4.622616
------------------------------------------------------------------------------
```

Note that the intercept is 4.34 (the mean of the bid factor in the "control" group) and the coefficient of N6 is 1.01 (the difference in means between the "treatment" and "control" groups, i.e. the effect size). Note also that the t-statistic is identical (in magnitude) to the t-statistic obtained using the independent samples t-test in the last sub-section. Hence the conclusion from this regression test is exactly the same: when N increases, the bid factor increases, meaning that bids decrease.

### 4.4.3 Accounting for dependence: the ultra-conservative test

The tests and regressions performed so far all assume independence between observations. In the current setting (and in most settings in experimental economics) the independence assumption is not met. First of all, since subjects are being observed in a sequence of 30 auction periods, there is dependence within subjects. Since some subjects are more "aggressive" bidders than others (or just more susceptible to winner's curse), their complete set of bids will tend to be higher than those of others. We sometimes describe this phenomenon as "clustering" at the level of the individual subject. There may also be clustering at the level of the session: some sessions may be characterised by aggressive bidding, others by reserved bidding. The presence of clustering means that the tests performed above are invalid.

The most obvious way of dealing with the dependence problem is to start by taking the averages over all dependent observations within each cluster, and then performing standard treatment tests on these averages. The resulting data set has only one observation for each independent unit, and is therefore free of the problem of dependence. The obvious drawback from using this procedure is that the sample size becomes very small, and hence the power of the test is limited. This is the reason why the procedure is described as "ultra conservative"; whatever the observed significance (p-value) of this test, we expect any other test to result in stronger significance (i.e. a smaller p-value).



In the present context, application of the ultra-conservative test involves finding the average bid factor for each session (using the STATA command `table session, contents(n y mean y)`). The resulting means are shown in the following table.

| Sessions | $N$ | $\epsilon$ | RNNE bid factor | Mean bid factors by session |
|---|---|---|---|---|
| 1–4 | 4 | 12 | 6 | 6.25, 5.69, 2.58, 2.36 |
| 5–8 | 6 | 12 | 8 | 7.32, 5.72, 4.31, 3.68 |
| 9–12 | 4 | 24 | 12 | 10.74, 10.73, 9.40, 6.96 |
| 13–16 | 6 | 24 | 16 | 15.78, 12.99, 12.73, 10.63 |

Once again we restrict attention to Sessions 1–8, for which the uncertainty parameter $\epsilon$ is fixed at 12. We have eight independent observations, four with $N = 4$ and four with $N = 6$. We may apply conventional treatment tests to these eight observations, and we obtain the following results.

```
. ranksum mean_y, by(n6)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

          n6 |      obs    rank sum    expected
-------------+---------------------------------
           0 |        4          15          18
           1 |        4          21          18
-------------+---------------------------------
    combined |        8          36          36

unadjusted variance       12.00
adjustment for ties        0.00
                     ----------
adjusted variance         12.00

Ho: mean_y(n6==0) = mean_y(n6==1)
             z =  -0.866
    Prob > |z| =   0.3865

. ttest mean_y, by(n6)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |       4        4.22      1.0178      2.0356    .9809063    7.459094
       1 |       4      5.2575    .8090156    1.618031    2.682851    7.832149
---------+--------------------------------------------------------------------
combined |       8     4.73875    .6329902    1.790367    3.241966    6.235534
---------+--------------------------------------------------------------------
    diff |            -1.0375    1.300163               -4.218884    2.143883
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -0.7980
Ho: diff = 0                                     degrees of freedom =        6

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2276         Pr(|T| > |t|) = 0.4553          Pr(T > t) = 0.7724
```

Both the Mann-Whitney test and independent samples t-test find no evidence that the number of players has any effect on bids. Remembering that we expect the bid factor to be higher when $N = 6$ than when $N = 4$, we may perform one-tailed tests.

The one-tailed p-values are 0.19 for the Mann-Whitney test (0.38 divided by 2), and 0.23 for the t-test. However, for the reason given above, we must treat these p-values as "ultra conservative", and, on the assumption that a treatment effect is actually present, we anticipate smaller p-values when different types of test are used to test the same hypothesis.

### *4.4.4 Accounting for dependence in a regression*

In Section 4.4.3 we demonstrated a test which explicitly avoids the dependence problem. However, it should be obvious that the procedure is far from satisfactory in view of the large amount of information that is being discarded in the process of averaging observations. Clearly, we wish to utilise a testing procedure that makes use of all of the available information contained in the data set, but which at the same time is robust to the data's dependence structure.

With this objective, we return to the regression framework introduced in Section 4.4.2. In the context of regression, dependence gives rise to a non-diagonal covariance matrix of the error term. This is a violation of one of the classical assumptions of the linear regression model, which requires that the covariance matrix of the error vector is diagonal. The covariance matrix is in fact block-diagonal as a result of the "clustering" of observations by subject or session.

One major advantage of the regression framework for conducting treatment tests is that there are well-established routines for correcting the results to allow for dependence between observations. It is well known (see, for example, Greene, 2008) that whenever the error covariance matrix is non-diagonal, the formula routinely used to compute standard errors is incorrect. There is usually a correct formula, and the choice of formula depends on the nature of the non-diagonality. In this situation of block-diagonality, the appropriate formula is applied by using the `vce(cluster clustvar)` option in STATA, where `clustvar` is the variable specifying to which "cluster" each observation belongs (e.g. the subject identifier, $i$).

Applying this correction to the treatment test peformed in Section 4.4.2, assuming that the clustering is at the level of subject id (i), we obtain the results:

```
. regress y N6, vce(cluster i)

Linear regression                                      Number of obs =    2121
                                                       F( 1,     79) =    1.90
                                                       Prob > F      =  0.1725
                                                       R-squared     =  0.0133
                                                       Root MSE      =  4.2901

                                    (Std. Err. adjusted for 80 clusters in i)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          N6 |   1.010457   .7339144     1.38   0.172    -.4503632    2.471277
       _cons |   4.337543   .5999469     7.23   0.000     3.143379    5.531707
------------------------------------------------------------------------------
```



Note that the coefficients are the same as before. All that is changing is the standard errors (and the various quantities that are computed from the standard errors). The standard error of the coefficient of N6 has risen from 0.189 to 0.734 as a result of dealing with clustering. This means that the t-test statistic for the treatment effect is much lower than before, and, unfortunately, the effect is no longer significant (p=0.172).

Adjusting for dependence has transformed a strongly significant treatment effect into an insignificant treatment effect. This example emphatically serves to illustrate the importance of adjusting for dependence in carrying out such tests.

### *4.4.5 Accounting for dependence: the block bootstrap*

As explained in Chapter 3, "bootstrap tests" are very popular in Experimetrics because they provide a means of carrying out standard parametric tests (e.g. the t-test) without relying on the distributional assumptions that are normally required.

Recall that, in the case of *independent* observations, the bootstrap procedure consists of the following steps:

1. Apply the chosen parametric test on the data set, obtaining a test statistic, $\hat{t}$.
2. Generate a healthy number, B, of "bootstrap samples". These are samples of the same size as the original sample. They are also drawn from the original sample, but the key point is that they are drawn *with replacement*. For each bootstrap sample, compute the test statistic, $\hat{t}_j^*$, $j = 1, \ldots, B$.
3. Compute the standard deviation $s_B$ of the bootstrap test statistics $\hat{t}_j^*$, j=1,..., $B$. This will be the bootstrap standard error.

In the presence of clustering, the above procedure will fail, since it fails to replicate the dependence in the data. The *block bootstrap* attempts to replicate the dependence by re-sampling *blocks* of data rather than single observations.

To apply the block bootstrap in STATA, we introduce `bootstrap` into the brackets after the `vce` option. We also specify the number of bootstrap samples. The results are as follows.

```
. reg y N6, vce(bootstrap, rep(999) cluster( i) )
(running regress on estimation sample)

Bootstrap replications (999)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
..................................................   550
..................................................   600
..................................................   650
```

```
..................................................   700
..................................................   750
..................................................   800
..................................................   850
..................................................   900
..................................................   950
..................................................

Linear regression                               Number of obs    =      2121
                                                Replications     =       999
                                                Wald chi2(1)     =      1.73
                                                Prob > chi2      =    0.1886
                                                R-squared        =    0.0133
                                                Adj R-squared    =    0.0128
                                                Root MSE         =    4.2901

                                   (Replications based on 80 clusters in i)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                       Normal-based
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          N6 |   1.010457   .7685028     1.31   0.189    -.4957809    2.516695
       _cons |   4.337543   .6176459     7.02   0.000     3.126979    5.548107
------------------------------------------------------------------------------
```

On this occasion, the result from the bootstrap test is not very different from that with the cluster robust standard error in Section 4.4.4.

## 4.5 Multiple Regression with Auction Data

All of the regressions considered so far have contained just one dummy explanatory variable. A further obvious advantage of the regression framework is that many determinants of behaviour can be investigated simultaneously. A model with more than one explanatory variable is a "multiple regression model". To such models we now turn.

### *4.5.1 Introducing the effect of uncertainty*

Recall that the parameter $\epsilon$ is a design parameter which represents the level of uncertainty, in the sense that the higher the value of $\epsilon$, the wider the range of private signals around the true value, and hence the less confidence that an individual bidder can have that his or her private signal is close to the true value. Once again appealing to (4.1), we see that RNNE theory predicts $\epsilon$ to have a negative effect on bids; that is, if the value of $\epsilon$ is increased and nothing else changes, we expect lower bidding. This is another comparative static prediction of RNNE theory. In the experiment, there are two different values of $\epsilon$, 12 and 24. This variation enables us to test the comparative static prediction, in a manner similar to our tests of the effect of the number of bidders.

Note that there are now two "treatment variables": N6, which we have already considered, and "eps24", a dummy variable taking the value 1 when $\epsilon = 24$, zero



otherwise. At this stage it is useful to reproduce the table shown earlier that summarises the features of the experiment.

| Sessions | Number of subjects | $N$ | $\epsilon$ | RNNE bid function | RNNE bid factor |
|---|---|---|---|---|---|
| 1–4 | 8 | 4 | 12 | $b(x) = x - 6$ | 6 |
| 5–8 | 12 | 6 | 12 | $b(x) = x - 8$ | 8 |
| 9–12 | 8 | 4 | 24 | $b(x) = x - 12$ | 12 |
| 13–16 | 12 | 6 | 24 | $b(x) = x - 16$ | 16 |

With this information, it is possible to write down a ("true") regression model involving both treatment dummies that fully captures the RNNE predictions. This model is:

$$bid = x - 6 - 2.N6 - 6.eps24 - 2.N6 \star eps24 \tag{4.5}$$

For the dependent variable, we are now using the bid itself, not the bid factor. This is possible as long as the private signal ($x$) is included as an explanatory variable, and we obviously expect its coefficient to take the value one. Note that in addition to the two treatment dummies, the equation includes the interaction term $N6 \star eps24$ formed as the product of the two dummies. It will be possible to estimate the coefficient of this interaction variable only because a "full factorial design" has been employed.

Equation (4.5) can be estimated (using the complete data set of all 16 sessions) using a multiple regression. Having estimated the equation, a test command may be used to perform a Wald test of the joint hypothesis that the coefficients are in agreement with the RNNE coefficients given in (4.5). Note that this amounts to a further (more stringent) test of RNNE. This is an F-test, since five separate equalities are being tested simultaneously, one for each of the terms in (4.5).

```
. regress bid x N6 eps24 N6eps24, vce(cluster i)

Linear regression                                      Number of obs =    4234
                                                       F(  4,   159) =       .
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9997
                                                       Root MSE      =  4.3029

                                      (Std. Err. adjusted for 160 clusters in i)
------------------------------------------------------------------------------
             |               Robust
         bid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.000129   .0002183  4581.20   0.000     .9996974     1.00056
          N6 |  -1.012588   .7326532    -1.38   0.169    -2.459576    .4343989
       eps24 |  -5.112904   .8103825    -6.31   0.000    -6.713407   -3.512402
     N6eps24 |  -2.661983   1.024279    -2.60   0.010    -4.684929   -.6390368
       _cons |  -4.401871   .6015577    -7.32   0.000    -5.589945   -3.213797
------------------------------------------------------------------------------

. test (x=1) (_cons=-6) (N6=-2) (eps24=-6) (N6eps24=-2)

 ( 1)  x = 1
 ( 2)  _cons = -6
 ( 3)  N6 = -2
 ( 4)  eps24 = -6
 ( 5)  N6eps24 = -2

       F(  5,   159) =   23.01
            Prob > F =    0.0000
```

All of the coefficients appear to have the "correct" signs, and most are significant despite the use of cluster-robust standard errors. In particular, note that when the level of uncertainty ($\epsilon$) rises from 12 to 24, *ceteris paribus*, the bid falls by 5.11, not too far from the prediction of 6.0. In fact, since the 95% confidence interval for this parameter contains the RNNE prediction of 6.0, we may conclude that this comparative static prediction is confirmed by the data.

However, the Wald test that tests all of the RNNE predictions simultaneously results in a strong rejection of the RNNE hypothesis, with a p-value of 0.0000.

### 4.5.2 Introducing the effect of experience

The role of experience is accounted for using the variable representing the period number. We expect experience to "improve" bidders' performance, in the sense of moving them closer to the RNNE prediction. We also expect a pattern of "convergence", with bidding reducing steeply in the early periods, but settling down to a stable bidding pattern in later periods. This is captured by including the reciprocal of period number, $1/period$, as an explanatory variable. A positive coefficient on this variable will indicate that the learning process takes the form of a trend towards RNNE. The model's other parameters may be interpreted in terms of a long term equilibrium.

Adding $1/period$ (named `rec_period` in the data set) to the regression gives the following results.

```
. regress bid x rec_period N6 eps24 N6eps24, vce(cluster i)

Linear regression                                      Number of obs =    4234
                                                       F(  5,    159) =       .
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9997
                                                       Root MSE      =  4.2658

                                   (Std. Err. adjusted for 160 clusters in i)
------------------------------------------------------------------------------
             |               Robust
         bid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.000184   .0002142  4670.16   0.000     .9997613    1.000607
  rec_period |   2.911282   .3183578     9.14   0.000     2.282527    3.540038
          N6 |   -1.04156   .7324525    -1.42   0.157     -2.48815    .4050315
       eps24 |  -5.126073   .8112263    -6.32   0.000    -6.728242   -3.523905
     N6eps24 |  -2.619187   1.024981    -2.56   0.012     -4.64352    -.594853
       _cons |  -4.821608   .6031499    -7.99   0.000    -6.012827   -3.630389
------------------------------------------------------------------------------

. test (x=1) (_cons=-6) (N6=-2) (eps24=-6) (N6eps24=-2)

 ( 1)  x = 1
 ( 2)  _cons = -6
 ( 3)  N6 = -2
 ( 4)  eps24 = -6
 ( 5)  N6eps24 = -2

       F(  5,   159) =   15.81
            Prob > F =    0.0000
```

The variable $1/period$ has a strongly positive coefficient as expected. The test that follows is again a Wald test of RNNE. However, note that this time it is a test of the



closeness of "long run" behaviour (i.e. behaviour after learning) to the RNNE. It is interesting that, although RNNE is again strongly rejected, the rejection is weaker than before, with an F-statistic of 15.8 (compared with 23.0 in the previous model). This is consistent with bidders becoming closer to RNNE with experience.

### 4.5.3 Introducing the effect of cash balance

Recall that subjects start with 14 units, and then profits or losses are earned in each period. "Cash balance" is the amount of money that a subject has accumulated up to the current period. This variable may well have an effect on bidding behaviour for more than one reason. One is "limited liability": since subjects are aware that they will simply be excluded from further rounds if their balance falls below zero, there may be an incentive to bid over-aggressively when the balance becomes low, in a gamble to improve the balance. If this is the case, we expect the cash balance to have a negative effect on bids. However, we might also expect a form of "house money effect": bidders might be expected to bid more aggressively when their balance is high, since this provides a "cushion" against losses, and it is not their money anyway. If this is the case, cash balance is expected to affect bids positively.

When cash balance is included, the results are as follows:

```
. regress bid x balance rec_period N6 eps24 N6eps24, vce(cluster i)

Linear regression                                      Number of obs =    4234
                                                       F(  6,    159) =       .
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9997
                                                       Root MSE      =  4.2602

                                   (Std. Err. adjusted for 160 clusters in i)
------------------------------------------------------------------------------
             |               Robust
         bid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.000188   .0002148  4657.23   0.000     .9997643    1.000613
     balance |  -.0124342   .0096541    -1.29   0.200    -.0315011    .0066327
  rec_period |   2.600908   .3841625     6.77   0.000     1.842188    3.359627
          N6 |  -1.158259   .7417146    -1.56   0.120    -2.623142    .3066246
       eps24 |  -4.887118    .842503    -5.80   0.000    -6.551058   -3.223177
     N6eps24 |  -2.730037   1.032137    -2.65   0.009    -4.768504   -.6915695
       _cons |  -4.480833   .6760663    -6.63   0.000    -5.816061   -3.145605
------------------------------------------------------------------------------

. test (x=1) (_cons=-6) (N6=-2) (eps24=-6) (N6eps24=-2)

 ( 1)  x = 1
 ( 2)  _cons = -6
 ( 3)  N6 = -2
 ( 4)  eps24 = -6
 ( 5)  N6eps24 = -2

       F(  5,   159) =    9.57
            Prob > F =    0.0000
```

We see that cash balance has a negative effect, pointing to a "limited liability" effect, but note that this effect is not significant.

## 4.6 Panel Data Estimators

We have been considering methods for dealing with dependence in experimental data. However, so far we have only taken a single step in this direction, by using cluster standard errors. A superior approach is to handle the data set in a panel data framework, which explicitly recognises that $n$ subjects are observed making a decision in each of $T$ time periods. With this approach, it is possible to improve the estimates themselves (i.e. not only the standard errors), using a panel-data estimator instead of OLS. The two most popular panel data estimators are the fixed-effects and random-effects estimators. Both can be represented by the following equation:

$$\begin{aligned} bid_{it} &= \alpha + \beta' x_{it} + \gamma' z_i + u_i + \epsilon_{it} \\ i &= 1\ldots, n \quad t = 1\ldots, T \\ Var(u_i) &= \sigma_u^2 \\ Var(\epsilon_{it}) &= \sigma_\epsilon^2 \end{aligned} \tag{4.6}$$

Note that in (4.6) there are two types of explanatory variable. The vector $x_{it}$ contains variables which vary both between subjects and time periods, for example cash balance. The vector $z_i$ contains variables which vary only between subjects, and are fixed over time. Examples of variables appearing in $z_i$ are treatment dummies (where between-subject treatments are applied) and subject characteristics. Note also that there are two error terms: $\epsilon_{it}$ is the conventional equation error term, which is assumed to have mean zero and variance $\sigma_\epsilon^2$ ; $u_i$ is known as the subject-specific term; $u_i$ differs between subjects – hence the $i$-subscript – but it is fixed for a given subject. The two estimators differ in the way this term is interpreted.

The fixed effects estimator is essentially a linear regression which includes a set of $n-1$ dummy variables, one for each subject in the data set (with one excluded to avoid the dummy variable trap). The presence of such dummies has the consequence that the intercept is estimated separately for each subject: the intercept for subject $i$ is $\alpha + u_i, i = 1\ldots, n$.

The random effects estimator does *not* estimate the intercept for each subject. It simply recognises that they are all different, and sets out to estimate only their *variance*, $\sigma_u^2$. Note that random effects is more efficient than fixed effects, because there are far fewer parameters to estimate. We therefore prefer to use random effects if this model turns out to be acceptable.

Another reason for preferring random effects over fixed effects is that the effects of time-invariant variables are not identified under fixed effects. That is, the parameter vector $\gamma$ in (4.6) is not identified under fixed effects. This is important because the variables in which we are most interested, namely the treatment variables, are time-invariant, unless a within-subject design has been used.

Under normal circumstances, to decide between fixed effects (FE) and random effects (RE), the Hausman test is used. In the current situation, however, the FE model is not useful. This is because the variables in whose effects we are most interested (i.e. the treatment dummies N6 and eps24) do not vary within a given subject. Such variation is essential for fixed effects estimation.



The random effects model is, however, useful, and the results are presented below. Panel data commands in STATA can be recognised by the prefix `xt`. For example panel data (linear) regression is carried out using `xtreg`. The fixed effects and random effects estimators are carried out using this command with the options `fe` and `re` respectively. Here we use `re`. Note also that we need to start by declaring the data as panel data using `xtset`, specifying the panel variable followed by the time variable.

```
. xtset i period
       panel variable:  i (unbalanced)
        time variable:  period, 1 to 30, but with gaps
                delta:  1 unit

. xtreg bid x balance rec_period N6 eps24 N6eps24 , re

Random-effects GLS regression                   Number of obs      =      4234
Group variable: i                               Number of groups   =       160

R-sq:  within  = 0.9999                         Obs per group: min =         1
       between = 0.9957                                        avg =      26.5
       overall = 0.9997                                        max =        30

                                                Wald chi2(6)       =  3.22e+07
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         bid |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.000459   .0001763  5674.85   0.000     1.000114    1.000805
     balance |  -.0050899    .003371    -1.51   0.131    -.011697     .0015171
  rec_period |   2.453855   .2521596     9.73   0.000     1.959632    2.948079
          N6 |  -1.115177   .7052598    -1.58   0.114    -2.497461    .2671071
       eps24 |  -5.180012   .7765955    -6.67   0.000    -6.702112   -3.657913
     N6eps24 |  -2.594669   .9987159    -2.60   0.009    -4.552116   -.6372215
       _cons |  -4.482867   .5609601    -7.99   0.000    -5.582329   -3.383406
-------------+----------------------------------------------------------------
     sigma_u |  3.0182377
     sigma_e |  2.9870243
         rho |  .50519753   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. test (x=1) (_cons=-6) (N6=-2) (eps24=-6) (N6eps24=-2)

 ( 1)  x = 1
 ( 2)  _cons = -6
 ( 3)  N6 = -2
 ( 4)  eps24 = -6
 ( 5)  N6eps24 = -2

           chi2(  5) =   99.47
         Prob > chi2 =    0.0000
```

Note that the estimates of the two variance components are similar in magnitude: $\hat{\sigma}_u = 3.02$ and $\hat{\sigma}_\epsilon = 2.98$. This implies that between- and within-subject variance are of roughly equal importance. Otherwise, the results are not dissimilar from those of the regression with cluster standard errors presented in Section 4.5.3.

If the between-subject standard deviation, $\sigma_u$, were equal to zero, this would make the random effects model equivalent to a linear regression model. The fact that the estimate of $\sigma_u$ is large in magnitude is strongly suggestive of the superiority of the random effects model. It is possible to make this comparison formally using a likelihood ratio test. This will be done in the context of multi-level modelling in the next section.

## 4.7 Multi-level Modelling

Multi-level modelling is an extension of the random effects framework that allows for more levels of dependence, and also allows for random slopes as well as random intercepts.

The convention adopted here for counting and ordering model levels is similar to that used by Skrondal & Rabe-Hesketh (2004). A "one-level" model is a straightforward linear regression model with a fixed intercept and fixed slopes. For example, imagine that we have $T$ observations, $y_1 \dots y_T$ on a *single* subject. Then the sample consists of only one cluster, and this is the sense in which there is only one level of clustering. Next, if we have $T$ observations on each of $n$ subjects, $y_{it}, i = 1, \dots, n, t = 1, \dots, T$, then a "two-level" model is appropriate, with the subject indicator $i$ representing the second level of clustering. Next, if the $n$ subjects are divided into $J$ sessions, a typical observation is represented by $y_{ijt}$, and a "three-level" model is appropriate, with the session indicator $j$ representing the third (or "highest") level of clustering.

The three-level model just described is specified as follows.

$$
\begin{aligned}
bid_{ijt} &= \alpha + \beta' x_{it} + \gamma' z_i + u_i + v_j + \epsilon_{ijt} \\
i = 1 \dots, n \quad j &= 1 \dots, J \quad t = 1 \dots, T \\
Var(u_i) &= \sigma_u^2 \\
Var(v_j) &= \sigma_v^2 \\
Var(\epsilon_{it}) &= \sigma_\epsilon^2
\end{aligned} \tag{4.7}
$$

In (4.7) $u_i$ is once again the subject-specific random effect, and the new term $v_j$ is the session-specific random effect.

Next, assume that the slope on one of the explanatory variables varies between subjects. For simplicity, let us assume that there is only one variable contained in $x_{it}$, so that $x_{it}$ and the associated parameter $\beta$ are both scalars. The model is:

$$
\begin{aligned}
bid_{ijt} &= \alpha + \beta x_{it} + \gamma' z_i + u_{0i} + u_{1i} x_{it} + v_j + \epsilon_{ijt} \\
i = 1 \dots, n \quad j &= 1 \dots, J \quad t = 1 \dots, T \\
Var(u_{0i}) &= \sigma_{u0}^2 \\
Var(u_{1i}) &= \sigma_{u1}^2 \\
Var(v_j) &= \sigma_v^2 \\
Var(\epsilon_{it}) &= \sigma_\epsilon^2
\end{aligned} \tag{4.8}
$$

In (4.8) there are two between-subject variance parameters: $\sigma_{u0}$ represents between-subject variation in the intercept; $\sigma_{u1}$ represents between-subject variation in the slope on the variable $x_{it}$.

The STATA command for multi-level modelling is `xtmixed`. We will now demonstrate the use of this command in various ways, using the same set of explanatory variables as used in previous sections.



We first consider what would result if the `xtmixed` command contained the list of variables and nothing else. That is:

```
. xtmixed bid x balance rec_period N6 eps24 N6eps24
```

The answer is that this model is the "one-level" model which is identical to the linear regression model, and we would therefore expect coefficient estimates identical to those of the linear regression model as presented in Section 4.5.3.

Next, we introduce clustering at the subject level, giving the "two-level model". Note that this simply requires the addition of "|| $i$ : " at the end of the command line. This model is, of course, equivalent to the random effects model (4.6) and the results below are almost identical to those obtained using the `xtreg` command in Section 4.6. After estimation, the estimates are stored in a vector named "two_level", for later testing.

```
. xtmixed bid x balance rec_period N6 eps24 N6eps24  || i:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  -10909.6
Iteration 1:   log likelihood =  -10909.6

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      4234
Group variable: i                               Number of groups   =       160

                                                Obs per group: min =         1
                                                               avg =      26.5
                                                               max =        30

                                                Wald chi2(6)       =  3.24e+07
Log likelihood =  -10909.6                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         bid |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.000461   .0001759  5688.01   0.000     1.000116    1.000805
     balance |  -.0050284    .003367    -1.49   0.135    -.0116276    .0015708
  rec_period |   2.450794   .2516203     9.74   0.000     1.957627    2.943961
          N6 |  -1.116107   .7315565    -1.53   0.127    -2.549931    .3177177
       eps24 |  -5.181521   .8053375    -6.43   0.000    -6.759953   -3.603088
     N6eps24 |  -2.593384   1.035923    -2.50   0.012    -4.623756   -.5630121
       _cons |  -4.480923   .5807649    -7.72   0.000    -5.619202   -3.342645
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
i: Identity                  |
                   sd(_cons) |   3.142733   .1845878      2.800995    3.526165
-----------------------------+------------------------------------------------
                sd(Residual) |   2.986586   .0330984      2.922413    3.052167
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  2462.09 Prob >= chibar2 = 0.0000

. est store two_level
```

One piece of output from `xtmixed` that is not present in the `xtreg` output is the "LR test vs linear regression". This is a likelihood ratio (LR) test of the linear regression model as a restricted version of the random effects model. It is therefore a test for the presence of between-subject variation. The null hypothesis is that the between-subject variation ($\sigma_u^2$) is zero.

The likelihood ratio test is a testing procedure appropriate for nested models. It is computed as:

$$LR = 2(LogL_u - LogL_r) \tag{4.9}$$

where $LogL_u$ and $LogL_r$ are the maximised log-likelihoods from the unrestricted and restricted models respectively. The concept of the log-likelihood function will be explained in detail in Chapter 6. Under the null hypothesis (that the restricted model is true) the test statistic has a $\chi^2(k)$ distribution where $k$ is the number of restrictions. In the present case, only one restriction is being tested ($\sigma_u^2 = 0$), so the null distribution is $\chi^2(1)$. The very high value of the test statistic (2462.09) and the accompanying p-value of 0.0000 emphatically confirm the importance of between-subject variation, and also confirm the superiority of the random effects model over linear regression.

Next, we progress to the "three-level" model by clustering at the session level as well as the subject level. This requires the addition of "|| *session* :" to the command, but note that this must appear *before* "|| *i* :" because it is the higher level of clustering. Placing these two parts of the command the wrong way around will prevent the command from working. We save the results as "three_level".

```
. xtmixed bid x balance rec_period N6 eps24 N6eps24  || session: || i:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -10900.072
Iteration 1:   log likelihood = -10900.072

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      4234

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
        session |       16        182      264.6        350
              i |      160          1       26.5         30
-----------------------------------------------------------

                                                Wald chi2(6)       =  3.24e+07
Log likelihood = -10900.072                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         bid |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.000463   .0001759  5688.39   0.000     1.000118    1.000808
     balance |  -.0042531   .0033643    -1.26   0.206    -.0108471    .0023408
  rec_period |    2.46885   .2515514     9.81   0.000     1.975819    2.961882
          N6 |  -1.108771   1.248898    -0.89   0.375    -3.556566    1.339024
       eps24 |  -5.204657   1.284162    -4.05   0.000    -7.721568   -2.687747
     N6eps24 |  -2.576869   1.766896    -1.46   0.145    -6.039922    .8861837
       _cons |  -4.500616   .9152521    -4.92   0.000    -6.294477   -2.706755
------------------------------------------------------------------------------
```



```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
session: Identity            |
                   sd(_cons) |   1.512163   .3639928       .943422    2.42377
-----------------------------+------------------------------------------------
i: Identity                  |
                   sd(_cons) |   2.756427   .1716992      2.439635   3.114356
-----------------------------+------------------------------------------------
                sd(Residual) |   2.986448   .0330933      2.922285   3.052019
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =  2481.15   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. est store three_level
```

We see that adding the session-level random effect term has caused changes in both the coefficients and the standard errors. Some of the conclusions regarding significance also change. We also see that the estimate of the standard deviation of the session-level random effect ($\sigma_v$) is 1.51, and the confidence interval indicates that this estimate is significantly greater than zero.

Again we can use an LR test to confirm this. This time we are testing the two-level model as a restriction on the three-level model. To carry out this test in STATA, we use the `lrtest` command, as follows.

```
. lrtest three_level two_level

Likelihood-ratio test                                 LR chi2(1)  =     19.06
(Assumption: two_level nested in three_level)         Prob > chi2 =    0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```

As indicated by the note following the test result, this result is "conservative" since the value of ($\sigma_v^2$) under the null hypothesis is zero which is indeed on the boundary of the parameter space. We see that the test results in an overwhelming rejection of the null hypothesis, and we therefore conclude that there is overwhelming evidence of the presence of between-session variation, and that the three-level model should be used in preference to the two-level (random effects) model.

As mentioned at the start of this sub-section, the multi-level modelling approach also allows for random slopes. Let us return to the effect of cash balance, discussed in Section 4.5.3. It was suggested that the effect of cash balance on bids might be negative, for reasons of "limited liability", or it might be positive, due to a form of "house money effect". What we have concluded from all of the models that we have estimated so far is that cash balance has no significant effect on bids. However, it is possible that the effect of cash balance varies over the population, with perhaps some individuals displaying a negative effect, and others displaying a positive effect. In order to investigate this possibility, we allow the slope on cash balance to vary between subjects. Where previously we added "|| *i* :" at the end of the command to request a random intercept between subjects, we now add

"|| *i* : *balance*" to request a random slope as well as a random intercept. The results are:

```
. xtmixed bid x balance rec_period N6 eps24 N6eps24 || session: || i: balance

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -10896.454
Iteration 1:   log likelihood = -10896.046
Iteration 2:   log likelihood = -10896.041
Iteration 3:   log likelihood = -10896.041

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      4234

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
        session |       16        182      264.6        350
              i |      160          1       26.5         30
-----------------------------------------------------------

                                                Wald chi2(6)       =  3.25e+07
Log likelihood = -10896.041                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         bid |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.000464   .0001756  5697.79   0.000      1.00012    1.000808
     balance |  -.006549    .0043847    -1.49   0.135    -.0151429    .0020449
  rec_period |   2.418536   .2523934     9.58   0.000     1.923854    2.913218
          N6 |  -1.157555   1.243504    -0.93   0.352    -3.594777    1.279668
       eps24 |  -5.239497   1.280784    -4.09   0.000    -7.749787   -2.729207
     N6eps24 |  -2.494244   1.761641    -1.42   0.157    -5.946997    .9585088
       _cons |  -4.415022   .9132111    -4.83   0.000    -6.204883   -2.625161
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
session: Identity            |
                   sd(_cons) |   1.506631   .3628689      .9397166    2.415555
-----------------------------+------------------------------------------------
i: Independent               |
                sd(balance)  |   .0180286   .0050008      .0104677    .0310506
                  sd(_cons)  |   2.714525   .1734982      2.394912    3.076792
-----------------------------+------------------------------------------------
               sd(Residual)  |   2.975482   .0331918      2.911134    3.041253
------------------------------------------------------------------------------
LR test vs. linear regression:      chi2(3) =  2489.21   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. est store random_slope
```

The standard deviation of the random slope, which is represented by $\sigma_{u1}$ in (4.8), is estimated to be 0.018 with a standard error of 0.005. Once again we can use an LR test to test its significance. This time the test compares the three-level model without the random slope ("three_level") against the three-level model with the random slope ("random_slope").

```
. lrtest random_slope three_level

Likelihood-ratio test                                 LR chi2(1)  =      8.06
(Assumption: three_level nested in random_slope)      Prob > chi2 =    0.0045
```



The p-value of the LR test is less than 0.01, suggesting strong evidence that the slope on the cash balance indeed varies between subjects. Since the point estimate of the effect of cash balance is close to zero (−0.0065) we are led to conclude that the population divides fairly evenly between those whose bids depend negatively on cash balance (due to limited liability), and those whose bids depend positively on them (due to a house money effect).

This is a good example of between-subject heterogeneity which is one of the dominant themes of this book. Subjects do respond to changes in their cash balance, but they vary in their responsiveness in such a way that when this heterogeneity is ignored, the effect of cash balance is erroneously estimated to be close to zero.

## 4.8 Modelling Data from Contest Experiments

Partly for the sake of providing a second example of theory testing, in this section we consider the analysis of data from contest experiments.

Contests are similar to auctions in some ways. The essential difference is that in an auction, only the highest bidder can win, while in a contest, any player who bids a positive amount can win. Also, in most auction types (with the exception of the "all-pay auction"), only the winner pays; in contests, all bidders pay.

### *4.8.1 The Tullock contest*

Perhaps the most popular contest model is the "Tullock contest" proposed by Buchanan et al. (1980). It is sometimes referred to as a "rent-seeking contest". In such a contest, there are $n$ players competing for a prize of $v$. Each player $i$ invests an *effort* $e_i$, and player $i$'s probability of winning the prize is defined by the *contest success function* (CSF):

$$p_i(e_i, e_{-i}) = \frac{e_i}{\sum_{j=1}^{n} e_j} \tag{4.10}$$

Given CSF (4.10), the expected payoff for player $i$ is

$$E(\pi_i(e_i, e_{-i})) = p_i(e_i, e_{-i})v - c(e_i) \tag{4.11}$$

where $c(e)$ is the cost of applying effort level $e$. Assuming that the $n$ players are risk neutral and identical to each other, it can easily be shown that the Nash equilibrium effort level for each player ($e^*$) is given by the solution to the following equation:

$$c'(e^*)e^* = \frac{(n-1)}{n^2}v \tag{4.12}$$

In the normal situation in which costs are linear, that is $c(e) = e$, (4.12) simplifies to:

$$e^* = \frac{(n-1)}{n^2}v \tag{4.13}$$

A regular finding in contest experiments is that subjects provide effort levels that are systematically higher than the Nash equilibrium prediction (4.13). That is, as in auction experiments, we observe the phenomenon of "over-bidding".

### 4.8.2 A contest experiment

A survey of experimental findings on contests has been provided by Sheremeta (2013). Here, we analyse data from a particular contest experiment, conducted by Chowdhury et al. (2014).

The experiment consists of 12 sessions. Each session consists of 12 subjects, each participating in 30 contests. Subjects were matched into groups of $n = 4$, with random rematching after each contest. The value of the prize in all contests was $v = 80$. In each contest, subjects simultaneously selected an effort level between 0 and 80. Applying (4.12) we find that the equilibrium prediction for effort in this situation is $e^* = 15$.

There are two treatments. The first is linear (L) versus convex (C) costs. In sessions employing the linear cost function, the cost of effort was $c(e) = e$ as is standard. In sessions with the convex cost function, the cost of effort was $c(e) = \frac{e^2}{30}$. Because this cost function is such that $c'(15) = 1$, the equilibrium prediction for effort is the same as it is in the linear cost case, that is, 15. The purpose of this treatment is to investigate the extent to which over-bidding is a consequence of the flatness of the payoff function, as suggested (in the context of private value auctions) by Harrison (1989). If this is indeed a reason for over-bidding, we would expect bidding to be lower under the convex costs treatment since the payoff function is steeper under this treatment.

The second treatment is a probabilistic (P) versus share (S) rule for awarding the prize. The probabilistic treatment (P) is the standard situation in which there is a single indivisible prize with winning probabilities given by the CSF (4.10). Under the share treatment (S), the prize is divided between the contestants with shares determined exactly by the CSF (4.10). Again, this treatment does not change the equilibrium prediction of 15. The purpose of this treatment is to investigate the extent to which over-bidding is caused by a non-monetary utility for winning the contest (or "joy of winning"), as suggested by Sheremeta (2010), or by distorted perceptions of probabilities, as suggested by Goeree et al. (2002). If either of these explanations for over-bidding are valid, we would expect lower bidding under the share treatment than under the probabilistic treatment. This is because, under the share treatment, there is no clear winner, and there are no probabilities to be distorted.

Information on the treatments is summarised in the following table:

| Sessions | # subjects | $n$ | P/S | L/C | RNNE bid | Observations |
|---|---|---|---|---|---|---|
| 1–3 | 12 | 4 | P | C | 15 | 1080 |
| 4–6 | 12 | 4 | P | L | 15 | 1080 |
| 7–9 | 12 | 4 | S | C | 15 | 1080 |
| 10–12 | 12 | 4 | S | L | 15 | 1080 |



Note that it is a $2 \times 2$ full-factorial design, with all possible combinations of the two treatments being covered. This means that it will be possible to estimate the interaction effect between the two treatments in addition to the main effects.

### 4.8.3 Analysis of data from contest experiment

The data of Chowdhury et al. (2014) are contained in the file **chowdhury**. To give a feel for the data, we present a histogram of effort levels, for the 1,080 observations in the baseline treatment (PL), in Figure 4.4. It is seen that there is a large variation in effort levels, covering the entire permissible range of the variable. We also see that the distribution is multi-modal, with accumulation of effort levels at multiples of 5. Most importantly, we see a tendency for effort to exceed the Nash equilibrium prediction of 15 (i.e. a tendency for over-bidding). The mean effort level over the 1,080 observations in the baseline treatment is 26.2, implying that the mean over-bidding rate is 75%.

The analysis performed here is similar to that of Chowdhury et al. (2014). We define a variable $o_{it}$ to be the "over-bid", that is, the excess effort relative to the Nash equilibrium prediction, by subject $i$ in round $t$. This is simply:

$$o_{it} = e_{it} - 15 \tag{4.14}$$

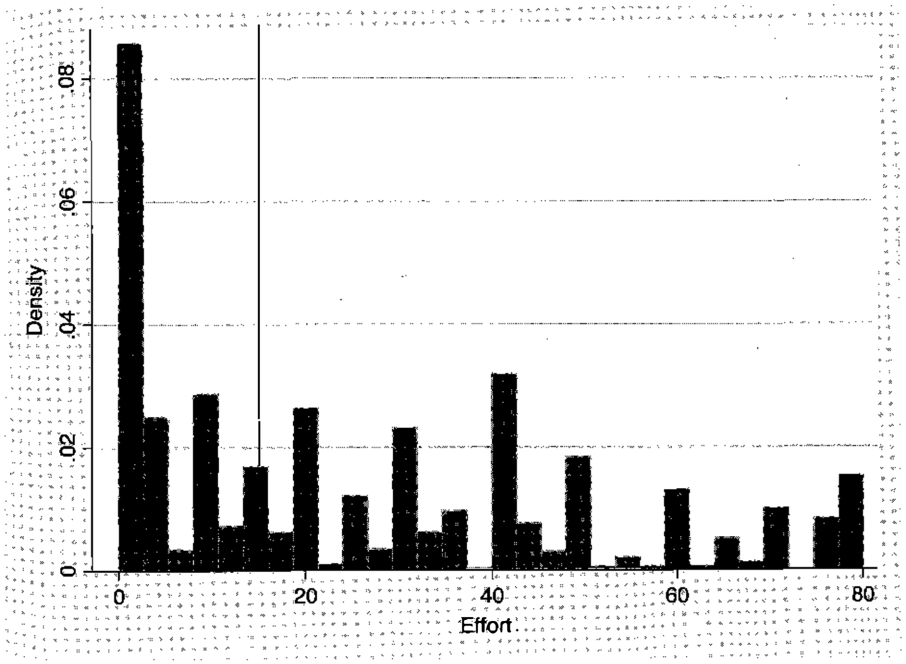



Figure 4.4: Distribution of effort levels in the experiment of Chowdhury et al.'s (2014). All 1,080 observations in the PL treatment. Vertical line drawn at Nash equilibrium prediction.

We then consider the following random effects model with $o_{it}$ as the dependent variable:

$$o_{it} = \beta_0 + \beta_1 S_i + \beta_2 C_i + \beta_3 S_i \star C_i + \beta_4(1/t) + u_i + \epsilon_{it} \tag{4.15}$$

$S_i$ and $C_i$ are treatment dummies for share and convex respectively, and $S_i \star C_i$ is the interaction variable formed as the product of the two. The intercept parameter $\beta_0$ has a clear interpretation as the expected level of over-bidding in the baseline treatment (PL) with experience (i.e. when $t$ is large).

Chowdhury et al. (2014) perform estimation separately by treatment, and also separately for the first 15 and last 15 rounds. The reason for estimating separately for the first 15 and last 15 rounds is that the (negative) effect of experience on effort appears to diminish markedly over the 30 rounds. They provide eight sets of estimation results in their Table 3. To reproduce their first set of results, we would use the following STATA code:

```
gen o = bid - 15
xtset i t
xtreg o s t if (c==1)&(t<=15)
```

When the command `xtreg` is used with no options, the random effects model is estimated as the default.

Here, we shall instead estimate model (4.15) using all data. By using all data, we are able to include both treatments together, both with and without an interaction. Note also that the use of the reciprocal of $t$ instead of $t$ as an explanatory variable in (4.6) is a means of capturing the diminishing effect of experience uncovered by Chowdhury et al. (2014). The code we use is therefore:

```
gen o = bid - 15
xtset i t
gen sc=s*c
gen rec_t=1/t
xtreg o s c rec_t
xtreg o s c sc rec_t
```

The results are:

```
. *Random effects model without interaction:
. xtreg o s c  rec_t

Random-effects GLS regression                   Number of obs      =      4320
Group variable: i                               Number of groups   =       144

R-sq:  within  = 0.0223                         Obs per group: min =        30
       between = 0.0883                                        avg =      30.0
       overall = 0.0497                                        max =        30

                                                Wald chi2(3)       =    108.72
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           o |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |  -7.234259   1.986886    -3.64   0.000    -11.12848   -3.340034
           c |  -1.249074   1.986886    -0.63   0.530      -5.1433    2.645151
       rec_t |   11.51388   1.180853     9.75   0.000     9.199449    13.82831
       _cons |   11.64993   1.727864     6.74   0.000     8.263383    15.03649
-------------+----------------------------------------------------------------
```



```
     sigma_u |  11.614135
     sigma_e |  14.727186
         rho |  .38344663   (fraction of variance due to u_i)
------------------------------------------------------------------------------

.
. *Random effects model with interaction:
. xtreg o s c sc rec_t

Random-effects GLS regression                   Number of obs      =      4320
Group variable: i                               Number of groups   =       144

R-sq:  within  = 0.0223                         Obs per group: min =        30
       between = 0.1150                                        avg =      30.0
       overall = 0.0608                                        max =        30

                                                Wald chi2(4)       =    113.26
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           o |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           s |  -3.194537   2.778254    -1.15   0.250    -8.639815    2.250741
           c |   2.790648   2.778254     1.00   0.315     -2.65463    8.235926
          sc |  -8.079444   3.929045    -2.06   0.040    -15.78023   -.3786583
       rec_t |   11.51388   1.180853     9.75   0.000     9.199449    13.82831
       _cons |   9.630073   1.970806     4.89   0.000     5.767365    13.49278
-------------+----------------------------------------------------------------
     sigma_u |  11.476361
     sigma_e |  14.727186
         rho |  .37781999   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Let us first interpret the results of the model without the interaction. The intercept is estimated to be +11.6 and the estimate is strongly significant. This simply tells us that in the baseline treatment (PL) with experience, subjects tend on average to over-bid by 11.6. This amounts to a rejection of the fundamental prediction of the theory.

We also see that the effect of experience is highly important. Effort declines rapidly but at a decreasing rate, as evidenced by the significantly positive coefficient on the reciprocal of the round number. Subjects appear to move in the direction of the equilibrium, but they converge to a point some distance short of it. In fact, the closeness of the coefficient of $1/t$ to the estimate of the intercept tells us that in the course of the experiment, subjects converge to a point roughly half way between the starting point and the equilibrium.

Turning to the treatment effects, we first see that share treatment significantly reduces over-bidding. This result is consistent with the "joy of winning" hypothesis; under the share rule, there is no clear winner, and hence any "joy of winning" motivation must be reduced. The result could also be explained in terms of the probability distortion hypothesis. The convex cost treatment appears to have no effect in the first model.

The second model includes the interaction variable "$s \star c$". We see that this variable has a significantly negative coefficient, indicating that the effect of the share treatment is stronger in the presence of convex costs.

Chowdhury et al. (2014) suggest that these findings regarding the drivers of out-of-equilibrium play are useful for the robust design of contests, and in particular that the interactions between these drivers may be important.

## 4.9 Meta-analysis

Meta-analysis is the name given to the set of methods for combining results from different studies, with the objective of identifying stronger patterns than can be seen in individual studies. In this section, we demonstrate that certain interesting experimental results (including comparative static predictions and effects of design features) can be confirmed (or refuted) using meta-analysis. We continue with the theme of experimental contests.

In the last section, data from a contest experiment conducted by Chowdhury et al. (2014) was analysed, and it was noted that the mean over-bidding rate in their baseline treatment was 75%. Sheremeta (2013) has collected a set of 39 such over-bidding rates (including that one) from experiments reported in 30 published articles. Along with the over-bidding rates, a number of features of each experiment are recorded, including number of players, prize, endowment, and matching protocol. The data set is presented in Table 1 of Sheremeta (2013), and it is reproduced in the file **sheremeta**. The first 20 rows of the data set are presented in Figure 4.5.

An obvious issue arising when meta-analysis is applied to contest results is that studies are not directly comparable in the sense that they use different units of measurement for the endowment and prize. To make the studies comparable, we therefore define the variable *endowment_rel* (relative endowment) as endowment divided by prize. Another issue is that, as seen in Figure 4.5, the data has an element of clustering, with some articles generating more than one row of the data. While this feature of meta-data may be very important in some situations, adjusting for clustering (using the techniques described earlier in the current chapter) makes very

| | obs | study | author | year | treatment | matching | endowment |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | millner_pratt | 1989 | lottery | random | 12 |
| 2 | 2 | 2 | millner_pratt | 1991 | less RA | random | 12 |
| 3 | 3 | 3 | shogren_baik | 1991 | lottery | fixed | 24 |
| 4 | 4 | 4 | davis_reilly | 1998 | lottery | fixed | . |
| 5 | 5 | 5 | potters_etal | 1998 | lottery | random | 15 |
| 6 | 6 | 6 | anderson_stafford | 2003 | homogeneous | one-shot | 5 |
| 7 | 7 | 6 | | . | | one-shot | 5 |
| 8 | 8 | 6 | | . | | one-shot | 5 |
| 9 | 9 | 6 | | . | | one-shot | 5 |
| 10 | 10 | 6 | | . | | one-shot | 5 |
| 11 | 11 | 7 | schmitt_etal | 2004 | static | random | 150 |
| 12 | 12 | 8 | schmitt_etal | 2005 | single-prize | one-shot | 20 |
| 13 | 13 | 9 | herrmann_orzen | 2008 | direct repeated | random | 16 |
| 14 | 14 | 10 | kong | 2008 | less RA | fixed | 300 |
| 15 | 15 | 11 | fonseca | 2009 | simultaneous | random | 300 |
| 16 | 16 | 12 | abbink_etal | 2010 | one:one | fixed | 1000 |
| 17 | 17 | 13 | sheremeta | 2010 | one-stage | random | 120 |
| 18 | 18 | 14 | sheremeta_zhang | 2010 | individual | random | 120 |
| 19 | 19 | 15 | ahn_etal | 2011 | individual | fixed | . |
| 20 | 20 | 16 | deck_jahedi | 2011 | baseline | one-shot | 5 |

Figure 4.5: First 21 rows of the data set of Sheremeta's (2013)



little difference to the analysis that follows. For this reason the clustering will be disregarded.

Equation (2) of Sheremeta (2013) presents the results of a linear regression with the over-bidding rate as the dependent variable and four explanatory variables: relative endowment; number of players (n); partners matching dummy; and one-shot dummy (the strangers matching dummy is excluded and represents the base case). The results are exactly reproduced as follows.

```
. reg overbid endowment_rel n partners one_shot

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  4,    34) =    8.73
       Model |  6.49674324     4  1.62418581           Prob > F      =  0.0001
    Residual |  6.32637945    34  .186069984           R-squared     =  0.5066
-------------+------------------------------           Adj R-squared =  0.4486
       Total |  12.8231227    38  .337450597           Root MSE      =  .43136

------------------------------------------------------------------------------
     overbid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
endowment_rel |    .431265   .2061099     2.09   0.044     .0123993    .8501307
           n |   .2036022   .0414801     4.91   0.000     .1193046    .2878999
    partners |  -.0778284   .1690991    -0.46   0.648    -.4214791    .2658223
    one_shot |   .2929277   .1984659     1.48   0.149    -.1104035    .6962588
       _cons |  -.4108494   .2713689    -1.51   0.139    -.9623374    .1406386
------------------------------------------------------------------------------
```

The $R^2$ of the regression is slightly more than 0.50, indicating that more than half of the variation in over-bidding rates is being explained by this simple model. The results also show that the over-bidding rate rises with the relative endowment ($p < 0.05$) and also rises with the number of players ($p < 0.01$). The two other coefficients indicate that the matching protocol is not important in explaining over-bidding rates. It is conceivable that "partners" matching might result in lower effort as a consequence of collusion between players, but this effect is not seen in this data. Baik et al. (2014) have already found evidence of this "non-result" in the context of a single experiment in which matching protocol varies between treatments. They stress the usefulness of the "non-result" in the sense that it implies that partners matching can reasonably be used in preference to strangers matching, with its attendant advantages including the greater number of independent observations for a given total number of subjects.

Let us investigate the effect of relative endowment more closely. In Figure 4.6, we present a scatter plot of over-bidding rates against relative endowment. Superimposed on the scatter plot is a "Lowess smoother", a form of non-parametric regression first developed by Cleveland (1979). This indicates that the relationship between endowment and effort may be non-linear, and in particular that there may be an "optimal" (from the viewpoint of the contest organiser at least!) relative endowment, around one, at which effort is maximised. Baik et al. (2014) have already found evidence of this result in the context of a single experiment in which endowment is a treatment variable. They attribute this result to the endowment acting as a constraint when it is low, but generating a wealth effect when it is larger. Wealth is hypothesised to bring about a reduction in "conflict intensity", that is, a reduction in effort.

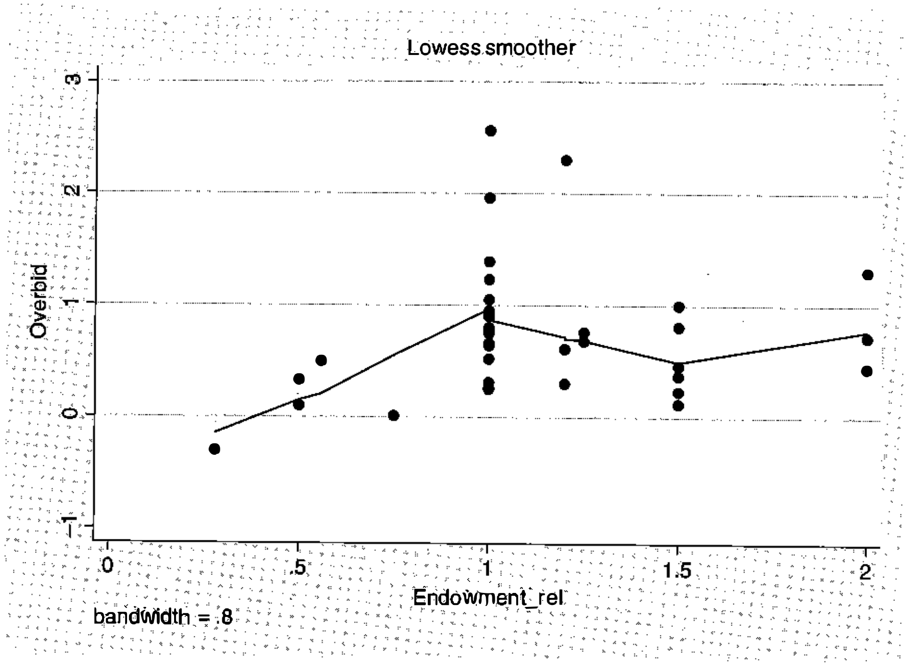



Figure 4.6: Scatterplot and Lowess smoother of over-bidding rate against relative endowment

The non-linear effect of endowment may be tested in the context of the meta-analysis by including the square of relative endowment (`end2` in the data set) as an explanatory variable. The results are as follows:

```
. reg  overbid endowment_rel end2 n partners one_shot

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  5,    33) =    9.86
       Model |  7.68145525     5  1.53629105           Prob > F      =  0.0000
    Residual |  5.14166744    33  .155808104           R-squared     =  0.5990
-------------+------------------------------           Adj R-squared =  0.5383
       Total |  12.8231227    38  .337450597           Root MSE      =  .39473

------------------------------------------------------------------------------
     overbid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
endowment_rel |    2.37303    .729003     3.26   0.003     .8898622    3.856198
        end2 |  -.8146338   .2954276    -2.76   0.009    -1.415686   -.2135818
           n |   .1986473   .0379999     5.23   0.000     .1213359    .2759587
    partners |   .0185187   .1586342     0.12   0.908     -.304225    .3412624
    one_shot |   .3408628   .1824413     1.87   0.071    -.0303169    .7120424
       _cons |  -1.471736   .4579111    -3.21   0.003    -2.403363   -.540109
------------------------------------------------------------------------------
```

We see that the coefficients of relative endowments and its square are respectively strongly positive and strongly negative, implying that there is indeed a value of relative endowment at which effort is maximised, confirming the result of Baik et al. (2014). It is easily verified that this optimal level can be computed as $-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$, where



$\hat{\beta}_1$ and $\hat{\beta}_2$ are respectively the coefficients on relative endowment and its square. This computation can be performed in STATA using a technique known as the *delta method*, which will be explained in detail in Chapter 6. The STATA command is `nlcom` and this is required immediately after the `regress` command. The command along with the results are as follows:

```
. nlcom end_star: -_b[endowment_rel]/(2*_b[end2])

    end_star:  -_b[endowment_rel]/(2*_b[end2])

------------------------------------------------------------------------------
     overbid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    end_star |   1.456501   .1503841     9.69   0.000     1.150542     1.76246
------------------------------------------------------------------------------
```

We see that the estimate of the "optimal endowment" is 1.46. A major attraction of using the delta method is that it also returns a standard error and 95% confidence interval. The confidence interval is [1.15, 1.76] which conveys evidence that the optimal relative endowment is somewhat greater than one.

## 4.10 Summary and Further Reading

In this chapter, we have tried to provide an overview of the types of methods used in the analysis of data from experiments in which subjects are required to submit bids. This set of techniques is applicable over a wide range of experimental settings. The settings on which we have focussed for illustrative purposes are auction experiments and contest experiments.

Readers wishing to learn about auction theory are referred to Krishna (2010). Studies in which auction data is modelled using the type of techniques introduced in this chapter include Kagel et al. (1995), Kagel & Levin (1986) and Ham et al. (2005).

Clustering has been a recurring theme. For a useful recent dicussion of the importance of clustering in economic experiments, see Fréchette (2012). For panel data estimation techniques, readers are referred to Baltagi (2008), and for multi-level modelling, they are referred to Rabe-Hesketh & Skrondal (2008).

Drichoutis et al. (2011) apply a dynamic panel estimation procedure to allow for bid interdependence (between rounds) in auctions. This sort of estimation procedure allows (for example) the bid in the previous round to influence the bid in the current round. Readers interested in dynamic panel data models should consult Roodman (2009). These estimation techniques have not been covered in this chapter, although the topic of dynamic panel estimation will be covered briefly in Section 9.3.3.

The technique of meta-analysis was briefly introduced in this chapter, and the meta-analytic results of Sheremeta (2013) in the context of contest experiments were reproduced and extended. Other meta-analyses that have been published in the experimental economics literature include Zelmer (2003) (public goods games), Engel (2011) (dictator games), and Johnson & Mislin (2011) (trust games).

## Exercises

1. Following each of the regression models estimated in Sections 4.5 and 4.6, the RNNE theory was tested using Wald tests. Conduct tests of the same hypothesis using LR tests. Do the results agree?
2. Consider the expected payoff for player $i$ in a Tullock contest with $n$ players, as given in (4.11). Substitute the contest success function (4.10) into (4.11), differentiate with respect to $e_i$, set to zero, and finally assume that all $n$ players invest the same effort, in order to find the Nash equilibrium effort level defined by (4.12).

# Chapter 5
# Modelling of Decision Times Using Regression Analysis

## 5.1 Introduction

In this chapter we consider the application of regression analysis to a topic that is currently growing in importance: the modelling of decision times. The decision time applied by a subject to a given task is often measured electronically with great accuracy, and is a reliable measure of the effort expended in performing the task. This sort of analysis is useful for a number of reasons, most importantly that it enables us to identify the features of a task that tend to increase the effort expended by subjects.

The decision-time example is also useful in the further demonstration of panel data estimators and in the highlighting of differences between estimation techniques. The example we use is particularly useful because all of the explanatory variables appearing in the model are time-varying. This means that, unlike with the auctions data in Chapter 4, the fixed-effects estimator can be used, and the Hausman test may be demonstrated as a test for adjudicating between fixed and random effects.

The example we consider in this chapter is decision times in a (simulated) risky choice experiment: linear regression models are used to identify the determinants of the length of time taken to choose between two lotteries. The results will be interpreted in terms of subjects' allocation of cognitive effort. A similar analysis has been performed by Moffatt (2005*b*). Camerer & Hogarth's (1999) "capital-labour-production" framework is relevant and useful here. "Capital" is the knowledge, skills, and experience which the subject brings to the experiment, and also includes the experience acquired during the experiment. "Labour" is the mental effort exerted while solving a task. "Production" is loosely represented by the subject's performance in the task, and is obviously determined by the levels of capital and labour inputs. Although there are reasons for believing that capital input is fixed at least in the context of a single experiment, labour input is fully flexible and potentially highly responsive to incentives and other factors.

Capital input as defined above is clearly hard to measure accurately. However, we shall see that the effects of knowledge and experience are seen indirectly on more than one occasion in estimation. In contrast, a measure of "labour input" is readily available: the time in seconds taken to make each choice.

A number of factors might be expected to influence labour input (i.e. effort). These will be introduced with the aid of a motivating example in Section 5.2. In Section 5.3, we develop a theoretical model of effort allocation, with a view to testing it in the later empirical analysis. Section 5.4 presents the data and considers estimation of the effort model using linear regression. Section 5.5 progesses to panel data estimation of the same model, and includes the demonstration of the Hausman test applied to adjudicating between random effects and fixed effects. Results are discussed in Section 5.6. Section 5.7 considers post-estimation issues, in particular a method for extracting the "posterior random effect", and why this is useful. Section 5.8 provides a summary.

## 5.2 The Decision-time Data

The data used in this chapter is simulated, and is from the same (simulated) experiment as used in Chapter 13. It is contained in the file **decision_times_sim**. The data is simulated in such a way as to resemble a real data set as closely as possible. This was achieved by basing the simulation, of both choices and decision times, on the results of Moffatt (2005*b*), who analysed the real data set of Hey (2001). Full details of the simulation can be found in Chapter 13.

The simulation is of the lottery choices of an imaginary sample of 60 subjects, each of whom faced a set of 50 choice problems on two different days. The total number of observations in the data set is therefore $60 \times 100 = 6{,}000$. The ordering of the problems changed between sessions and also between subjects. The probabilities defining the 50 choice problems are listed in Appendix C. All 50 problems involve the three outcomes: \$0, \$10, and \$20. We imagine that the random lottery incentive system was applied: at the end of the second session, one of the subject's 100 chosen lotteries was selected at random and played for real.

We assume that, for each choice made, decision time was electronically recorded. There are therefore 6,000 decision times in the sample. The penultimate column of the table in Appendix C shows the mean decision time (in seconds) for each choice problem.

Direct motivation for the analysis of decision times reported later is provided by briefly previewing two examples of risky choice problems from the experiment. The selected problems are presented diagrammatically, as they would be presented to subjects during the experiment (if it were real), in Figure 5.1. The mean of decision time is taken over the 120 observed decisions for each problem, and is shown in the figure.

The key feature of the example is that "Task 40" takes subjects almost three times as long to solve, on average, as "Task 14". There are many possible explanations for the difference. The most obvious is differing complexity: the second problem is clearly more complex than the first. This is not the most interesting determinant of effort, but it is clearly a factor that needs to be controlled for. An important issue to be addressed later is therefore how best to measure complexity. A second possible explanation of the difference in decision times is the difference



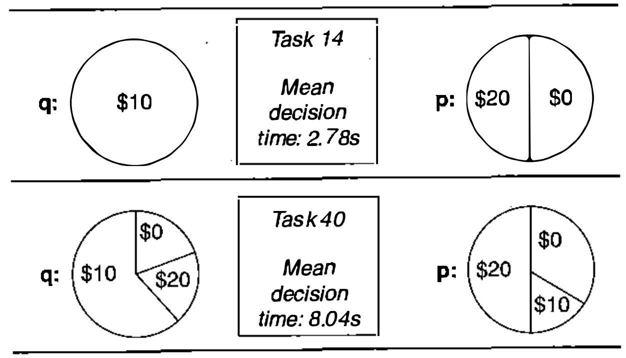



Figure 5.1: Typical choice tasks in the risky-choice experiment

in financial incentives: the expected values of the lotteries are higher in the second problem. A third, and more subtle, explanation for the difference is the possibility that subjects tend to be closer to indifference when faced with the second problem, than with the first. In order to investigate this possibility, we need to establish a framework for measuring closeness to indifference for each subject and for each problem. This framework is the parametric choice model developed and estimated in Chapter 13. A variable representing closeness to indifference is obtained using the method described in Section 13.4.5. This variable is used in the modelling in this chapter.

## 5.3 A Theoretical Model of Effort Allocation

In this section, we adopt a framework introduced by Moffatt (2005*b*) to analyse the effects of two of the factors that we expect to be important in the allocation of effort: the subject's closeness to indifference between the two lotteries; and the objective similarity between the two lotteries. It is important to realise that these two factors are not equivalent. If two lotteries are identical, then necessarily a subject will be indifferent between them, but the converse does not apply: indifference does not imply identicality. In order to isolate the effect of closeness to indifference, we therefore need to control for the effect of objective similarity.

As noted in the previous section, all choice problems in the experiment involve combinations of the three outcomes: \$0, \$10, \$20. We index the problems in the experiment by $t$ ($t = 1, \ldots, 100$). For most choice problems, one of the two lotteries may be classified as the "riskier" lottery, and the other as the "safer" lottery. When this is not possible, the problem is one of "dominance", since one lottery first-order stochastically dominates the other.[1] If task $t$ is a non-dominance problem, we will label the riskier lottery as $\mathbf{p}_t$, and the safer as $\mathbf{q}_t$. For a dominance problem, $\mathbf{p}_t$

[1] See Chapter 12 for definitions of these terms.

will be the dominating lottery, and $\mathbf{q}_t$ the dominated. $\mathbf{p}_t = (p_{1t}\ p_{2t}\ p_{3t})$ and $\mathbf{q}_t = (q_{1t}\ q_{2t}\ q_{3t})$ are vectors containing three probabilities corresponding to the three possible outcomes.

For closeness to indifference, we shall use the absolute valuation differential for subject $i$ in problem $t$, $|\hat{\Delta}_{it}|$. This variable is explained fully, and generated, in Chapter 13. In short, it is a non-negative variable which takes the value zero if subject $i$ is completely indifferent between the two lotteries in problem $t$, and takes a large positive value if the subject has a clear preference for one of the lotteries.

For objective similarity (of the lotteries in choice problem $t$) we shall use the following measure:

$$\Delta_t^o = \sum_{j=1}^{3} (q_{jt} - p_{jt})^2 \tag{5.1}$$

Note that $\Delta_t^o = 0$ for a problem in which the two lotteries $\mathbf{p}_t$ and $\mathbf{q}_t$ are identical, while $\Delta_t^o$ reaches a maximum of 2 if the two lotteries are certainties of different amounts.

Figure 5.2 shows a graph with a particular subject's absolute valuation differential on the horizontal, and objective difference on the vertical. First, we note that the feasible region is the triangle OAC. This is because when the two lotteries are identical, the subject is necessarily indifferent, so we must be at the origin. Also, when the objective difference is maximal, that is, when the two lotteries are certainties of different amounts, it is not possible for the subjective valuation differential to be zero, hence the point B is outside the feasible region. The upward sloping lines

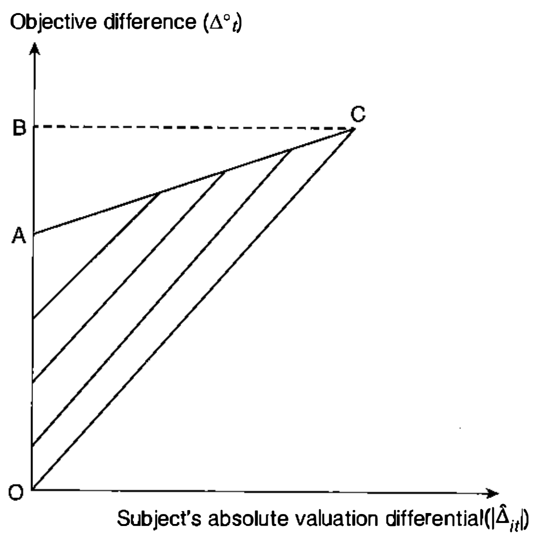



Figure 5.2: Iso-effort curves in $\left(|\hat{\Delta}_{it}|, \Delta_t^o\right)$ space



inside the triangle, including OC, can be interpreted as *iso-effort* curves. Along OC, effort is minimised, since here the subject's preference is as clear as is allowed for a pair of lotteries which differ objectively by a given amount. The other *iso-effort* curves represent higher levels of effort, and the highest effort is allocated at point A, where although the lotteries are objectively very different, the subject in question is indifferent between them.

The important predictions arising from the simple analysis depicted in Figure 5.2 are that, *ceteris paribus*, we expect effort to increase with the objective difference between the lotteries, but to decrease with the subject's absolute valuation differential. The econometric models of effort allocation reported in Sections 5.4–5.6 confirm these predictions, albeit on simulated data.

## 5.4 An Econometric Model of Effort Allocation

As mentioned in Section 5.1, our measure of "labour input", or effort expended in solving a problem, is simply the time taken to make a decision. The logarithm of this variable will be the dependent variable in the model we will estimate in this section.

To give a feel for the data, summary statistics of pooled decision times are shown in Table 5.1, and a histogram of the same variable is shown in Figure 5.3. We see that a typical response time is between 0 and 10 seconds, although there is a very long tail to the right. This calls for the use of the logarithmic transformation of the variable in econometric modelling. Multiplying the mean by 50, we see that a typical subject would spend around five minutes on each of the two days engaging in the experiment.

Like many experimental data sets, this one contains repeated observations for each subject. This should be taken into account in modelling. As explained in Chapter 4, it is natural to handle such data sets in a panel data framework, in which it is assumed that, say, $n$ subjects are observed making a decision in each of $T$ time periods. Clearly, subjects are expected to differ from each other. In the present context, it is expected that some subjects are by nature quick decision makers, and that others are slow. This clearly implies that, if we are treating decision time as the dependent variable in a regression model, we anticipate dependence at the level of the subject, and we need to deal with this dependence using the methods demonstrated in a different context in Chapter 4.

A very useful way of viewing panel data graphically is using the `xtline` command in STATA. First, the `xtset` command is used to declare the data set as a panel.

| Variable | n | mean | median | s.d. | min | max |
|---|---|---|---|---|---|---|
| Decision time (seconds) | 6000 | 5.098 | 3.808 | 4.6235 | 0.231 | 73.02 |

Table 5.1: Summary statistics of decision time in seconds

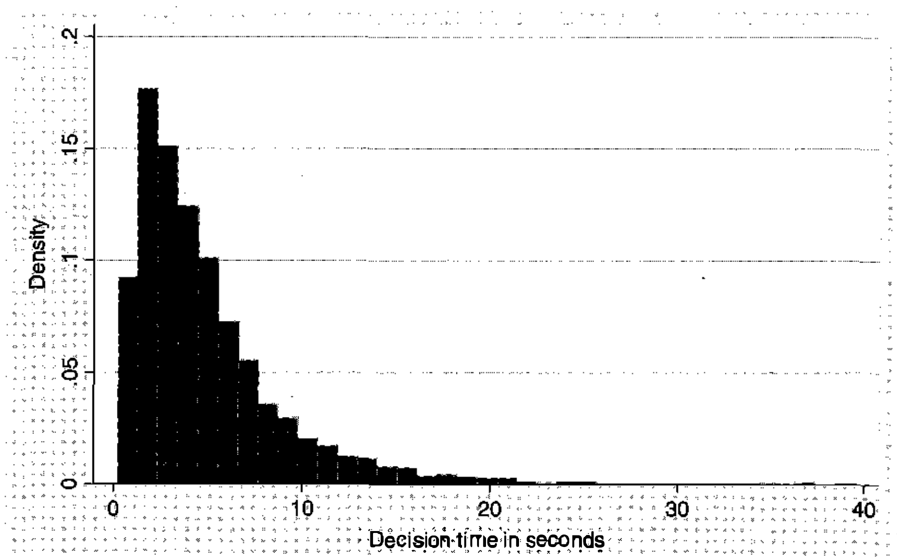



Figure 5.3: Histogram of decision time

Then we apply the `xtline` command to decision time, excluding the small number of observations for which decision time is greater than 20 seconds.

```
. xtset i t
       panel variable:  i (strongly balanced)
        time variable:  t, 1 to 100
                delta:  1 unit

. xtline dt if dt<20
```

The result is shown in Figure 5.4. We see that the `xtline` command produces a time series plot of the decision time variable for each of the 60 subjects in the sample. The resulting graph is useful for assessing the extent of between-subject variability. For some subjects, the time series is persistently very close to zero, indicating fast decision times. Subject 24 is in fact the fastest decision maker, with a mean decision time of 1.558 seconds.[2] For other subjects, decision time is persistently high, indicating slow decision making. The slowest decision maker appears to be subject 29, with a mean decision time of 9.947 seconds.

Next, we shall use graphical analysis to identify the determinants of decision time. In Figure 5.5 the logarithm of decision time is plotted against a variable representing the position of the problem in the sequence of 100 problems. Since it is hard to discern a relationship from the scatter alone, a non-parametric regression (Lowess; see Cleveland, 1979), sometimes called a "smoother", has been

[2] The mean of decision time for individual subjects are computed using the command:

```
table i, contents (mean dt)
```



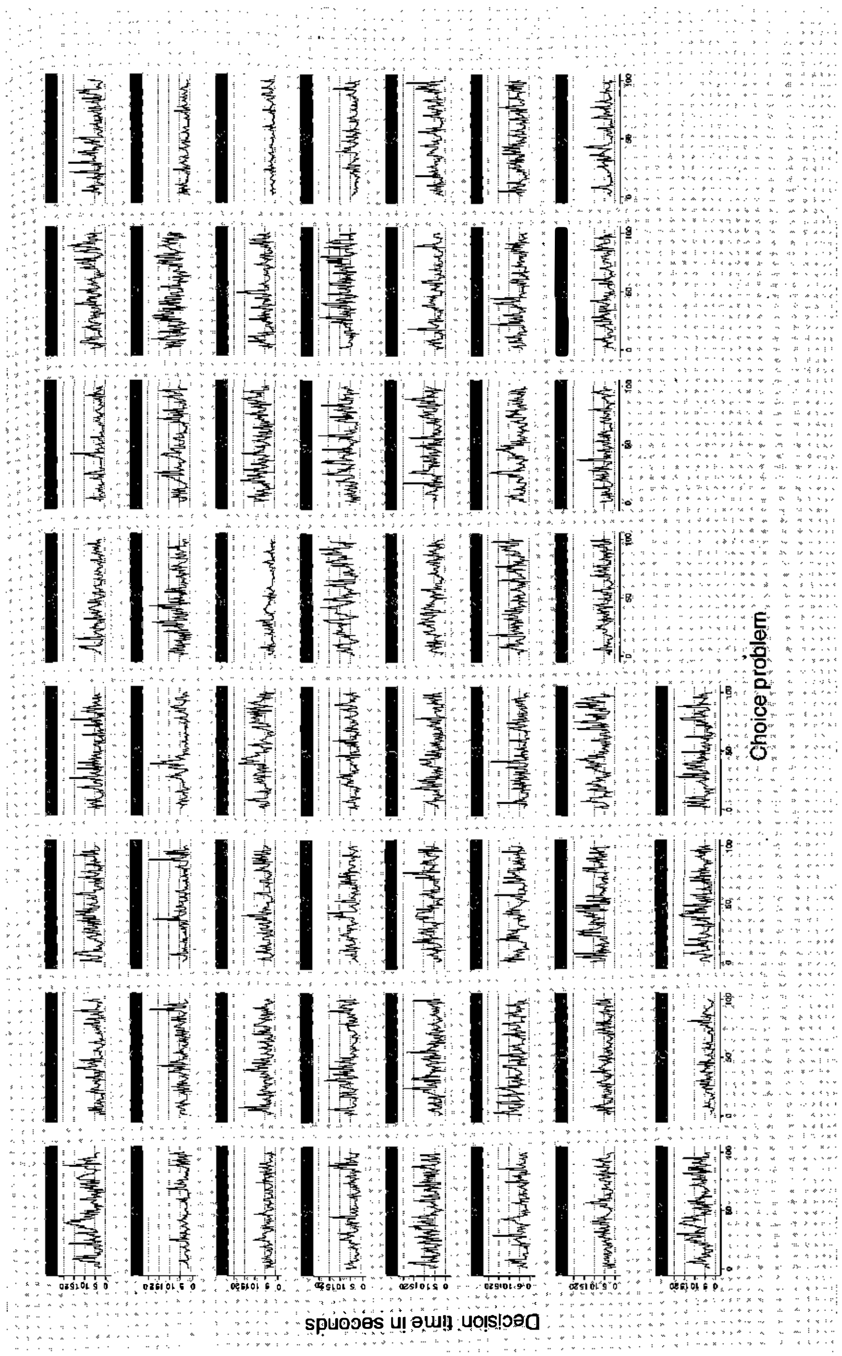



Figure 5.4: Time series graphs of decision time (in seconds) by subject

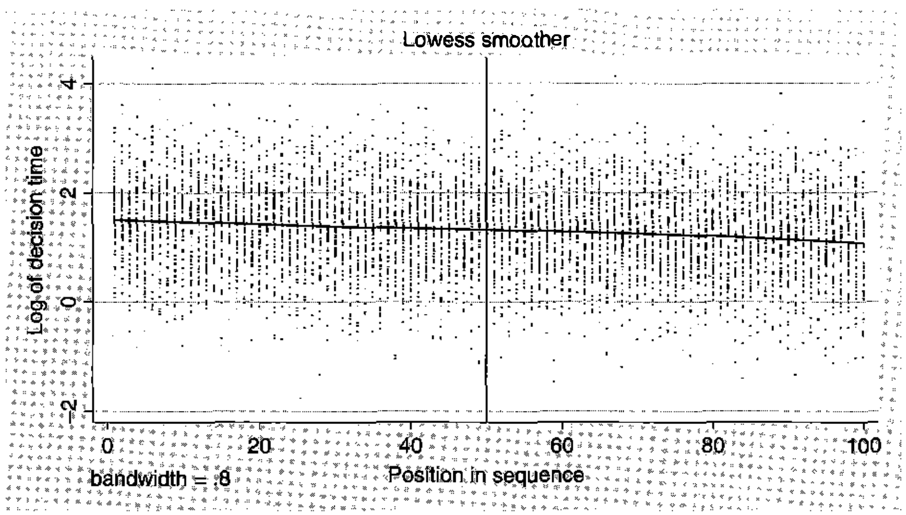



Figure 5.5: Logarithm of decision time against position in the sequence

superimposed. This smoother clearly reveals a tendency for decision times to diminish with the progress of the experiment. Also, remembering that 50 problems were solved on each day, a within-day reduction in decision times is also discernible in Figure 5.5 (this is seen most easily in the extremes of the distribution). While the overall decline over the experiment may be attributed to the accumulation of experience, any within-day decline is more likely to be the result of boredom. These two effects are captured separately in the model estimated in the next section.

In Figure 5.6 decision time in seconds is plotted against the absolute valuation differential (closeness-to-indifference), again with a smoother. The smoother clearly reveals that more effort is allocated to problems for which the subject is close to indifference. This relationship will be confirmed in the estimation results of the parametric model.

As mentioned in Section 5.2, it is very likely that effort expended depends upon the complexity of the problem, and therefore we must control for this in some way. We assume that subjects assess the complexity of a problem using a very simple rule: they count the number of outcomes appearing in the simpler of the two lotteries.[3] This rule gives rise to three levels of complexity, to which we shall refer as Level 1, Level 2, and Level 3. For specific examples, we may refer back to Figure 5.1, and note that task 14 is of complexity level 1, since one lottery involves only a single outcome of $10, while task 40 is of the higher level of complexity 3, since both lotteries involve this number of outcomes.

[3] A finer classification of problems was originally used, but the estimated model indicated that this simple classification into three levels is sufficient to explain the data. Hey (1995) and others have used a different measure: the mean taken over the two lotteries of the number of outcomes.



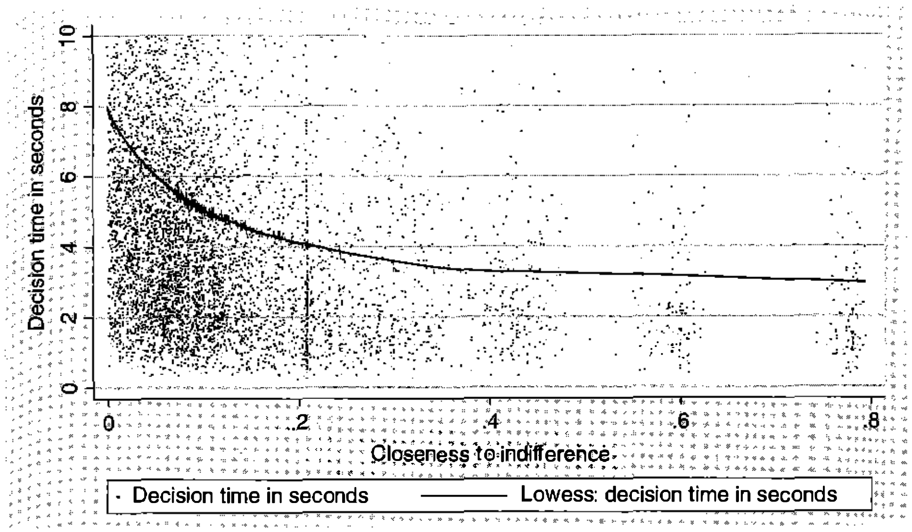



Figure 5.6: Decision time against closeness to indifference

In the following section, we will estimate the following (log-)linear regression model, with the logarithm of decision time as the dependent variable:

$$\begin{aligned}\log(\textit{decision time}_{it}) &= \alpha + \beta_1 complex2_t + \beta_2 complex3_t + \beta_3\tau^d_{it} + \beta_4\tau_{it} \\ &+\beta_5 log\,(EV_t) + \beta_6|\hat{\Delta}_{it}| + \beta_7|\hat{\Delta}_{it}|^2 + \beta_8|\hat{\Delta}_{it}|^3 + \beta_9\Delta^o_t + u_i + \epsilon_{it} \qquad (5.2) \\ &i = 1\ldots,n \quad t = 1\ldots,T \quad var(u_i) = \sigma^2_u \quad var(\epsilon_{it}) = \sigma^2_\epsilon\end{aligned}$$

In (5.2), $i$ indexes subjects, while $t$ indexes problems. Note that there are two stochastic terms: $u_i$ is the subject-specific effect, with mean zero and variance $\sigma^2_u$; $\epsilon_{it}$ is the equation error, with mean zero and variance $\sigma^2_\epsilon$. Equation (5.2) is either a fixed effects (FE) or a random effects (RE) model, depending on how the individual effect $u_i$ is interpreted. Both of these models will be estimated in the next section. We shall also estimate a model in which $u_i$ is assumed to be zero for all $i$. Such a model is known as a pooled regression model, since it disregards the panel structure of the data.

The first two explanatory variables in (5.2) are dummy variables indicating the level of complexity of problem $t$, according to the rule defined above. The excluded complexity level is the least complex, Level 1. The third and fourth explanatory variables represent the position of a problem in the sequence: $\tau^d_{it}$ is the position of problem $t$ within the day on which the problem was solved, so $\tau^d$ ranges from 1 through 50; $\tau_{it}$ is in contrast the position of problem $t$ in the complete sequence of problems faced by subject $i$, and therefore ranges from 1 through 100. The fifth explanatory variable represents the financial incentive associated with each problem. Many different measures could be used here; the one which is chosen is the logarithm of the expected value of the simpler of the two lotteries. The next three explanatory variables are the closeness to indifference of subject $i$ in choice problem $t$, as defined briefly in Section 5.3, and its square and cube. The purpose of these three variables is to allow closeness to indifference to have a non-linear effect on effort, as it appears to do on the evidence of the non-parametric regression shown in Figure 5.6. The final explanatory variable is our measure of the objective difference between the two lotteries, as defined in (5.1).

The various variables in the data set (**decision_times_sim**) are named as follows:

| | |
|---|---|
| log_dt: | natural logarithm of decision time in seconds; |
| complex: | level of complexity of choice problem (1,2, or 3); |
| tau_d: | position of choice problem within single day (1–50) ($\tau^d_{it}$); |
| tau: | position of choice problem in complete sequence of 100 ($\tau_{it}$); |
| log_ev: | natural logarithm of expected value of simpler lottery; |
| cti: | closeness-to-indifference ($\lvert\hat{\Delta}_t\rvert$); |
| obj_diff: | objective difference between the two lotteries ($\Delta^o_t$). |

## 5.5 Panel Data Models of Effort Allocation

We start by estimating (5.2) using a pooled model, that is, a regression using the complete sample of $nT$ observations, that attributes all unexplained variation in terms of within-subject randomness, and does not allow for any variation between subjects. In the context of (5.2), the pooled model is one in which $u_i = 0$ for all $i$. The following STATA command performs the required regression:

```
. regress  log_dt complex2 complex3 tau_d tau  logev cti cti2 cti3 obj_diff
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 576.37795 | 9 | 64.0419944 |
| Residual | 3288.28327 | 5990 | .548962149 |
| Total | 3864.66122 | 5999 | .644217573 |

| | |
|---|---|
| Number of obs = | 6000 |
| F( 9, 5990) = | 116.66 |
| Prob > F = | 0.0000 |
| R-squared = | 0.1491 |
| Adj R-squared = | 0.1479 |
| Root MSE = | .74092 |

| log_dt | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| complex2 | .2397661 | .0493898 | 4.85 | 0.000 | .1429443 | .336588 |
| complex3 | .3804073 | .0654843 | 5.81 | 0.000 | .2520345 | .5087801 |
| tau_d | -.002203 | .0007657 | -2.88 | 0.004 | -.0037041 | -.000702 |
| tau | -.0032846 | .0003826 | -8.58 | 0.000 | -.0040346 | -.0025345 |
| logev | .0609713 | .0562583 | 1.08 | 0.279 | -.0493151 | .1712578 |
| cti | -5.590682 | .3889394 | -14.37 | 0.000 | -6.353144 | -4.828221 |
| cti2 | 12.31351 | 1.443477 | 8.53 | 0.000 | 9.483773 | 15.14324 |
| cti3 | -8.576194 | 1.349663 | -6.35 | 0.000 | -11.22202 | -5.930369 |
| obj_diff | .1323685 | .0428824 | 3.09 | 0.002 | .0483035 | .2164335 |
| _cons | 1.648698 | .0793198 | 20.79 | 0.000 | 1.493202 | 1.804193 |

We see that nearly all of the explanatory variables have strongly significant effects on decision time. However, we shall avoid interpreting each individual effect at this stage. This is because we will shortly be reporting estimates from panel data models



that are found to be superior to this pooled model. We shall interpret the results of the most preferred model in the next section.

On the evidence of Figure 5.4, we have reason to believe that there is strong dependence at the subject level, or "subject-level clustering". As in Chapter 4, we shall start to address this problem by simply correcting the standard errors for this sort of dependence. Recall that this requires using the `vce(cluster i)` option with the regress command.

```
regress  log_dt complex2 complex3 tau_d tau logev cti cti2 cti3 obj_diff ///
 , vce(cluster i)
```

Linear regression

| | |
|---|---|
| Number of obs = | 6000 |
| F( 9, 59) = | 196.50 |
| Prob > F = | 0.0000 |
| R-squared = | 0.1491 |
| Root MSE = | .74092 |

(Std. Err. adjusted for 60 clusters in i)

| log_dt | Coef. | Robust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| complex2 | .2397661 | .0395952 | 6.06 | 0.000 | .1605363 | .318996 |
| complex3 | .3804073 | .0529175 | 7.19 | 0.000 | .2745197 | .486295 |
| tau_d | -.002203 | .0008415 | -2.62 | 0.011 | -.0038868 | -.0005192 |
| tau | -.0032846 | .000355 | -9.25 | 0.000 | -.0039949 | -.0025743 |
| logev | .0609713 | .047967 | 1.27 | 0.209 | -.0350105 | .1569531 |
| cti | -5.590682 | .3573599 | -15.64 | 0.000 | -6.305758 | -4.875607 |
| cti2 | 12.31351 | 1.283165 | 9.60 | 0.000 | 9.745901 | 14.88111 |
| cti3 | -8.576194 | 1.178483 | -7.28 | 0.000 | -10.93433 | -6.218055 |
| obj_diff | .1323685 | .0340484 | 3.89 | 0.000 | .0642379 | .2004992 |
| _cons | 1.648698 | .0713508 | 23.11 | 0.000 | 1.505925 | 1.79147 |

Note that the estimates themselves are identical to those obtained in the previous regression in which no adjustment was made for clustering. Only the standard errors have changed. Nearly all of the corrected standard errors are smaller than their uncorrected counterparts. This means that the t-statistics are larger in magnitude, and therefore greater significance is detected, as a result of making the correction. However, in this case, none of the changes are great enough to result in reversals of conclusions regarding significance of effects.

Using cluster-robust standard errors is only a first step in addressing the issue of the panel structure of the data. It is also possible to improve the estimates themselves, using a panel-data estimator instead of OLS. As discussed in Section 4.6 the two most popular panel data estimators are the fixed-effects and random-effects estimators. Both are represented by equation (5.2). Recall that in (5.2) there are two error terms: $\epsilon_{it}$ is the conventional equation error term, which is assumed to have mean zero and variance $\sigma^2_\epsilon$; $u_i$ is known as the subject-specific term. $u_i$ differs between subjects – hence the $i$-subscript – but it is fixed for a given subject. The two estimators differ in the way this term is interpreted.

The difference between fixed effects and random effects was explained in Section 4.6, and will be explained only briefly here. The fixed effects estimator is essentially a linear regression which includes a set of $n - 1$ dummy variables, one for each subject in the data set (with one excluded to avoid the dummy variable trap). Hence a different intercept is estimated for each subject. The random effects estimator does *not* estimate the intercept for each subject. It simply recognises that they are all different, and sets out to estimate only their *variance*, $\sigma_u^2$. Note that random effects is more efficient than fixed effects, because there are far fewer parameters to estimate. We therefore prefer to use random effects if this model turns out to be acceptable.

The first question we might wish to ask is: is a panel data model required at all? If there are no differences between subjects, the pooled regression model estimated above is the correct model. The most obvious way of testing for differences within subjects is to test for equality of the subject fixed effects in the fixed-effects model. The null hypothesis for this test is:

$$H_0 : u_1 = u_2 = \cdots = u_n = 0$$

This null hypothesis embodies $n - 1$ restrictions on the model. The test is routinely performed as an F-test in the estimation of the fixed-effects model. If this null hypothesis is rejected (it nearly always is) we conclude that there are significant differences between subjects which make the pooled regression model invalid and necessitate the use of panel data estimation.

To decide between fixed effects (FE) and random effects (RE), the Hausman test is used. This test is based on a comparison of the two sets of estimates. The test will be explained in greater detail in Section 7.6. Roughly expressed the reasoning is as follows. The assumptions underlying RE are more stringent than those underlying FE. If the two sets of estimates are close to each other, this implies that the assumptions underlying both sets of estimates are valid, and RE is preferred because it is a more efficient estimator. If the two sets of estimates are very different from each other, this implies that only FE can be correct, and the assumptions underlying RE must be false. Hence, if the Hausman test results in a rejection, FE is preferred, while a failure to reject leads us to favour RE.

As mentioned in Section 4.6 panel data commands in STATA always start with the prefix `xt`. For example panel data (linear) regression is carried out using `xtreg`. The fixed effects and random effects estimators are implemented using this command with the options `fe` and `re` respectively.

The two models are estimated, and the Hausman test performed, with the following sequence of commands:

```
. xtset i t
       panel variable:  i (strongly balanced)
        time variable:  t, 1 to 100
                delta:  1 unit

. xtreg log_dt complex2 complex3 tau_d tau logev cti cti2 cti3 obj_diff, fe

Fixed-effects (within) regression               Number of obs      =      6000
Group variable: i                               Number of groups   =        60

R-sq:  within  = 0.2004                         Obs per group: min =       100
       between = 0.0002                                        avg =     100.0
       overall = 0.1491                                        max =       100

                                                F(9,5931)          =    165.12
corr(u_i, Xb)  = -0.0059                        Prob > F           =    0.0000
```



```
------------------------------------------------------------------------------
      log_dt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    complex2 |   .2374196   .0417019     5.69   0.000     .1556687    .3191706
    complex3 |   .3770243   .0552954     6.82   0.000     .2686252    .4854234
       tau_d |  -.0022045   .0006465    -3.41   0.001    -.0034718   -.0009372
         tau |  -.0032846    .000323   -10.17   0.000    -.0039179   -.0026513
       logev |   .0601018   .0474991     1.27   0.206    -.0330137    .1532173
         cti |  -5.671949    .330508   -17.16   0.000    -6.319865   -5.024033
        cti2 |   12.46305    1.22361    10.19   0.000     10.06433    14.86177
        cti3 |  -8.645294   1.142589    -7.57   0.000    -10.88518   -6.405403
    obj_diff |   .1324667   .0362049     3.66   0.000     .0614919    .2034415
       _cons |   1.656612   .0670205    24.72   0.000     1.525227    1.787996
-------------+----------------------------------------------------------------
     sigma_u |  .40495036
     sigma_e |  .62554574
         rho |  .29531259   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(59, 5931) =    41.90            Prob > F = 0.0000

. est store fe

. xtreg log_dt complex2 complex3 tau_d tau logev cti cti2 cti3 obj_diff, re

Random-effects GLS regression                   Number of obs      =      6000
Group variable: i                               Number of groups   =        60

R-sq:  within  = 0.2004                         Obs per group: min =       100
       between = 0.0002                                        avg =     100.0
       overall = 0.1491                                        max =       100

                                                Wald chi2(9)       =   1486.20
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      log_dt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    complex2 |   .2374744   .0416953     5.70   0.000     .1557531    .3191957
    complex3 |   .3771036   .0552865     6.82   0.000      .268744    .4854631
       tau_d |  -.0022045   .0006464    -3.41   0.001    -.0034713   -.0009376
         tau |  -.0032846    .000323   -10.17   0.000    -.0039176   -.0026516
       logev |   .0601223   .0474916     1.27   0.206    -.0329595    .1532041
         cti |  -5.670031   .3304062   -17.16   0.000    -6.317615   -5.022446
        cti2 |   12.45948   1.223303    10.19   0.000     10.06185    14.85711
        cti3 |  -8.643612   1.142337    -7.57   0.000    -10.88255   -6.404672
    obj_diff |   .1324644   .0361992     3.66   0.000     .0615152    .2034136
       _cons |   1.656426   .0851303    19.46   0.000     1.489574    1.823278
-------------+----------------------------------------------------------------
     sigma_u |  .40678176
     sigma_e |  .62554574
         rho |   .2971941   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. est store re

. hausman fe re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      fe           re          Difference          S.E.
-------------+----------------------------------------------------------------
    complex2 |    .2374196     .2374744       -.0000548         .0007429
    complex3 |    .3770243     .3771036       -.0000792         .0009906
       tau_d |   -.0022045    -.0022045       -3.38e-08         .0000115
         tau |   -.0032846    -.0032846        8.98e-17         5.72e-06
       logev |    .0601018     .0601223       -.0000205         .0008433
         cti |   -5.671949    -5.670031       -.0019184         .0082029
        cti2 |    12.46305     12.45948          .00357         .0274092
        cti3 |   -8.645294    -8.643612       -.0016819         .0239862
    obj_diff |    .1324667     .1324644        2.26e-06         .0006416
------------------------------------------------------------------------------
```

```
                b = consistent under Ho and Ha; obtained from xtreg
         B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.17
                Prob>chi2 =      1.0000
```

We see that the two sets of results are indeed very similar, and it is not surprising that the Hausman test leads us to favour the random effects model (with a p-value of 1.0000). In the following section, we shall therefore focus on the results of the random effects model in interpretation.

## 5.6 Discussion of Results

For ease of comparison, we present the results from all estimated models in Table 5.2.

Since the F-test strongly rejects the pooled OLS model ($p = 0.0000$), and the Hausman test indicates acceptance of the random effects model ($p = 1.0000$), we shall interpret the results from the random effects model, shown in the final column of Table 5.2.

| Log decision time | OLS | Cluster OLS | Fixed effects | Random effects |
|---|---|---|---|---|
| Constant | 1.649(0.073) | 1.649(0.071) | 1.656(0.067) | 1.656(0.085) |
| Complexity 1 (base) | — | — | — | — |
| Complexity 2 | 0.240(0.049) | 0.240(0.040) | 0.237(0.042) | 0.237(0.042) |
| Complexity 3 | 0.380(0.065) | 0.380(0.053) | 0.377(0.055) | 0.377(0.055) |
| $\tau^d$ | −0.002(0.0008) | −0.002(0.0008) | −0.002(0.0006) | −0.002(0.0006) |
| $\tau$ | −0.003(0.0004) | −0.003(0.0004) | −0.003(0.0003) | −0.003(0.0003) |
| $Log(EV)$ | 0.061(0.056) | 0.061(0.048) | 0.060(0.047) | 0.060(0.047) |
| $cti$ | −5.591(0.389) | −5.591(0.357) | −5.672(0.331) | −5.670(0.330) |
| $cti^2$ | 12.313(1.443) | 12.313(1.283) | 12.463(1.223) | 12.459(1.223) |
| $cti^3$ | −8.576(1.350) | −8.576(1.178) | −8.645(1.143) | −8.643(1.142) |
| $\Delta^o$ | 0.132(0.043) | 0.132(0.034) | 0.132(0.036) | 0.132(0.036) |
| $\sigma_\epsilon$ | 0.741 | 0.741 | 0.625 | 0.625 |
| $\sigma_u$ | — | — | 0.405 | 0.407 |
| $n$ | 60 | 60 | 60 | 60 |
| $T$ | 100 | 100 | 100 | 100 |
| F-test (59, 5931) | | 41.90 ($p$=0.0000) | | |
| Hausman test $\chi^2(9)$ | | | 0.17($p$=1.0000) | |

Table 5.2: Results from log-linear regression models for decision time
Note: Standard errors in parentheses.



Firstly we note from the estimate of the intercept that the predicted decision time for the "easiest" type of problem, that is, a problem of complexity Level 1, for which the two lotteries are in fact identical ($\Delta^o = 0$), is exp(1.656) = 5.238 seconds. Of course, this problem must also be assumed to be right at the start of the experiment ($\tau^d = \tau = 0$) and this might explain why the prediction appears so high for such a simple problem.

Secondly we note that all included explanatory variables, except one, show strong significance. The effect of complexity is as expected: problems involving more outcomes take longer to solve. We do notice that the extra effort induced in moving from Level 2 to Level 3 is lower than that induced in the move from 1 to 2. This might be interpreted as evidence that subjects are discouraged by complex tasks, and this interpretation could be extended to predict a reduction of effort when complexity reaches intolerable levels. However, there is no evidence that such levels of complexity are being encountered in this experiment.

The effect of experience is quite dramatic. Both $\tau$ and $\tau^d$ are seen to have a strongly significant negative effect on decision time. As previously suggested, the first of these is interpretable as an experience effect, and the second as a boredom effect.

The coefficient of the variable $log(EV)$, 0.060, may be interpreted as an elasticity of effort with respect to financial incentives: if all prizes were doubled, we would expect response times (effort) to rise by around 6%. This is consistent with the view of many economists concerning incentives: that higher incentives bring about an increase in effort expended. However, we see that this effect is not actually significant, so we do not have statistical confirmation of this prior belief.

Finally, and perhaps most importantly, we see a strong negative and non-linear effect of closeness to indifference coupled with a strong positive effect of objective difference. This is exactly what we expected on the basis of our theoretical model of effort allocation described in Section 5.3. Figure 5.7 demonstrates the first effect clearly. Here, we have used the model estimates to predict the decision time at each value of closeness-to-indifference, for different complexity levels, and with other explanatory variables set to representative values. We see from Figure 5.7 that as the subject approaches indifference, the decision time rises steeply, and when a subject is actually indifferent (i.e. when closeness-to-indifference = 0), the predicted response time is more than double what it would be if one alternative were clearly preferred.

Of course, the pattern seen in Figure 5.7 is not surprising given the very similar shape seen in the non-parametric regression in Figure 5.6. This similarity is consistent with the correctness of the specification of our effort model (5.2)

## 5.7 Post-estimation

In Section 5.6 we established that the random effects model is the most suitable model for analysing the data set on decision times, and we interpreted the results from this model. The random effects approach is a natural means of capturing

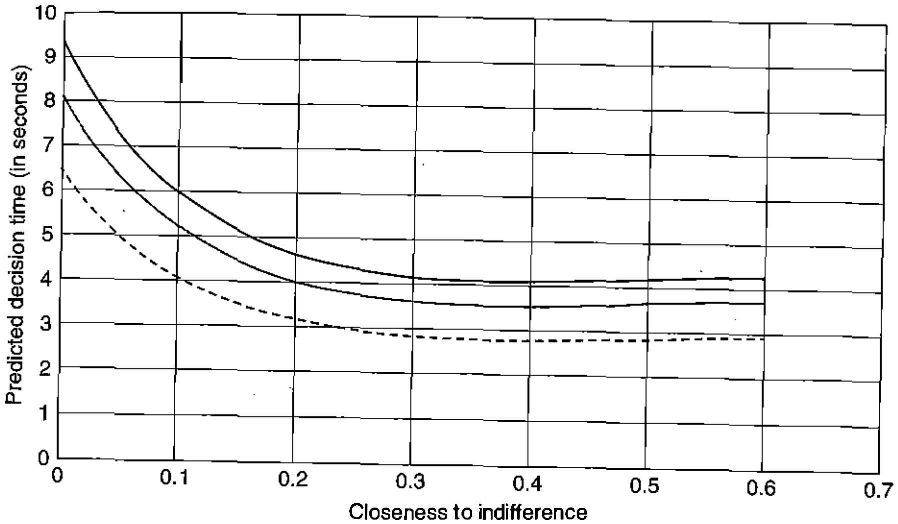



Figure 5.7: Predicted decision time against absolute valuation differential, at different complexity levels

Notes: $\tau$ set to 0; EV set to 10; $\Delta^0$ set to 0.5.

between-subject heterogeneity, and it is an approach that is followed repeatedly throughout the book. Sometimes, the random effect term has a clear interpretation. For example, in the risky choice model estimated in Chapter 13, the random effect term may be interpreted as the subject's coefficient of risk aversion.

Having estimated a random effects model, a natural question to ask is: what is the (estimated) value of the random effect for each subject? We shall refer to these values as the posterior random effects. This is a task that is carried out routinely following estimation. It usually involves the application of Bayes' rule to the estimated parameters together with the data. After using the `xtreg` command, only one additional command is required to generate the posterior random effects. The command is:

```
predict u_hat, u
```

This command stores the posterior random effects in the new variable u_hat. We present descriptive statistics of this variable below, and a histogram of u_hat is shown in Figure 5.8. Note that, for these purposes, we only require one observation for each subject; hence, the use of `if t==1` in the command.

```
. summ u_hat if t==1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       u_hat |        60    5.59e-10    .3955946  -1.093639   .7443957
```



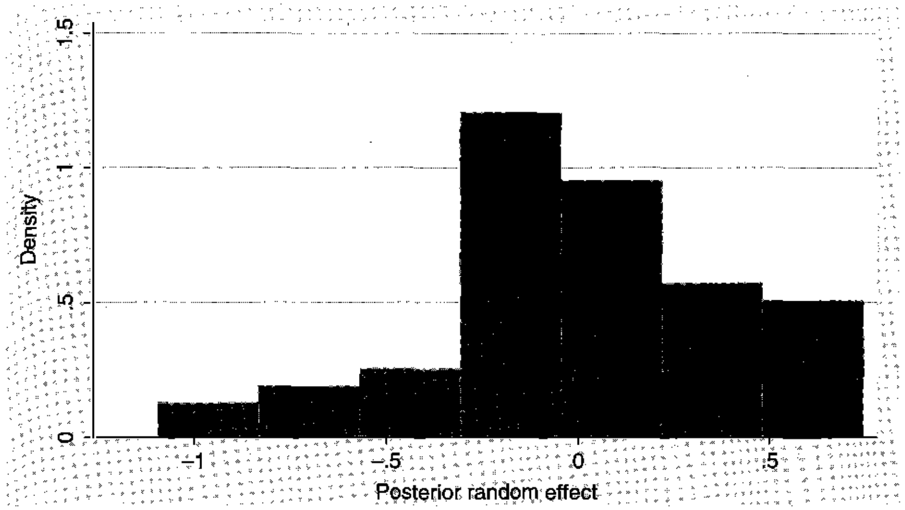



Figure 5.8: A histogram of posterior random effects

As expected, the posterior random effect has a mean of zero, and a standard deviation very close to the estimate of $\sigma_u^2$ reported in the random effects results, which was 0.407.

In the discussion of Figure 5.4, it was noted that subject 24 is the fastest decision maker, with a mean decision time of 1.558 seconds. It is therefore not surprising that this subject has the lowest posterior random effect, of $-1.09$. It was also noted that subject 29 is the slowest decision maker, with a mean decision time of 9.947 seconds. Again, it is not surprising that this subject has the highest posterior random effect, of +0.744.

## 5.8 Summary and Further Reading

This chapter has applied regression analysis to the modelling of decision times, with a view to identifying the determinants of effort in experimental tasks. Similar goals have been pursued by Wilcox (1994), who finds evidence that the use of simple rules in valuation tasks is associated with a reduction in decision time, suggesting that the use of such rules are motivated by a desire to save effort. Hey (1995) identifies factors which influence decision time in risky choice tasks, and then estimates choice models which allow the "noisiness" of response to depend on these factors. Decision times have also been analysed by Buschena & Zilberman (2000), Moffatt (2005*b*), and Alos-Ferrer et al. (2012).

The data used in this chapter was simulated. However, the simulation was created to resemble the real data set of Hey (2001) analysed by Moffatt (2005*b*). Readers interested to know how the simulation was performed are referred to Sections 13.3.2 and 13.4.6 of this book.

This seems to be a good example of an area in which the two disciplines of experimental economics and psychology overlap. Economists have traditionally been interested in the decisions made, not in the process by which the decision is reached. Psychologists are interested in the process (see, for example, Busemeyer & Townsend, 1993). The recent surge of interest in the analysis of decision times in the experimental economics literature is perhaps a sign that economists are becoming more interested in the decision-making process.

Of course there are more informative ways of analysing the decision-making process than simply observing decision times. One type of analysis that is starting to become popular among experimental economisis is eye-tracking (Hohnqvist et al., 2011).

## Exercise

Little (1949, p. 92) posed the following question: "how long must a person dither before he is pronounced indifferent?" How can the results presented in Section 5.6, and particularly Figure 5.7, be invoked to provide an answer to this question?

# Chapter 6
# Dealing with Discreteness in Experimental Data

## 6.1 Introduction

There is at least an element of discreteness in most data that results from economic experiments. In many settings, a typical task requires the subject to make a straightforward choice between a number of discrete alternatives, and discrete choice models are required for the analysis of the data. The number of alternatives is often two, in which case binary data models are required. Even in situations in which the decision variable is continuous, the variable is often censored at the lower and/or upper limits, or there may be an accumulation of data at some other "focal point"; such data features are best modelled as forms of discreteness.

This chapter is concerned with this sort of modelling. It starts by covering models appropriate when the decision variable is binary. These models are simple probit and simple logit. A number of examples are given, including the choice between lotteries and the analysis of the responder's decision in an ultimatum game.

The chapter will also cover a variety of other data types: exact data; censored data; and interval data. Again, examples are used of each.

This chapter also introduces the distinction between the various competing approaches to stochastic modelling that have appeared in the recent literature. These are: the Fechner approach, in which optimal behaviour simply has a mean-zero, symmetrically distributed error term appended; the heterogeneous agent approach, which attributes variation in behaviour to variation over the population in the parameters representing preferences; and, finally, the "tremble" approach, in which it is assumed that, at any point in time, there is a small probability that an individual loses concentration, and their behaviour is determined in a random fashion. It should be noted that it is possible to combine these approaches, and many models do this.

Note that in every example presented in this chapter, it is assumed that there is only one observation per subject. Methods appropriate in the more common setting of multiple observations per subject have already been covered in Chapters 4 and 5 for linear models, and will be covered in later chapters for discrete data.

Section 6.2 introduces the concept of maximum likelihood estimation in the context of binary data modelling. Section 6.3 introduces the `ml` routine in STATA, which is used many times throughout the book. Sections 6.4 and 6.5 introduce various approaches to structural modelling, by which we mean the estimation of the parameters of the decision-maker's utility function. Section 6.6 covers various other data types, such as interval and censored data. Section 6.7 applies some of the modelling approaches to ultimatum game data, for example estimating a model that treats the proposer's decision as a problem of choice under risk.

## 6.2 Binary Data

The standard method of estimation for binary data models, and for most of the models introduced in this chapter, and indeed for most of the models introduced throughout the book, is the method of Maximum Likelihood (ML). This section starts by presenting a straightforward example involving coloured balls, which is ideal for conveying both the concept of maximum likelihood estimation, and at the same time raises a number of important practical issues. We then apply the method to binary data in two different experimental contexts.

Section 6.2.1 could safely be skipped by readers who are already familiar with the concepts of the likelihood function and maximum likelihood estimation.

### *6.2.1 Likelihood and log-likelihood explained*

Suppose that we have an urn containing 1,000 balls, coloured either red or white. We do not know how many of the balls are red or how many are white. Our objective is to estimate the proportion of the balls in the urn that are red. Let this proportion be $p$. The proportion of white balls is then $1 - p$.

We draw a sample of ten balls *with replacement*. That is, after each draw, we replace the drawn ball. This guarantees independence between the ten draws. Let us suppose that our ten draws result in seven red balls and three white balls. This is our *sample*. The sample might be presented as:

RRRWRRWRWR

For most purposes, the order in which the ten balls have been drawn is irrelevant.

We write down the *likelihood function* for the sample. Think of this as "the probability of observing the sample that we observe". This is:

$$L(p) = p^7 (1 - p)^3 \qquad (6.1)$$

Note that (6.1) is formed as simply the product of ten probabilities, as a consequence of the assumption of *independence* between observations. If the observations were not independent (e.g. if the sample had been drawn without replacement) the likelihood function would be somewhat more complicated than (6.1).

The maximum likelihood estimate (MLE) of $p$ is the value that maximises the likelihood function (6.1). In other words, it's the value which makes the observed



sample most likely. We denote the MLE by $\hat{p}$. One way of obtaining $\hat{p}$ is by differentiating (6.1) with respect to $p$, setting equal to zero, and solving for $p$:

$$\begin{aligned}\frac{\partial L}{\partial p} &= 7p^6(1-p)^3 - 3p^7(1-p)^2 \\ &= p^6(1-p)^2\left[7(1-p) - 3p\right] \\ &= p^6(1-p)^2\left[7 - 10p\right] = 0\end{aligned} \qquad (6.2)$$

This gives three solutions: $p = 0$; $p = 1$; $p = 7/10$. So there are three turning points. Figure 6.1 shows a graph of $L$ against $p$. We can see that there is a maximum at $p = 7/10$, and minima at 0 and 1. So the MLE of p is:

$$\hat{p} = \frac{7}{10} \text{ (the proportion of reds in the sample)} \qquad (6.3)$$

Figure 6.2 shows a graph of $LogL$ in addition to $L$. Note that $L$ appears very flat as a result of the change in scale. An important point is that maximising the log of $L$ gives the same answer as maximising $L$ itself. This is obvious because log is a monotonically increasing function. Let us verify that it gives the same answer. The log of (6.1) is:

$$LogL = 7log(p) + 3log(1 - p) \qquad (6.4)$$

Differentiating (6.4), we have:

$$\frac{\partial LogL}{\partial p} = \frac{7}{p} - \frac{3}{(1-p)} \qquad (6.5)$$

Setting (6.5) equal to zero gives $3p = 7(1 - p)$, which implies $\hat{p} = \frac{7}{10}$, the same answer as equation (6.3).

The reason why this is important is that in practice the sample size is large, so the likelihood function, $L$, being a product of a large number of probabilities, is very close to zero, so is a very flat function. This makes it difficult, in practice, to locate the maximum of $L$. Taking the log of $L$ has the effect of "stretching" the function vertically, making it easier to locate the maximum. For this reason, we nearly always maximise $Log\,L$, not $L$.

The log-likelihood function has a number of other practical uses. One is that its curvature may be used to obtain the standard error associated with the MLE. Differentiating (6.5) with respect to $p$, we obtain the second derivative of the log-likelihood with respect to the parameter:

$$\frac{\partial^2 LogL}{\partial p^2} = -\frac{7}{p^2} - \frac{3}{(1-p)^2} \qquad (6.6)$$

Equation (6.6) is usually referred to as the "Hessian matrix", although, in this case, since the parameter vector consists of only one element, it is a scalar. The first thing

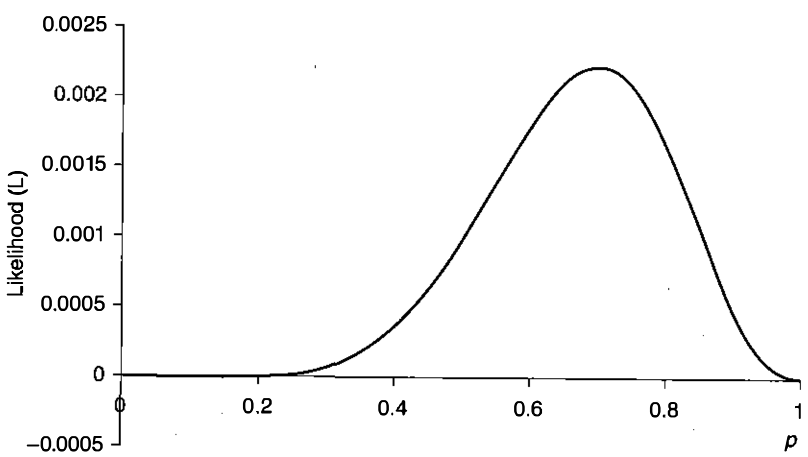



Figure 6.1: The likelihood function

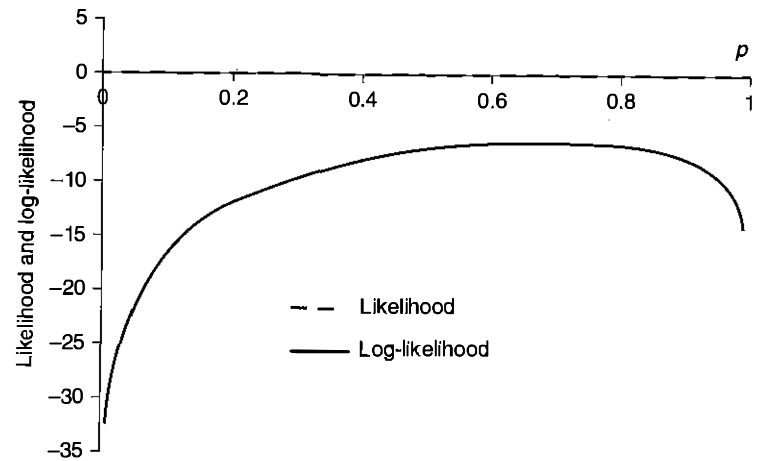



Figure 6.2: The likelihood and log-likelihood functions

to notice about (6.6) is that it is strictly negative for all $p$. This is important because it implies that the log-likelihood function is globally concave, and therefore that (in this situation) the MLE is unique. When the Hessian matrix really is a matrix, the necessary condition for uniqueness of the MLE is that the Hessian matrix is negative definite. This condition is satisfied for some, but certainly not all, models.

The estimated (asymptotic) covariance matrix of the MLE vector is usually obtained as the inverse of the negative of the Hessian matrix, evaluated at the MLE. In this case, this results in the scalar:

$$\hat{V}\left(\hat{p}\right) = \left[-\frac{\partial^2 LogL}{\partial p^2}\right]^{-1}\Bigg|_{p=\hat{p}} = \left[\frac{7}{\hat{p}^2} + \frac{3}{\left(1-\hat{p}\right)^2}\right]^{-1} \qquad (6.7)$$



Inserting our MLE, $\hat{p} = 0.7$, into (6.7), we obtain:

$$\hat{V}\left(\hat{p}\right) = \left[\frac{7}{0.7^2} + \frac{3}{0.3^2}\right]^{-1} = \underline{0.021} \qquad (6.8)$$

Finally, taking the square root of the asymptotic variance estimate (6.8), we obtain an estimate of the standard error of $\hat{p}$. This is:

$$a.s.e\left(\hat{p}\right) = \sqrt{0.021} = \underline{\underline{0.145}} \qquad (6.9)$$

We refer to (6.9) as the "asymptotic standard error" of the MLE of the parameter $p$. The word "asymptotic" is being used because the theory that establishes the formula (6.7) is asymptotic (i.e. large sample) theory.

It is interesting to see what happens when the same exercise is carried out with a larger sample size. Let us suppose that instead of drawing ten balls, we draw 100 (again, with replacement), of which 70 are red and 30 are white. The MLE of $p$ will be 0.7 again. However, the standard error will be different:

$$a.s.e\left(\hat{p}\right) = \sqrt{\left[\frac{70}{0.7^2} + \frac{30}{0.3^2}\right]^{-1}} = \underline{0.046} \qquad (6.10)$$

Note that the ten-fold increase in the sample size has resulted in a reduction in the standard error of a factor around one-third. This is because the log-likelihood function becomes more peaked in the neighbourhood of the MLE when the sample size increases. In fact, this is a manifestation of the "root-n" law that is a centre-piece of asymptotic theory: the precision of an estimate improves with the square root of the sample size.

### *6.2.2 Modelling choices between lotteries (the "house money effect")*

In this sub-section, we consider a very popular application of binary data models: risky choice experiments. We will use the models to test a particular hypothesis relating to behaviour in this context. The hypothesis of interest is the "house money effect", that is, the phenomenon of choices becoming more risk-seeking when the initial endowment is higher (see Thaler & Johnson, 1990; Keasey & Moon, 1996).

Consider the choice problem presented in Figure 6.3, where the two circles represent lotteries, and the areas within them represent probabilities of the stated outcomes (the same lottery choice example was used to demonstrate the use of non-parametric tests in Chapter 3). The left-hand lottery is the "safe" lottery and it pays \$5 with certainty. The right-hand lottery is the "risky lottery" and represents a 50:50 gamble involving the outcomes \$0 and \$10.

Clearly, by choosing between the lotteries in Figure 6.3, a subject is conveying some information about his or her attitude to risk. What is of interest here is whether previously endowing a subject with an amount of money has an effect on this choice.

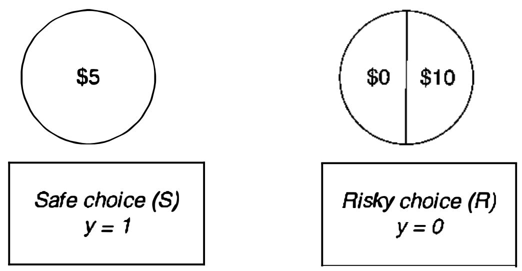



Figure 6.3: A lottery choice problem

Let us define the "house money effect" as the phenomenon of agents becoming less risk averse (i.e. more likely to choose the risky lottery) when their initial endowment (i.e. "house money") increases.

Suppose we have a sample of 1,050 subjects. We endow each subject ($i$) with a different wealth level ($w_i$); we then immediately ask them to choose between the two lotteries shown in Figure 6.3. We then define the binary variable $y$ to take the value 1 if the safe lottery is chosen, and 0 if risky is chosen. The results of this (imaginary) experiment are contained in the file **house_money_sim**. Here is some summary information about the data:

```
. table w, contents(n y mean y)
----------------------------------
        w |       N(y)     mean(y)
----------+-----------------------
        0 |         50         .92
       .5 |         50         .88
        1 |         50         .88
      1.5 |         50         .84
        2 |         50         .84
      2.5 |         50          .9
        3 |         50         .84
      3.5 |         50         .72
        4 |         50         .78
      4.5 |         50          .7
        5 |         50          .7
      5.5 |         50         .74
        6 |         50         .72
      6.5 |         50         .72
        7 |         50          .5
      7.5 |         50         .64
        8 |         50          .5
      8.5 |         50         .48
        9 |         50         .56
      9.5 |         50          .5
       10 |         50          .5
----------------------------------
```

The final column of the table shows the mean of the binary variable for different wealth levels. Since the mean of a binary variable is the proportion of ones in the sample, the numbers in this column represent the proportion choosing the safe lottery at each wealth level. The tendency for this proportion to fall as wealth rises is consistent with the house money effect.

Next we set out to confirm this using a parametric model. A natural model to start with is the probit model, defined as follows:



$$P(y_i = 1|w_i) = \Phi(\beta_0 + \beta_1 w_i) \tag{6.11}$$

where $\Phi(.)$ is the standard normal cdf.[1] The likelihood function for the probit model is:

$$L = \prod_{i=1}^{n} [\Phi(\beta_0 + \beta_1 w_i)]^{y_i} [1 - \Phi(\beta_0 + \beta_1 w_i)]^{1-y_i} \tag{6.12}$$

and the log-likelihood is:

$$LogL = \sum_{i=1}^{n} \left[ y_i ln(\Phi(\beta_0 + \beta_1 w_i)) + (1 - y_i) ln(1 - \Phi(\beta_0 + \beta_1 w_i)) \right] \tag{6.13}$$

An important property of the cdf (6.11) defining the probit model is symmetry. By this, we mean that $\Phi(-z) = 1 - \Phi(z)$. This property also applies to the distribution underlying the logit model (see Exercise 1). This feature of the underlying distribution is useful because it allows the log-likelihood function to be written more compactly as follows. If we recode the binary variable as:

$$yy_i = 1 \text{ if } S \text{ is chosen}$$
$$yy_i = -1 \text{ if } R \text{ is chosen}$$

then the log-likelihood (6.13) can be written as:

$$LogL = \sum_{i=1}^{n} ln(\Phi(yy_i \times (\beta_0 + \beta_1 w_i))) \tag{6.14}$$

We maximise LogL defined in (6.14) to give **MLEs** of the two parameters $\beta_0$ and $\beta_1$. This task is performed using the `probit` command in STATA, as follows:

```
. probit y w

Iteration 0:   log likelihood =  -634.4833
Iteration 1:   log likelihood = -584.91375
Iteration 2:   log likelihood =  -584.5851
Iteration 3:   log likelihood = -584.58503
Iteration 4:   log likelihood = -584.58503

Probit regression                                 Number of obs   =       1050
                                                  LR chi2(1)      =      99.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -584.58503                       Pseudo R2       =     0.0786

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1409882   .0145377    -9.70   0.000    -.1694816   -.1124948
       _cons |   1.301654   .0911155    14.29   0.000     1.123071    1.480237
------------------------------------------------------------------------------
```

[1] If a random variable $Z$ has a standard normal distrubution, its density function is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$, and its cumulative distribution function (cdf) is $\Phi(z) = P(Z < z) = \int_{-\infty}^{z} \phi(z)dz$.

The first thing we might do when we have obtained the results is to test for the presence of the house money effect. This test is done for us. The asymptotic t-statistic associated with wealth is $z = -9.70$, and the associated p-value is 0.000. This tells us that there is strong evidence that wealth has a negative effect on the probability of choosing the safe lottery. In other words, there is strong evidence of the house money effect in this data.

There is a STATA command `test` that can be used immediately after estimation of a model. Using this for the test just performed, we obtain:

```
. test w=0

 ( 1)  [y]w = 0

           chi2(  1) =   94.05
         Prob > chi2 =    0.0000
```

This is a Wald test of the house money effect. The Wald test statistic is the square of the asymptotic t-test statistic $[94.05 = (-9.70)^2]$, and has a $\chi^2(1)$ distribution under the null hypothesis of no effect. The Wald test is equivalent to the asymptotic t-test and the two tests will always have the same p-value.

The next thing we might wish to do is to predict the probability of making the safe choice at each wealth level. We would do this in Excel. The formula we need to program is $\Phi(1.302 - 0.141w)$, and this is done in the Excel spreadsheet **house money calculations** which is presented below. Note that the intercept and slope estimates from the probit model appear at the far right, in cells E1 and E2 respectively, and these are the inputs to the algorithm. These two values are used in the column headed "P(SAFE)" to work out the probability of the safe choice at each wealth level appearing in the first column. The Excel function NORMSDIST gives the standard normal distribution function of the quantity in parentheses.

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | w | P(SAFE) | | b0: | 1.3017 |
| 2 | 0 | =NORMSDIST(E$1+E$2*A2) | | b1: | -0.141 |
| 3 | 1 | =NORMSDIST(E$1+E$2*A3) | | | |
| 4 | 2 | =NORMSDIST(E$1+E$2*A4) | | | |
| 5 | 3 | =NORMSDIST(E$1+E$2*A5) | | | |
| 6 | 4 | =NORMSDIST(E$1+E$2*A6) | | | |
| 7 | 5 | =NORMSDIST(E$1+E$2*A7) | | | |
| 8 | 6 | =NORMSDIST(E$1+E$2*A8) | | | |
| 9 | 7 | =NORMSDIST(E$1+E$2*A9) | | | |
| 10 | 8 | =NORMSDIST(E$1+E$2*A10) | | | |
| 11 | 9 | =NORMSDIST(E$1+E$2*A11) | | | |
| 12 | 10 | =NORMSDIST(E$1+E$2*A12) | | | |
| 13 | 11 | =NORMSDIST(E$1+E$2*A13) | | | |
| 14 | 12 | =NORMSDIST(E$1+E$2*A14) | | | |
| 15 | 13 | =NORMSDIST(E$1+E$2*A15) | | | |

We may then plot the resulting predicted probabilities against the wealth level. The result is shown in Figure 6.4. We see that when the initial endowment is 0, there is a high probability that the safe alternative will be chosen, that is, subjects



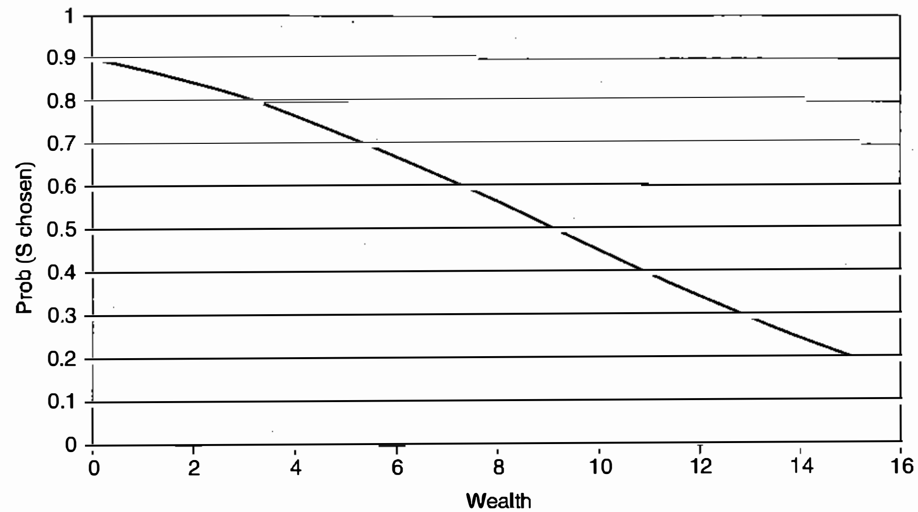



Figure 6.4: Predicted probabilities of safe choice against wealth level from the probit model

appear to be highly risk averse. We also see that as the initial endowment rises, the probability of choosing the safe alternative falls fairly steeply. Since a probability of 0.5 is associated with risk-neutrality, and, remembering that $\Phi^{-1}(0.5) = 0$, it appears that, in order to induce risk-neutrality in subjects, it is necessary to endow them with an amount $1.3016/0.1410 = \$9.23$. When the initial endowment is above this amount, risk-seeking behaviour is predicted, since the predicted probability of the safe choice is then lower than 0.5. (See Exercise 1 for further analysis of the "risk-neutral initial endowment".)

### 6.2.3 *Marginal effects*

Something else that is sometimes useful after estimating a probit model is to obtain conditional marginal effects. This is the predicted change in the probability resulting from a small change in the explanatory variable starting from a particular value. For example, if we wish to know how much the probability of $S$ changes when $w$ rises from 0, we use:

```
. margins, dydx(w) at(w=0)

Conditional marginal effects                      Number of obs   =       1050
Model VCE    : OIM

Expression   : Pr(y), predict()
dy/dx w.r.t. : w
at           : w               =           0

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |   -.024109    .0013299   -18.13   0.000    -.0267155   -.0215026
------------------------------------------------------------------------------
```

We see that the conditional marginal effect is −0.024, implying that, roughly speaking, if $w$ rises from 0 to 1, the probability of $S$ will fall by 2.4 percentage points. If we condition on a higher value of $w$, we obtain a different result:

```
. margins, dydx(w) at(w=10)

Conditional marginal effects                    Number of obs   =       1050
Model VCE    : OIM

Expression   : Pr(y), predict()
dy/dx w.r.t. : w
at           : w               =          10

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.0559177    .0053804   -10.39   0.000    -.0664631   -.0453724
------------------------------------------------------------------------------
```

This higher (in magnitude) marginal effect (−0.056) simply reflects the fact that the curve shown in Figure 6.4 is steeper at w=10 than at w=0. Finally, if we use the `margins` command without the `at( )` option, we obtain the *average marginal effect*.

```
. margins, dydx(w)

Average marginal effects                        Number of obs   =       1050
Model VCE    : OIM

Expression   : Pr(y), predict()
dy/dx w.r.t. : w

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.0444259    .0039929   -11.13   0.000    -.0522518      -.0366
------------------------------------------------------------------------------
```

The average marginal effect is seen to be −0.044. This is simply the average of the marginal effects over all of the observations in the sample.

### 6.2.4 Wald tests and LR tests

The test used in Section 6.2.2 for the significance of the variable $w$ was a Wald test. It was demonstrated that this test can be conducted using the `test` command, and the Wald test statistic is the square of the asymptotic t-test statistic.

There is yet another way of testing the same hypothesis: the likelihood ratio (LR) test. This test is based on a comparison of the maximised log-likelihood in two different models. The test statistic is computed using:

$$LR = 2(LogL_U - LogL_R) \tag{6.15}$$

where $LogL_U$ is the maximised log-likelihood from the unrestricted model, and $LogL_R$ is the same for the restricted model. In the present case, the unrestricted model is the model that has been estimated (probit model with $w$), while the restricted model is a probit model with $w$ removed, that is, a model with an intercept only. Estimation of this restricted model gives:

```
. probit y

Iteration 0:   log likelihood = -634.4833
Iteration 1:   log likelihood = -634.4833

Probit regression                                 Number of obs   =       1050
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -634.4833                        Pseudo R2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5464424   .0408516    13.38   0.000     .4663746    .6265101
------------------------------------------------------------------------------
```

We see that the restricted log-likelihood is −634.48. This is used to compute the LR test statistic using (6.15):

$$LR = 2(LogL_U - LogL_R) = 2(-584.59 - (-634.48)) = 99.8 \tag{6.16}$$

Under the null hypothesis of no house money effect, the statistic given by (6.16) comes from a $\chi^2(1)$ distribution. Therefore we reject the null because $99.8 > \chi^2_{1,0.05} = 3.84$.

In fact, there is a way of computing the LR test statistic directly in STATA. The estimates from the two models are stored, and then the `lrtest` command is applied. The required sequence of commands, and the results, are as follows:

```
probit y w
est store with_w

probit y
est store without_w

lrtest with_w without_w

Likelihood-ratio test                                 LR chi2(1)  =     99.80
(Assumption: without_w nested in with_w)              Prob > chi2 =    0.0000
```

Reassuringly the result is exactly the same as (6.16). An advantage of using STATA to perform the test is that a p-value is provided in addition to the test statistic. In this case the p-value (0.0000) conveys overwhelming evidence of the house money effect.

Finally, note that the LR test statistic (99.80) is fairly close to the Wald test statistic (94.05) for the same hypothesis. This similarity is not surprising since the two tests are asymptotically equaivalent.

### 6.2.5 Analysis of ultimatum game data

The ultimatum game, introduced by Güth et al. (1982), was described in Section 2.5. The reader is referred back to that section for a reminder of the structure of the game.

The file **ug_sim** contains (simulated) data from 200 subjects who participated in an ultimatum game, in which the size of the pie is 100 units. Each subject plays twice, once as proposer, and once as responder, with a different opponent each time. The variables are:

i: proposer ID;
j: responder ID;
male_i: 1 if proposer is male; 0 otherwise;
male_j: 1 if responder is male; 0 otherwise;
y: proposer's offer;
d: responder's decision: 1 if accept; 0 if reject.

In Section 3.6, we analysed the proposer's offers in this data set, and we tested for a gender effect. In this section, we will turn to the responder's decision. This is a binary decision, so binary data models are required to identify its determinants.

We first consider simply how many of the subjects rejected offers. For this we obtain a tabulation of the binary variable, from which we see that 51 of the 200 subjects (approximately one-quarter of them) rejected offers.

```
. tab d

          d |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         51       25.50       25.50
          1 |        149       74.50      100.00
------------+-----------------------------------
      Total |        200      100.00
```

The main determinant of the responder's decision is the proposer's offer (y). Sometimes it is useful to plot binary data. The command `lowess d y` produces the graph shown in Figure 6.5. Lowess (locally weighted scatter-plot smoother)

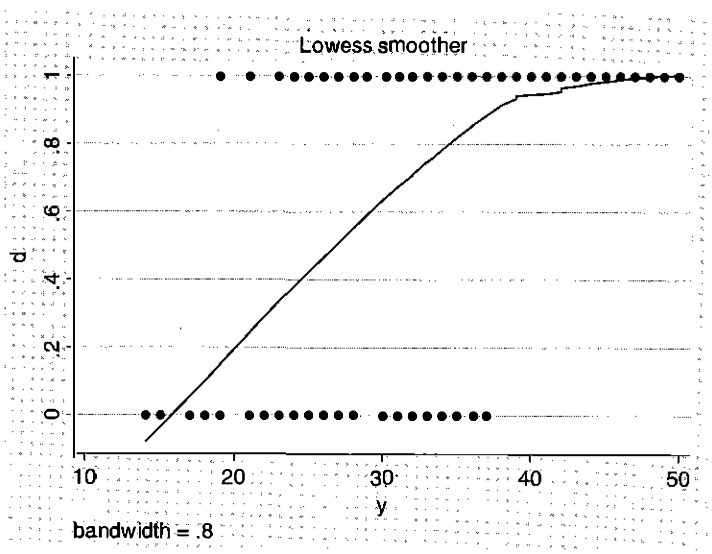



Figure 6.5: Responder's decision (d) against proposer's offer (y), with smoother



is a form of non-parametric regression that has been used previously in the book. Roughly speaking, it shows the mean value of $d$ conditional on different values of $y$. Since the mean of $d$ is closely related to the probability of the offer being accepted, the graph is telling us that the probability of acceptance rises sharply as the offer rises, approaching 1 as the offer approaches 50.

In complete contrast to "Lowess", the probit model introduced in Section 6.2.2 is an example of a fully parametric estimation procedure. The probit model is defined as follows:

$$P\,(d = 1|y) = \Phi\,(\beta_0 + \beta_1 y) \tag{6.17}$$

where $\Phi(.)$ is the standard normal cdf. The results are as follows:

```
. probit d y

Iteration 0:   log likelihood = -113.55237
Iteration 1:   log likelihood = -70.230335
Iteration 2:   log likelihood = -66.806698
Iteration 3:   log likelihood = -66.738058
Iteration 4:   log likelihood = -66.738049
Iteration 5:   log likelihood = -66.738049

Probit regression                                 Number of obs   =        200
                                                  LR chi2(1)      =      93.63
                                                  Prob > chi2     =     0.0000
Log likelihood = -66.738049                       Pseudo R2       =     0.4123

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           y |   .1439157   .0212804     6.76   0.000     .1022069    .1856244
       _cons |  -3.855266    .631443    -6.11   0.000    -5.092872   -2.617661
------------------------------------------------------------------------------
```

From the results, we can deduce a formula for the predicted probability of an offer ($y$) being accepted:

$$\hat{P}\,(d = 1|y) = \Phi\,(-3.855 + 0.144y) \tag{6.18}$$

In this situation, it is useful to consider the equation *underlying* the probit model:

$$d^* = \beta_0 + \beta_1 y + \epsilon \qquad \epsilon \sim N\,(0, 1) \tag{6.19}$$

In (6.19), $d^*$ is the propensity of the responder to accept the offer. If this propensity is greater than zero, the offer is accepted:

$$d = 1 \Leftrightarrow d^* > 0 \Leftrightarrow \beta_0 + \beta_1 y + \epsilon > 0 \Leftrightarrow \epsilon > -\beta_0 - \beta_1 y \tag{6.20}$$

Hence the probability of the offer being accepted is:

$$P\,(d = 1) = P\,(\epsilon > -\beta_0 - \beta_1 y) = \Phi\,(\beta_0 + \beta_1 y) \tag{6.21}$$

which is the probability formula (6.17) on which the probit model is based. The reason why (6.19) is useful is because it enables us to compute the "minimum acceptable offer (MAO)" for a typical subject. Disregarding the error term, we have:

$$d^* = \beta_0 + \beta_1 y \tag{6.22}$$

A typical subject is indifferent between accepting and rejecting an offer when (6.22) is zero:

$$\beta_0 + \beta_1 y = 0 \Rightarrow y = -\frac{\beta_0}{\beta_1} \tag{6.23}$$

We compute this from the estimates as follows:

$$y^{MAO} = -\frac{-3.855}{0.144} = \underline{\underline{26.79}} \tag{6.24}$$

The MAO (6.24) can also be computed in STATA with the nlcom command:

```
. nlcom MAO: -_b[_cons]/_b[y]

        MAO:  -_b[_cons]/_b[y]

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         MAO |   26.78837   .9268278    28.90   0.000     24.97182    28.60492
------------------------------------------------------------------------------
```

The `nlcom` procedure uses a technique known as the "delta method" which is considered in more detail in Section 6.5. The major benefit from applying this technique is that it returns a standard error and confidence interval for MAO, in addition to the point estimate. The point estimate of 26.79 is telling us that a "typical" responder (typical in the sense of having an error term $\epsilon$ equal to the mean value of 0) would say "no" to the offer of 26, but "yes" to 27.

### 6.2.6 The strategy method

The strategy method was explained in Section 2.6.2. It has been used in the context of the ultimatum game by Solnick (2001), among others. The proposer makes an offer as before. Let this offer be $y$. Meanwhile, in a different room, the responder is asked to state their minimum acceptable offer ($y^{MAO}$). Then y is compared to $y^{MAO}$. If $y \geq y^{MAO}$, the offer is taken as being accepted, both players receive their pay-offs. If $y < y^{MAO}$, the offer is taken as being rejected, and both players receive zero.

Under this approach, the responder is not only being asked for a decision, but for their *strategy*. Note that it is in their interest to state their MAO truthfully; for this reason, the strategy method is said to be *incentive compatible*.

The standard version of the ultimatum game (i.e. as described in Section 2.5.1) is known as the "direct decision approach". The strategy method has a considerable advantage over the direct decision approach. The data is much more informative. Clearly, it is more useful to know the responder's minimum acceptable offer than it is simply to know whether they have accepted a particular offer. This is particularly so in the cases where proposers offer 50% of their endowment. When a proposer offers 50%, the offer is almost certain to be accepted, and very little is learned, despite a significant cost to the experimenter. The strategy method enables useful information to be learned from all responders.



Let us imagine that the strategy method has been applied to the 200 subjects instead of the "direct decision approach", and that the data set consists of:

| | |
|---|---|
| i: | proposer ID; |
| j: | responder ID; |
| male_i: | 1 if proposer is male; 0 otherwise; |
| male_j: | 1 if responder is male; 0 otherwise; |
| y: | proposer's offer; |
| MAO: | responder's minimum acceptable offer; |
| d: | outcome: 1 if $y \geq y^{MAO}$; 0 if $y < y^{MAO}$. |

A (simulated) data set containing these variables is contained in the file **ug_sm_sim**. With this data, we carry out the following simple analysis:

```
. ci MAO

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
         MAO |        200      31.375    .6666664        30.06036    32.68964

. tab d

          d |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         87       43.50       43.50
          1 |        113       56.50      100.00
------------+-----------------------------------
      Total |        200      100.00
```

The straightforward command `ci MAO` has been used to obtain a 95% confidence interval for the population mean of $y^{MAO}$. This confidence interval is clearly narrower than the one obtained in Section 6.2.5 from the direct-decision data (24.97182 → 28.60492). This simply confirms that we are able to estimate parameters more precisely when the strategy method has been used.

However, note also that the MAO appears to be around five units higher when the strategy method is being used. This has the consequence that the number of "rejections" is higher (87, compared with 51 under the direct-decision approach). This is a common finding. Eckel & Grossman (2001) explain the higher rejection rate under the strategy method in terms of subjects' failure to understand the simultaneous nature of the decision and their attempt to signal a "tough" bargaining position.

There may also be a perception on the part of the responder that their statement of MAO is *hypothetical* (even though it is not, since it determines their pay-off). Asking a subject about how they *would* act in different situations is sometimes referred to as a "cold" treatment, to be contrasted with the "hot" treatment that arises when an offer is actually placed in front of the responder, and all they need to do is "accept".

The message here is perhaps that superior data should be obtained using the strategy method, but that an adjustment should be applied to the MAO data in order for it to be applicable to the "direct decision" situation. On the evidence above, the stated values of MAO would need to be reduced by around five units.

## 6.3 The `ml` Routine in STATA

There is another way to estimate the probit model. This is to specify the log-likelihood function ourselves, and ask STATA to maximise it. We shall return to the **house_money_sim** data set used in Section 6.2.2.

The following code defines a program called `myprobit` which computes the log-likelihood. It then reads the data and calls on the `ml` program to perform the maximisation of the log-likelihood. The formula being programmed is the one given in (6.14) above:

$$LogL = \sum_{i=1}^{n} ln\left(\Phi\left(yy_i \times (\beta_0 + \beta_1 w_i)\right)\right) \qquad (6.25)$$

```
* LOG-LIKELIHOOD EVALUATION PROGRAM "myprobit" STARTS HERE:

program define myprobit

* SPECIFY NAME OF QUANTITY WHOSE SUM WE WISH TO MAXIMISE (logl)
* AND ALSO PARAMETER NAMES (EMBODIED IN xb)
* PROVIDE LIST OF TEMPORARY VARIABLES (p ONLY)

args logl xb
tempvar p

* GENERATE PROBABILITY OF CHOICE MADE BY EACH SUBJECT (p):

quietly gen double `p'=normal(yy*`xb')

* TAKE NATURAL LOG OF p AND STORE THIS AS logl

quietly replace `logl'=ln(`p')

* END "myprobit" PROGRAM:

end

* READ DATA

use "house money_sim", clear

* GENERATE (INTEGER) yy FROM y:

gen int yy=2*y-1

* SPECIFY LIKELIHOOD EVALUATOR (lf), EVALUATION PROGRAM (myprobit),
* AND EXPLANATORY VARIABLE LIST.
* RUN MAXIMUM LIKELIHOOD PROCEDURE

ml model lf myprobit ( = w)
ml maximize
```

The line `args logl xb` is important. It indicates that the quantity that we wish to maximise is the sum over the sample of the variable named `logl`, and that the parameters with respect to which we wish to maximise it are implicit in the variable `xb`, which corresponds to $\beta_0 + \beta_1 w$ in the formula. `logl` and `xb` are examples of "local variables", being variables which exist within the program but not outside it. Any other local variables need to be declared by the `tempvar` command. Whenever temporary variables are referred to within the program, they need to be placed inside a particular set of quotation marks:

```
`p'
```

The quote before the p is the left single quote; you will find it in the upper left corner of most keyboards, below the "escape" key. The quote after the p is the right single quote; you will find it somewhere near the "enter" key.

Variables appearing without quotes are "global" variables, meaning that they also exist outside the program. In this example, yy (the binary dependent variable) is a global variable.

The last two lines of the above code are the lines that cause the program to run. The `ml` command specifies that the `lf` likelihood evaluator will be used. `lf` stands for "linear form", which essentially means that the likelihood evaluation program returns one log-likelihood contribution for each row of the data set. A situation in which the linear form restriction is not met is in the context of a panel data model, for which the likelihood evaluation program will return one contribution for each *block* of rows. In such a situation the *d-family* evaluators are required in place of `lf`. These will be introduced at an appropriate time in later chapters.

The results from running the above code are:

```
. ml model lf myprobit ( = w)
. ml maximize

initial:       log likelihood = -727.80454
alternative:   log likelihood =  -635.1321
rescale:       log likelihood =  -635.1321
Iteration 0:   log likelihood =  -635.1321
Iteration 1:   log likelihood = -584.84039
Iteration 2:   log likelihood = -584.58503
Iteration 3:   log likelihood = -584.58503

                                              Number of obs   =       1050
                                              Wald chi2(1)    =      94.05
Log likelihood = -584.58503                   Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1409882   .0145377    -9.70   0.000    -.1694816   -.1124948
       _cons |   1.301654   .0911155    14.29   0.000     1.123071    1.480237
------------------------------------------------------------------------------
```

Note that the results are identical to the results obtained using the command `probit y w` presented in Section 6.2.2.

See Exercise 1 for more practice with the `ml` routine.

## 6.4 Structural Modelling

In the present context, a structural model is one that is expressed in terms of the individual's utility function. The models analysed in Section 6.2 were not structural; they simply attempted to explain the data. Here, and in the next section, we consider the estimation of some simple structural models.

Firstly, let us assume that all individuals have the same utility function, which is:

$$\begin{aligned} U(x) &= \frac{x^{1-r}}{1-r} \qquad r \neq 1 \\ &= ln(x) \qquad r = 1 \end{aligned} \tag{6.26}$$

Equation (6.26) is known as the constant relative risk aversion (CRRA) utility function, because the parameter $r$ is the coefficient of relative risk aversion:[2] the higher is $r$, the more risk averse the subject. Note that $r$ can be negative, indicating risk-seeking.

Secondly, let us assume that individuals maximise expected utility. We shall continue to use the "house money effect" example introduced in Section 6.2.2. The expected utilities from choosing the safe and risky choices are:

$$EU(S) = \frac{(w+5)^{1-r}}{1-r} \tag{6.27}$$

$$EU(R) = 0.5\frac{(w)^{1-r}}{1-r} + 0.5\frac{(w+10)^{1-r}}{1-r} \tag{6.28}$$

Thirdly, let us assume that when an individual computes the expected utility difference, they make a computational error $\epsilon$, where $\epsilon \sim N(0, \sigma^2)$. This type of error has come to be known in risk modelling as a *Fechner error term* after Fechner (1860).

Given these three assumptions, the safe choice is made if:

$$EU(S) - EU(R) + \epsilon > 0$$

where $EU(S)$ and $EU(R)$ are defined in (6.27) and (6.28) respectively.

The probability of the safe choice being made is therefore:

$$\begin{aligned} P(S) &= P\left[EU(S) - EU(R) + \epsilon > 0\right] \\ &= P\left[\epsilon > EU(R) - EU(S)\right] \\ &= P\left[\frac{\epsilon}{\sigma} > \frac{EU(R) - EU(S)}{\sigma}\right] \\ &= 1 - \Phi\left[\frac{EU(R) - EU(S)}{\sigma}\right] \\ &= \Phi\left[\frac{EU(S) - EU(R)}{\sigma}\right] \end{aligned} \tag{6.29}$$

Substituting (6.27) and (6.28) into (6.29), and using the "yy trick" introduced in Section 6.2.2, the log-likelihood function may be written:

[2] See Section 12.2 for a complete definition of CRRA.



$$LogL = \sum_{i=1}^{n} ln\Phi\left[yy_i \times \frac{\frac{(w_i+5)^{1-r}}{1-r} - \left(0.5\frac{(w_i)^{1-r}}{1-r} + 0.5\frac{(w_i+10)^{1-r}}{1-r}\right)}{\sigma}\right] \tag{6.30}$$

We maximise (6.30) to obtain estimates of the two parameters $r$ and $\sigma$. The challenge is that there is no STATA command that does this for us. We need to program it and use the `ml` command.

The required program, and the commands required to run the program, are as follows. For infomation about the syntax, the reader should refer back to the example provided in Section 6.3 in which each step was explained.

```
program drop structural
program structural
args logl r sig
tempvar eus eur diff p

quietly gen double `eus'=(w+5)^(1-`r')/(1-`r')
quietly gen double `eur'=0.5*w^(1-`r')/(1-`r')+0.5*(w+10)^(1-`r')/(1-`r')
quietly gen double `diff'=(`eus'-`eur')/`sig'
quietly gen double `p'=normal(yy*`diff')
quietly replace `logl'=ln(`p')
end

ml model lf structural /r  /sig
ml maximize
```

The line `args logl r sig` is again important. Here, it indicates that the quantity we are seeking to maximise is named `logl`, and that the parameters with respect to which we wish to maximise it are `r` and `sig`. One difference from the code in Section 6.3 is that the two parameters (`r` and `sig`) are named in the `ml` command. This is appropriate because these two parameters are stand-alone parameters, unlike those in the example in Section 6.3 which were regression parameters. Providing parameter names in the `ml` command is useful because it causes the same names to be included in the results table:

```
. ml model lf structural /r  /sig

. ml maximize

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -601.45646
rescale:       log likelihood = -601.45646
rescale eq:    log likelihood = -600.78259
Iteration 0:   log likelihood = -600.78259
Iteration 1:   log likelihood =  -595.2424
Iteration 2:   log likelihood = -595.22797
Iteration 3:   log likelihood = -595.22739
Iteration 4:   log likelihood = -595.22739

                                                  Number of obs   =       1050
                                                  Wald chi2(0)    =          .
Log likelihood = -595.22739                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
r            |
       _cons |     .21765   .0976928     2.23   0.026     .0261757    .4091244
-------------+----------------------------------------------------------------
sig          |
       _cons |   .3585733   .1046733     3.43   0.001     .1534174    .5637292
------------------------------------------------------------------------------
```

We see that the following estimates are obtained for the two parameters:

$$\hat{r} = 0.2177$$
$$\hat{\sigma} = 0.3586$$

So, on the basis of the assumptions of this model, it appears that every individual is operating with the same utility function:

$$U(x) = \frac{x^{1-0.2177}}{1-0.2177} = \frac{x^{0.7823}}{0.7823}$$

and also that, when individuals compute the difference between the expected utilities of the two lotteries, they make a random computational error with mean zero and standard deviation 0.3586.

## 6.5 Further Structural Modelling

### *6.5.1 The heterogeneous agent model*

We continue to assume that subjects have the CRRA utility function:

$$U(x) = \frac{x^{1-r}}{1-r} \qquad r \neq 1$$

In Section 6.4, we assumed that all individuals had the same risk attitude, i.e. all had the same value of $r$. We attributed variation in choices to errors in the computation of expected utilities.

Here, we shall adopt a different approach. We shall assume (more realistically) that each subject has his or her own value of $r$, and we shall refer to the model as the "heterogeneous agent model". We just need to make an assumption about how $r$ varies over the population. An obvious choice is:

$$r \sim N\left(\mu, \sigma^2\right) \tag{6.31}$$

We ask each subject to make a choice between two lotteries, S and R. We shall use the popular Holt & Laury (2002) design, which is presented in Table 6.1.

In Table 6.1, there are ten problems listed in order. In Problem 1, we expect all subjects to choose $S$; in Problem 10, we expect all subjects to choose $R$ (in fact, $R$ stochastically dominates[3] in Problem 10). What is interesting is *where* in the sequence a subject switches from $S$ to $R$, since this will indicate their

[3] See Chapter 12 for a definition of stochastic dominance.



| Problem | Safe(S) | Risky(R) | $r^*$ |
|---|---|---|---|
| 1 | (0.1, \$2.00; 0.9, \$1.60) | (0.1, \$3.85; 0.9, \$0.10) | −1.72 |
| 2 | (0.2, \$2.00; 0.8, \$1.60) | (0.2, \$3.85; 0.8, \$0.10) | −0.95 |
| 3 | (0.3, \$2.00; 0.7, \$1.60) | (0.3, \$3.85; 0.7, \$0.10) | −0.49 |
| 4 | (0.4, \$2.00; 0.6, \$1.60) | (0.4, \$3.85; 0.6, \$0.10) | −0.15 |
| 5 | (0.5, \$2.00; 0.5, \$1.60) | (0.5, \$3.85; 0.5, \$0.10) | 0.15 |
| 6 | (0.6, \$2.00; 0.4, \$1.60) | (0.6, \$3.85; 0.4, \$0.10) | 0.41 |
| 7 | (0.7, \$2.00; 0.3, \$1.60) | (0.7, \$3.85; 0.3, \$0.10) | 0.68 |
| 8 | (0.8, \$2.00; 0.2, \$1.60) | (0.8, \$3.85; 0.2, \$0.10) | 0.97 |
| 9 | (0.9, \$2.00; 0.1, \$1.60) | (0.9, \$3.85; 0.1, \$0.10) | 1.37 |
| 10 | (1.0, \$2.00; 0.0, \$1.60) | (1.0, \$3.85; 0.0, \$0.10) | $\infty$ |

Table 6.1: The Holt and Laury design, with threshold risk aversion parameter for each choice problem

attitude to risk. The content of Table 6.1 is sometimes called a "multiple price list" (MPL).

In the fourth column of Table 6.1, a value $r^*$ is shown. This is known as the "threshold risk attitude" for the problem. It is the risk attitude (i.e. the coefficient of relative risk aversion) that would (assuming EU) make a subject indifferent between S and R for the choice problem. It can be worked out using Excel (see the spreadsheet: **risk aversion calculations**) as shown below.

| | | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | r: | -1.72 | -0.95 | -0.49 | -0.15 | 0.15 | 0.41 | 0.68 | 0.97 | 1.37 |
| 2 | prob of higher outcome: | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| 3 | | | | | | | | | | |
| 4 | x | U(x) | U(x) | U(x) | U(x) | U(x) | U(x) | U(x) | U(x) | U(x) |
| 5 | 0.1 | 0.000701 | 0.005754 | 0.021718 | 0.061561 | 0.166181 | 0.43566 | 1.495719 | 31.10848 | -6.33575 |
| 6 | 1.6 | 1.320193 | 1.282329 | 1.351925 | 1.492932 | 1.754216 | 2.236551 | 3.632189 | 33.80667 | -2.2713 |
| 7 | 2 | 2.422327 | 1.981408 | 1.885161 | 1.929686 | 2.120589 | 2.551266 | 3.901033 | 34.03374 | -2.0913 |
| 8 | 3.85 | 14.38415 | 7.105814 | 5.002075 | 4.098095 | 3.700179 | 3.754653 | 4.81058 | 34.70904 | -1.64126 |
| 9 | | | | | | | | | | |
| 10 | eu(S): | 1.430406 | 1.422145 | 1.511896 | 1.667634 | 1.937403 | 2.42538 | 3.82038 | 33.98832 | -2.1093 |
| 11 | eu(R): | 1.439046 | 1.425766 | 1.515825 | 1.676174 | 1.93318 | 2.427056 | 3.816122 | 33.98893 | -2.11071 |
| 12 | | | | | | | | | | |
| 13 | cert equiv (S): | 1.6478 7 | 1.687207 | 1.724713 | 1.761618 | 1.798329 | 1.835619 | 1.873599 | 1.91 33 | 1.954209 |
| 14 | cert equiv (R): | 1. 51519 | 1.689409 | 1.72772 | 1.76946 | 1.793719 | 1.83777 | 1.867081 | 1.914063 | 1.950689 |

As an example, if a subject chooses $S$ on problems 1–6, and chooses $R$ on problems 7–10, they are revealing that (assuming EU) their risk attitude ($r$) is somewhere between 0.41 and 0.68.

Here, we assume that each subject is only asked to solve one of the ten problems. Each problem is solved by ten subjects, so we have 100 subjects in total. The data is contained in the file **holtlaury_sim**.

Assume that subject $i$ is presented with a choice problem with threshold risk level $r_i^*$. Let $y_i = 1$ if S is chosen, and $y_i = 0$ if $R$ is chosen. The probability of subject $i$ choosing $S$ is (using the normal distribution of $r$ specified in (6.31)):

$$P(y_i = 1) = P\left(r_i > r_i^*\right) = P\left(z > \frac{r_i^* - \mu}{\sigma}\right) = P\left(z < \frac{\mu - r_i^*}{\sigma}\right)$$
$$= \Phi\left(\frac{\mu - r_i^*}{\sigma}\right) = \Phi\left(\frac{\mu}{\sigma} - \left(\frac{1}{\sigma}\right) r_i^*\right) \qquad i = 1, \ldots, n \quad (6.32)$$

In (6.32) we again have a probit model with dependent variable $y$. The explanatory variable is the threshold risk attitude for the problem being solved, $r^*$.

The intercept is $\frac{\mu}{\sigma}$ and the slope is $-\frac{1}{\sigma}$. Therefore from the probit estimates we are able to deduce estimates of $\mu$ and $\sigma$. This is done in STATA using the `nlcom` command, which is a way of implementing the delta method (see next sub-section).

The output from the probit model is as follows:

```
. probit y rstar

Iteration 0:   log likelihood = -68.994376
Iteration 1:   log likelihood = -32.754689
Iteration 2:   log likelihood = -31.899974
Iteration 3:   log likelihood = -31.896643
Iteration 4:   log likelihood = -31.896643

Probit regression                                 Number of obs   =        100
                                                  LR chi2(1)      =      74.20
                                                  Prob > chi2     =     0.0000
Log likelihood = -31.896643                       Pseudo R2       =     0.5377

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rstar |  -1.826082   .3481266    -5.25   0.000    -2.508398   -1.143767
       _cons |   .7306556   .2264169     3.23   0.001     .2868867    1.174424
------------------------------------------------------------------------------
Note: 10 failures and 0 successes completely determined.

.
. nlcom (mu: -_b[_cons]/_b[rstar]) (sig: -1/_b[rstar])

          mu:  -_b[_cons]/_b[rstar]
         sig:  -1/_b[rstar]

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          mu |    .400122   .0978294     4.09   0.000     .2083799    .5918641
         sig |   .5476205    .104399     5.25   0.000     .3430021    .7522389
------------------------------------------------------------------------------
```

Note that we have estimates of $\mu$ and $\sigma$. But this time, as we have estimated a heterogeneous agent model, the interpretation is as follows: every individual has a different "coefficient of relative risk aversion", drawn from the following distribution:

$$r \sim N\left(0.4001, 0.5476^2\right)$$



Having drawn their risk aversion parameter, they use this in the expected utility calculation, which they perform without error.

### 6.5.2 *The delta method*

The delta method (`nlcom` in STATA) is used to obtain *standard errors* of the estimates of $\mu$ and $\sigma$ in (6.32).

Let the probit estimates be $\hat{\beta}$ and $\hat{\alpha}$. We might refer to these estimates as the *reduced form* estimates. Using STATA, we can obtain an estimate of the variance matrix of these estimates:

$$\hat{V}\begin{pmatrix}\hat{\beta}\\ \hat{\alpha}\end{pmatrix} = \begin{pmatrix} var\left(\hat{\beta}\right) & cov\left(\hat{\alpha}, \hat{\beta}\right) \\ cov\left(\hat{\alpha}, \hat{\beta}\right) & var\left(\hat{\alpha}\right) \end{pmatrix} \tag{6.33}$$

The square roots of the diagonal elements of this matrix are the standard errors that we see in the STATA output from the `probit` command.

If you wanted to see $\hat{V}$ having estimated the probit model, you would do it as follows, and this is what you would see:

```
. mat V=e(V)
. mat list V

symmetric V[2,2]
                   y:          y:
               rstar       _cons
y:rstar    .12119211
y:_cons   -.04842685   .05126459
```

The parameters that we are interested in are functions of $\alpha$ and $\beta$.

$$\alpha = \frac{\mu}{\sigma}; \quad \beta = -\frac{1}{\sigma} \Rightarrow \mu = -\frac{\alpha}{\beta}; \quad \sigma = -\frac{1}{\beta} \tag{6.34}$$

We would refer to $\mu$ and $\sigma$ as the *structural parameters*, being parameters of the utility function underlying behaviour.

We require the matrix D, where:

$$D = \begin{pmatrix} \frac{\partial\mu}{\partial\beta} & \frac{\partial\mu}{\partial\alpha} \\ \frac{\partial\sigma}{\partial\beta} & \frac{\partial\sigma}{\partial\alpha} \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{\beta^2} & -\frac{1}{\beta} \\ \frac{1}{\beta^2} & 0 \end{pmatrix} \tag{6.35}$$

Let $\hat{D}$ be the matrix D with parameters replaced by MLEs. The variance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is:

$$\hat{V}\begin{pmatrix}\hat{\mu}\\ \hat{\sigma}\end{pmatrix} = \hat{D}\left[\hat{V}\begin{pmatrix}\hat{\beta}\\ \hat{\alpha}\end{pmatrix}\right]\hat{D}' \tag{6.36}$$

The required standard errors are the square roots of the diagonal elements of this matrix.

Note that the delta method is applied using the `nlcom` command in STATA. This command will be used again. The example presented at the end of Section 6.5.1 makes clear the required syntax of the command.

### 6.5.3 *Another example using the delta method*

In Section 6.2.1, we considered the problem of estimating the proportion ($p$) of red balls in an urn based on a sample of balls drawn with replacement. In the example analysed, ten balls were drawn, of which seven were red. We derived the MLE of $p$ to be 0.7, and we also derived the asymptotic standard error of this estimate to be 0.145.

In this sub-section, we will demonstrate how it is possible to estimate this parameter and its standard error using the probit model. The dependent variable in the probit model is the binary variable indicating whether ($y = 1$) or not ($y = 0$) the drawn ball is red. We therefore require a column named $y$ containing seven ones and three zeros, in any order. We then estimate the probit model with *no explanatory variables*, that is, one in which only an intercept, $\beta_0$, is estimated:

$$P(y = 1) = \Phi(\beta_0) \tag{6.37}$$

The results obtained are as follows:

```
. probit y

Iteration 0:   log likelihood = -6.108643
Iteration 1:   log likelihood = -6.108643

Probit regression                               Number of obs   =         10
                                                LR chi2(0)      =       0.00
                                                Prob > chi2     =          .
Log likelihood = -6.108643                      Pseudo R2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5244005    .416787     1.26   0.208     -.292487    1.341288
------------------------------------------------------------------------------
```

Note that the estimate of the intercept is 0.5244. To deduce the parameter $p$ (the probability of a red ball), we need to apply (6.37) to this estimate. We may use the delta method to obtain this.

```
. nlcom p_hat: normal(_b[_cons])

       p_hat:  normal(_b[_cons])

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       p_hat |         .7   .1449138     4.83   0.000     .4159742    .9840258
------------------------------------------------------------------------------
```

Reassuringly, the delta method gives an estimate of $p$ of exactly 0.7, and also gives an asymptotic standard error of 0.145, the same as the one we obtained



using a different method in Section 6.2.1. Note also that the maximised log-likelihood (−6.1086) is the same as that which we see in Figure 6.2 (in which the log-likelihood function was plotted using Excel).

The parameter $p$ could also be estimated using the logit model with no explanatory variables. See Exercise 2.

## 6.6 Other Data Types

### 6.6.1 *Interval data: the interval regression model*

Let us return to the Holt & Laury (2002) design (Table 6.1). We continue to assume that subjects have the CRRA utility function:

$$U(x) = \frac{x^{1-r}}{1-r} \qquad r \neq 1$$

Recall that in the fourth column of Table 6.1, we show the value of $r$ (the coefficient of relative risk aversion) that would make a subject indifferent between the two lotteries. Recall also that in Problem 1 we expect all subjects to choose S; in Problem 10, we expect all subjects to choose R. In Section 6.5.1, we considered ways of estimating the distribution of $r$ over the population, when the available data consists of choices between pairs of lotteries.

In this section, we assume that the available information is more precise. We ask each subject to solve each choice problem in order, starting with Problem 1, thus revealing where in the list they switch from $S$ to $R$. Under the assumption of EU, knowledge of where a subject switches gives us an interval for $r$ for that subject. For example, an EU-maximising subject switching between Problems 5 and 6 is revealing that their coefficient of relative risk aversion is *between 0.15 and 0.41*.

The sort of data that results is known as "interval data". We are interested in the appropriate method for estimating the distribution of $r$ over the population when interval data is available.

The file **interval_data_sim** contains this information for 100 subjects (as well as information on subject characteristics).

As in Section 6.4, we assume that the distribution of $r$ over the population is:

$$r \sim N\left(\mu, \sigma^2\right) \tag{6.38}$$

For each subject, $i$, we have a lower bound ($l_i$) and an upper bound ($u_i$) for his or her $r$-value. The likelihood contribution for each subject is the probability of them being in the interval in which they are observed. So:

$$L_i = P(l_i < r < u_i) = P(r < u_i) - P(r < l_i) = \Phi\left(\frac{u_i - \mu}{\sigma}\right) - \Phi\left(\frac{l_i - \mu}{\sigma}\right) \tag{6.39}$$

So the sample log-likelihood is:

$$LogL = \sum_{i=1}^{n}\left[\Phi\left(\frac{u_i-\mu}{\sigma}\right) - \Phi\left(\frac{l_i-\mu}{\sigma}\right)\right] \tag{6.40}$$

Equation (6.40) is maximised to give MLEs of $\mu$ and $\sigma$. This is called the interval regression model (although at present there are no explanatory variables). To estimate it in STATA, use the command:

```
intreg  rlower rupper
```

where `rlower` and `rupper` are the variables containing the lower and upper bounds for each observation.

The results are as follows:

```
. intreg rlower rupper

Fitting constant-only model:

Iteration 0:   log likelihood = -199.07231
Iteration 1:   log likelihood = -198.96851
Iteration 2:   log likelihood = -198.96849

Fitting full model:

Iteration 0:   log likelihood = -198.96849
Iteration 1:   log likelihood = -198.96849

Interval regression                          Number of obs   =        100
                                             LR chi2(0)      =       0.00
Log likelihood = -198.96849                  Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .613146   .0597808    10.26   0.000     .4959777    .7303143
-------------+----------------------------------------------------------------
    /lnsigma |  -.5323404   .0764651    -6.96   0.000    -.6822092   -.3824716
-------------+----------------------------------------------------------------
       sigma |    .587229   .0449025                      .505499    .6821733
------------------------------------------------------------------------------

  Observation summary:         0  left-censored observations
                               0     uncensored observations
                               6 right-censored observations
                              94       interval observations
```

Estimates of the parameters of interest can be read directly. The distribution of risk-attitude over the population is estimated to be:

$$r \sim N\left(0.613, 0.587^2\right)$$

Next, suppose that we wish to allow risk attitude to vary according to subject characteristics. For example:



$$\begin{aligned} r_i &= \beta_0 + \beta_1 age_i + \beta_2 male_i + \epsilon_i \\ &= x_i'\beta + \epsilon_i \\ \epsilon_i &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{6.41}$$

In the second line of (6.41) we are adopting the convention of collecting all explanatory variables pertaining to observation $i$, including a constant, into the vector $x_i$. The vector $\beta$ contains the three corresponding parameters $\beta = (\beta_0 \quad \beta_1 \quad \beta_2)'$.

With this generalisation, (6.38) becomes:

$$r_i \sim N\left(x_i'\beta, \sigma^2\right) \tag{6.42}$$

and the log-likelihood function becomes:

$$LogL = \sum_{i=1}^{n} ln\left[\Phi\left(\frac{u_i - x_i'\beta}{\sigma}\right) - \Phi\left(\frac{l_i - x_i'\beta}{\sigma}\right)\right] \tag{6.43}$$

To estimate an interval regression model with explanatory variables, we do as follows:

```
. intreg rlower rupper age male

Fitting constant-only model:

Iteration 0:   log likelihood = -199.07231
Iteration 1:   log likelihood = -198.96851
Iteration 2:   log likelihood = -198.96849

Fitting full model:

Iteration 0:   log likelihood = -197.24143
Iteration 1:   log likelihood = -197.17109
Iteration 2:   log likelihood = -197.17108

Interval regression                          Number of obs   =        100
                                             LR chi2(2)      =       3.59
Log likelihood = -197.17108                  Prob > chi2     =     0.1657

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |     .02213   .0196956     1.12   0.261    -.0164727    .0607327
        male |  -.2165679   .1341118    -1.61   0.106    -.4794222    .0462864
       _cons |   .1592841   .4565128     0.35   0.727    -.7354646    1.054033
-------------+----------------------------------------------------------------
    /lnsigma |  -.5507208   .0764747    -7.20   0.000    -.7006085   -.4008332
-------------+----------------------------------------------------------------
       sigma |   .5765341   .0440903                     .4962832    .6697618
------------------------------------------------------------------------------

  Observation summary:         0  left-censored observations
                               0     uncensored observations
                               6 right-censored observations
                              94       interval observations
```

From these results, we now have an equation which determines the risk-aversion parameter for an individual of a given age and gender:

$$\hat{r}_i = 0.159 + 0.022age_i - 0.217male_i$$

However, note that neither of these explanatory variables appear to have significant effects on risk attitude. "Male" is close to being significant, and the negative sign would tell us that males are less risk-averse (or more risk-seeking) than females.

### 6.6.2 Continuous (exact) data

Yet another way of eliciting risk attitude is to present a subject with a single lottery, and ask them for their "certainty equivalent", that is, the amount of money such that they would be exactly indifferent between receiving this sum of money and playing the lottery.

For example, if the lottery is:

$$(0.3, \$3.85; 0.7, \$0.10)$$

and the subject claims that their certainty equivalent is \$0.75, then we can deduce that their coefficient of relative risk aversion is exactly 0.41. How do we know this? Because:[4]

$$0.3\frac{3.85^{1-0.41}}{1-0.41}+0.7\frac{0.10^{1-0.41}}{1-0.41}=\frac{0.75^{1-0.41}}{1-0.41}$$

A very important question is: how do you elicit a subject's certainty equivalent? You can simply ask them, and hope that they give an honest answer. But, according to some, there needs to be an *incentive* for the subject to report their certainty equivalent correctly. The scheme used to elicit the certainty equivalent needs to be *incentive compatible*.

One popular method for doing this, which is under reasonable assumptions incentive compatible, is the Becker-DeGroot-Marschak (BDM; Becker et al., 1964) incentive mechanism. BDM is described as follows. The individual is asked to place a valuation on a lottery (i.e. to report their certainty equivalent). They are told that after they have done this a random "price" will be generated. If the randomly generated price is higher than their reported valuation, they will be given an amount of money equal to this price, and they will not play the lottery; if the price is lower than their valuation, they will play the lottery.

In the file **exact_ data_sim**, we have the values of $r$ elicited in this way for 100 subjects. This is "exact" data in the sense that the value of $r$ is exactly observed. It is also "continuous" data as opposed to "discrete" (binary and interval data are both forms of "discrete" data). The distribution of $r$ over the sample of 100 subjects is shown in Figure 6.6.

We return to the assumption:

$$r \sim N\left(\mu, \sigma^2\right) \tag{6.44}$$

[4] This may be verfied easily using the Excel sheet "risk aversion calculations".



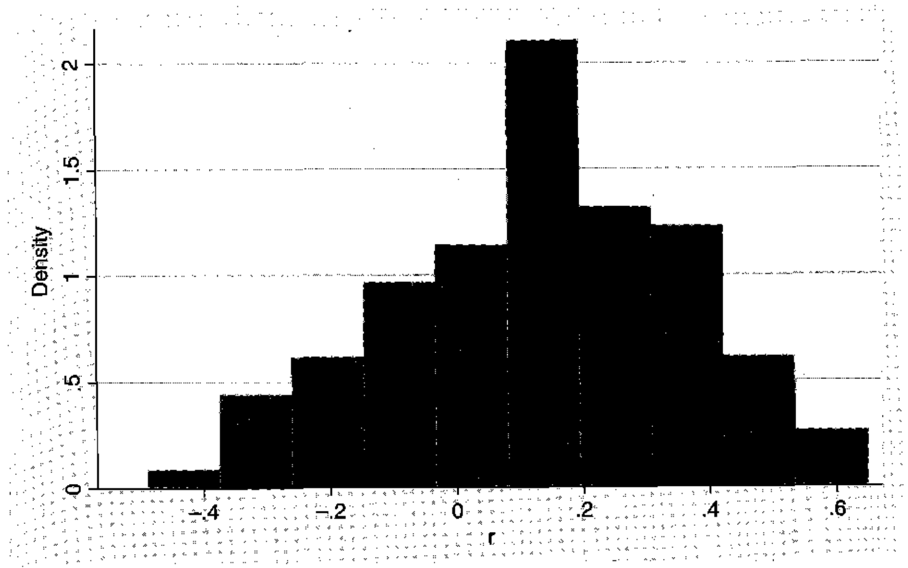



Figure 6.6: The distribution of r over 100 subjects

How do we estimate $\mu$ and $\sigma$ when exact data is available? First, let us consider what happens when we try to do this using maximum likelihood.

Consider the density associated with a particular observation $r_i$:

$$f\left(r_i;\mu,\sigma\right)=\frac{1}{\sigma\sqrt{2\pi}}exp\left(-\frac{\left(r_i-\mu\right)^2}{2\sigma^2}\right)=\frac{1}{\sigma}\phi\left(\frac{r_i-\mu}{\sigma}\right) \tag{6.45}$$

Equation (6.45) is a typical likelihood contribution, so the sample log-likelihood function is given by:

$$LogL=\sum_{i=1}^{n} ln\left[\frac{1}{\sigma}\phi\left(\frac{r_i-\mu}{\sigma}\right)\right] \tag{6.46}$$

To program (6.46), we would do as follows:

```
program define exact
args lnf xb sig
tempvar y p

quietly gen double `y'=$ML_y1
quietly gen double `p'=(1/`sig')*normalden((`y'-`xb')/`sig')
quietly replace `lnf'=ln(`p')
end

ml model lf exact (r= ) ()
ml maximize
```

The results are as follows:

```
. ml maximize

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -60.251905
rescale:       log likelihood = -7.5739988
rescale eq:    log likelihood =  3.1167494
Iteration 0:   log likelihood =  3.1167494
Iteration 1:   log likelihood =  3.2682025
Iteration 2:   log likelihood =  3.6372157
Iteration 3:   log likelihood =   3.637384
Iteration 4:   log likelihood =   3.637384

                                              Number of obs   =        100
                                              Wald chi2(0)    =          .
Log likelihood =   3.637384                   Prob > chi2     =          .

------------------------------------------------------------------------------
           r |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
       _cons |   .1340463   .0233327     5.74   0.000     .0883149    .1797776
-------------+----------------------------------------------------------------
eq2          |
       _cons |   .2333275   .0164987    14.14   0.000     .2009905    .2656644
------------------------------------------------------------------------------
```

and we see that the maximum likelihood estimates are:

$$\hat{\mu} = 0.134$$
$$\hat{\sigma} = 0.233$$

Of course, there is a much easier way of obtaining the MLEs of $\mu$ and $\sigma$ when exact data is available:

```
summ r
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           r |       100    .1340463    .2345029  -.4884877   .6499107
```

The maximum likelihood estimates of $\mu$ and $\sigma$ are just the sample mean and sample standard deviation respectively, of the variable $r$. The slight difference between the MLE of $\sigma$ and the sample standard deviation arises because the former uses $n$ as the divisor, while the latter uses $n - 1$. Asymptotically, the two will be equal.

Note that the main purpose of using ML above is to remind ourselves of the structure of the log-likelihood function in a situation in which the data are *continuous*. This is particularly important in the next sub-section, in which we consider *censored data*, which usually takes the form of a mixture of discrete and continuous data.

What is interesting about the results is that the estimate of $\mu$ (0.134) is much closer to zero than the estimates obtained using all other methods so far (which have varied between 0.400 and 0.613). This suggests that, while we know that most subjects are risk averse when choosing between lotteries, they tend towards risk neutrality when asked for certainty equivalents (when $r$=0, we would have risk neutrality). Another way of saying this is that when a subject is asked for a certainty



equivalent, there is a tendency for them to compute the *expected value* of the lottery, and to report something close to this.

The tendency towards risk-neutrality in valuation problems is an obvious explanation for the well-known "preference reversal" phenomenon discussed in Section 3.7.2. This is the tendency for subjects to prefer the safer lottery (the "P-bet") when asked to choose between them, but to place a higher valuation on the riskier lottery (the "$-bet).

### *6.6.3 Censored data: the Tobit model*

Tobit-type models, or censored regression models, originally due to Tobin (1958), are required when the dependent variable is censored, that is, when there is an accumulation of observations at the limits of the range of the variable. The lower limit of the range is usually zero, and censoring is usually "zero censoring", although sometimes we are required to deal with upper censoring, where there is an accumulation of observations at the maximum.

As an illustration of censored data, we shall use real data from a *public goods game*. The method of elicitation described here is an example of what is known as the voluntary-contribution mechanism (VCM), which was described in Section 2.5.4. The principal feature of such a setting is that while the unique Nash equilibrium is for all subjects in a group to "free ride", each contributing zero to the public fund, what is typically observed in experiments is a significant proportion of positive contributions along with a high proportion of zeros. The focus of investigations is usually to identify the motivations underlying positive contributions.

We will start by describing the data, and then we will explain the application of the Tobit model. The file **bardsley** contains data from 98 subjects who were arranged into 14 groups of $n = 7$, and each performed 20 tasks, in an experiment conducted by Bardsley (2000). An important feature of this experiment is that the seven subjects in a group take turns to contribute, and each subject observes the previous contributions by others. This is useful because it allows us to test directly for the motivation of "reciprocity" in contributions.

The variables to be used here are:

| | |
|---|---|
| y: | subject's contribution to the public fund; |
| med: | median of contributions by others in the group who have already contributed; |
| tsk: | task number. |

Figure 6.7 presents a histogram of contribution. Here we clearly see that there is a large accumulation of observations at the minimum possible contribution of zero, and a less prominent accumulation at the maximum possible contribution of 10. We say that contribution ($y$) is a "doubly censored" variable. This is because it is "censored from below" at 0, and "censored from above" at 10.

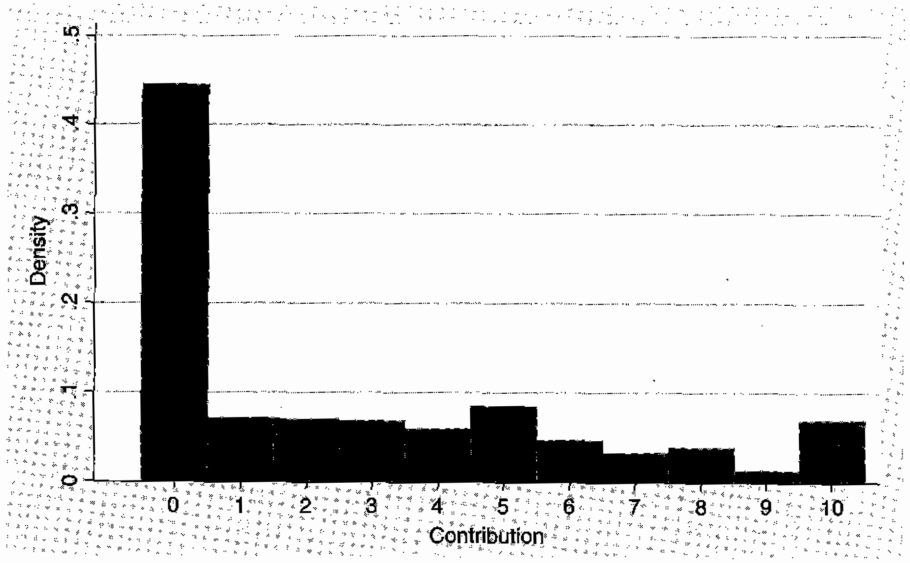



Figure 6.7: A histogram of contribution

Figure 6.8 presents scatterplots of contribution ($y$) against the median of others' previous contributions (med), and against task number (tsk).[5] The commands that have been used to obtain these graphs are:

```
lowess y med, msize(0.4) jitter(1) bwidth(0.7) xlabel(0(1)10)
lowess y tsk, msize(0.4) jitter(1) bwidth(0.7) xlabel(0(5)30)
```

The scatters themselves are not highly informative. One problem is that as a consequence of the discreteness in the variables, most combinations are represented by points. For this reason, we include "jitter"[6] in the scatterplot to make it possible at least to see which locations contain the most points. We also include Lowess smoothers, which plot the estimated conditional mean of contribution. The smoother in the left-hand plot makes it clear that previous contributions by others have a positive effect on contribution, providing non-parametric evidence of reciprocity. The smoother in the right-hand plot indicates a downward trend in contribution over the course of the experiment, and this is usually interpreted in terms of subjects "learning to play Nash". Next, we estimate models that allow us to conduct parametric tests of reciprocity and learning.

[5] Note that task number goes from 1 to 30 even though subjects only engage in 20 tasks. This is because the sequence of tasks also contained other tasks unrelated to the public goods experiment.

[6] A "jittered" scatterplot is a scatterplot in which the position of each point has been perturbed slightly. An alternative that has been used by experimental economists is the "bubble chart", in which the number of observations at a particular point is represented by the size of a circle drawn around that point. A possible drawback with bubble charts is we cannot be sure which dimension of a circle (e.g. radius, area) is being perceived by the viewer as representing its size. With jitter, there is no ambiguity because every observation appears as a distinct point in the plot.



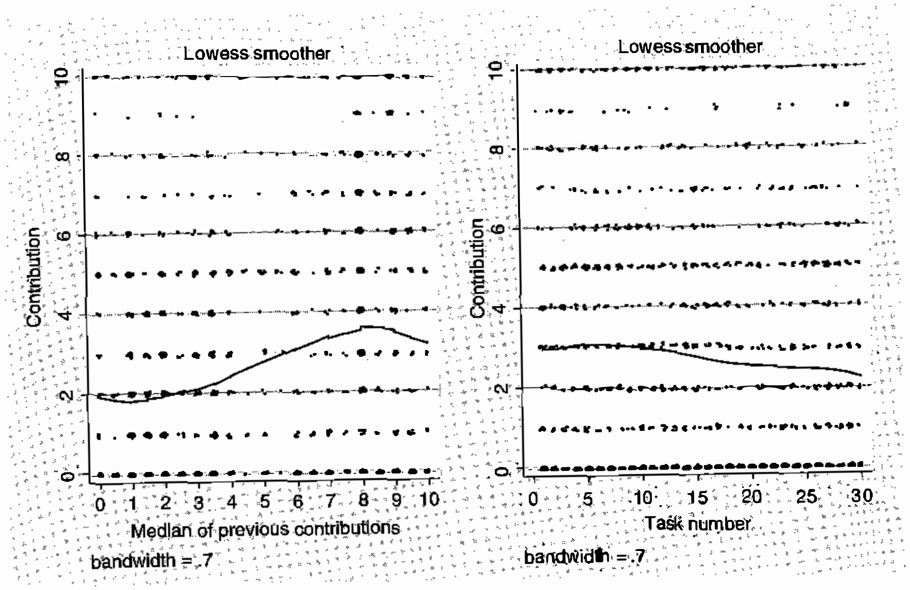



Figure 6.8: Left pane: contribution against median of previous contributions; right pane: contribution against task number; both with smoothers

Construction of the Tobit model commences with a linear specification for *desired* contribution. For the time being we shall disregard the panel structure of the data. This means that in the following regression model, the index $i$ simply represents observation number, not subject number. An obvious model incorporating both reciprocity and learning is:

$$\begin{aligned} y_i^* &= \beta_0 + \beta_1 med_i + \beta_2 tsk_i + \epsilon_i \\ &= x_i'\beta + \epsilon_i \\ \epsilon_i &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{6.47}$$

where $y_i^*$ is *desired* contribution underlying observation $i$.

The important feature of $y_i^*$ defined in (6.47) is that it can be negative, or it can be greater than the maximum possible contribution, i.e. the endowment. Subjects are permitted to *desire* to contribute a negative amount, or to *desire* to contribute an amount greater than their endowment. Of course, if a subject does desire to contribute a negative amount, most experimental designs would constrain them to contribute zero;[7] likewise, if they desire to contribute an amount that exceeds their endowment, their observed contribution is likely to be the amount of their

[7] An interesting recent development is the emergence of "take games", discussed previously in Section 3.7.3, dictator games in which some treatments allow dictators to take money away from the recipient, i.e. to "give" less than zero. See Bardsley (2008) and List (2007).

endowment. However, if they desire to contribute any amount between zero and the endowment, this amount will be their actual observed contribution.

These considerations amount to what are known as a *censoring rules*, defined as follows:

$$
\begin{aligned}
y_i &= 0 \text{ if } y_i^* \leq 0 \\
y_i &= y_i^* \text{ if } 0 < y_i^* < 10 \\
y_i &= 10 \text{ if } y_i^* \geq 10
\end{aligned}
\tag{6.48}
$$

where $y_i$ is the actual, observed, contribution for observation $i$.

We sometimes say that there are three *regimes*. To obtain the likelihood contributions, we consider each regime in turn:

$$y_i = 0: P(y_i = 0) = P(y_i^* \leq 0) = \Phi\left(-\frac{x_i'\beta}{\sigma}\right) \tag{6.49}$$

$$0 < y_i < 10: f(y_i) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - x_i'\beta)^2}{\sigma\sqrt{2\sigma^2}}\right) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \tag{6.50}$$

$$y_i = 10: P(y_i = 10) = P(y_i^* \geq 10) = \Phi\left(\frac{x_i'\beta - 10}{\sigma}\right) \tag{6.51}$$

Putting the three together, the sample log-likelihood becomes:

$$
\begin{aligned}
LogL = \sum_{i=1}^{n}\Bigg(I_{y_i=0}\ln\left[\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right] + I_{0<y_i<10}\ \ln\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] \\
+ I_{y_i=10}\ \ln\left[\Phi\left(\frac{x_i'\beta - 10}{\sigma}\right)\right]\Bigg)
\end{aligned}
\tag{6.52}
$$

$I_{(.)}$ is the indicator function, taking the value one if the subscripted expression is true, and zero otherwise.

The model developed in (6.47) through (6.52) is known as the two-limit tobit model (Nelson, 1976). It is estimated in STATA as follows:

```
. tobit y med tsk, ll(0) ul(10)

Tobit regression                                  Number of obs   =       1960
                                                  LR chi2(2)      =     109.34
                                                  Prob > chi2     =     0.0000
Log likelihood = -3821.3745                       Pseudo R2       =     0.0141

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         med |    .428854   .0441072     9.72   0.000      .342352    .5153559
         tsk |  -.0659412    .015992    -4.12   0.000    -.0973044    -.034578
       _cons |  -.1716389   .3612068    -0.48   0.635    -.8800291    .5367512
-------------+----------------------------------------------------------------
      /sigma |   5.692231   .1497394                      5.398566    5.985896
------------------------------------------------------------------------------
  Obs. summary:        872  left-censored observations at y<=0
                       952     uncensored observations
                       136 right-censored observations at y>=10
```



The strongly significant estimate of +0.429 for the slope associated with med strongly confirms the existence of the reciprocity phenomenon in this data. The fact that the estimate is smaller than one is consistent with the hypothesis of "biased reciprocity" – that subjects, although influenced positively by the contributions of others, tend to donate less than the levels contributed by others. Remember, however, that the model estimated here assumes that *all* agents are reciprocators, and the estimated effect is being diluted by the presence of subjects who are not responsive to others' contributions. When we come to estimate a "mixture model" in Chapter 8, in which we assume that only a proportion of the population are reciprocators, we will find that the effect of med is considerably larger.

The strongly negative coefficient associated with tsk confirms the existence of a learning process: through learning about the structure of the game, subjects contribute less as experience accumulates. We will discuss this learning effect much more when we return to the analysis of this data in Chapter 8.

Finally, we consider the consequences of disregarding the censoring. If we apply a straighforward linear regression, we obtain the following results:

```
. regress y med tsk

      Source |       SS       df       MS              Number of obs =    1960
-------------+------------------------------           F(  2,  1957) =   70.64
       Model |  1381.29031     2  690.645157           Prob > F      =  0.0000
    Residual |  19133.2627  1957  9.77683329           R-squared     =  0.0673
-------------+------------------------------           Adj R-squared =  0.0664
       Total |  20514.5531  1959  10.4719515           Root MSE      =  3.1268

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         med |   .2439314   .0218857    11.15   0.000     .2010097    .2868531
         tsk |  -.0343927   .0080305    -4.28   0.000    -.0501419   -.0186434
       _cons |   2.016418   .1779047    11.33   0.000     1.667515     2.36532
------------------------------------------------------------------------------
```

We see that there is a huge downward bias (i.e. bias towards zero) in the estimation of both the reciprocity parameter and the learning parameter, when the censoring is disregarded. For example, the OLS estimate of the reciprocity parameter is 0.24, which is almost 50% lower than the estimate of 0.43 obtained from Tobit.

The important point here is that the slopes, and therefore the extent of reciprocity and learning, are seriously under-estimated if the censoring at 0 and 10 is not taken into account. This may be explained with reference to Figure 6.8. Consider the left-hand scatterplot (contribution against med). The lower-censored observations (i.e. the zero contributions) are concentrated more highly at the left-hand end of the graph (i.e. at low levels of previous contributions), and the presence of these observations have the effect of pushing the lower end of the linear regression line upwards. The upper-censored observations (i.e. the contributions of ten) are concentrated at the right-hand side of the graph and their presence has the effect of pulling the top end of the regression line downwards. The combined effect of both groups of censored observations is a severe downward bias in the regression slope.

Throughout this sub-section, we have disregarded the panel structure of the data. Given that the data set consists of the decisions of 98 subjects in 20 tasks, the data set is a panel, and it is preferable to use a panel data model, of the type introduced in Chapters 4 and 5. The "panel tobit" estimator is implemented simply by using the command `xttobit` in place of `tobit`. This will be done in Section 10.5.

## 6.7 The Ultimatum Game: Further Analysis

### 6.7.1 Further tests of gender effects

In Section 3.6, we considered the effect of gender on the responder's offer in the ultimatum game, using simple treatment tests. Here we shall return to the same problem, but using more sophisticated techniques. We will use the **ug_sim** data again.

As obvious way of looking for a gender effect is using regression analysis. The important advantage of regression analysis is that it enables us to estimate different effects at the same time.

For example, we might wish to do the following. We start by generating a dummy variable indicating whether a male proposer is giving to a female responder.

```
. gen m_to_f=male_i*(1-male_j)

. regress y male_i male_j m_to_f

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  3,   196) =    3.37
       Model |  976.185392     3  325.395131           Prob > F      =  0.0195
    Residual |  18901.4946   196   96.436197           R-squared     =  0.0491
-------------+------------------------------           Adj R-squared =  0.0346
       Total |    19877.68   199  99.8878392           Root MSE      =  9.8202

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      male_i |  -4.519608   1.885099    -2.40   0.017     -8.23729   -.8019261
      male_j |   3.744608   2.074081     1.81   0.073    -.3457722    7.834988
      m_to_f |   2.381863    2.80275     0.85   0.396    -3.145557    7.909282
       _cons |     35.275   1.552709    22.72   0.000     32.21284    38.33716
------------------------------------------------------------------------------
```

These results tell us that:

1. Male proposers tend to offer 4.5 units less than female proposers, *ceteris paribus*.
2. Proposers tend to offer 3.7 units more when the responder is male, than when the responder is female, *ceteris paribus*. Note that this effect is only marginally significant.
3. Male proposers tend to offer 2.38 units more when the responder is female than when the responder is male, *ceteris paribus*. Eckel & Grossman (2001) refer to this effect as the"chivalry effect". Note that the effect is not statistically significant in this sample.

In consideration of conclusion 2, that proposers offer more when the responder is male, we might ask whether it is rational to do so. It is rational to offer more to male responders if males are more likely to reject offers. To see if this is the case, we go back to the probit model of Section 6.2.5, and add gender of responder as an explanatory variable in addition to proposer's offer. The results from doing so are:



```
. probit d y male_j

Iteration 0:   log likelihood = -113.55237
Iteration 1:   log likelihood = -68.373743
Iteration 2:   log likelihood = -64.187937
Iteration 3:   log likelihood = -64.116934
Iteration 4:   log likelihood = -64.116904
Iteration 5:   log likelihood = -64.116904

Probit regression                                 Number of obs   =        200
                                                  LR chi2(2)      =      98.87
                                                  Prob > chi2     =     0.0000
Log likelihood = -64.116904                       Pseudo R2       =     0.4354

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           y |   .1567836   .0231961     6.76   0.000       .11132    .2022472
      male_j |  -.5976406   .2668131    -2.24   0.025    -1.120585   -.0746966
       _cons |  -3.933341   .6589175    -5.97   0.000    -5.224796   -2.641886
------------------------------------------------------------------------------
```

We see that there is indeed evidence ($p = 0.025$) that a male is less likely to accept an offer of a given size than a female. So we may conclude that it is rational for the proposer to offer more to male responders.

The obvious follow-up question is: how much more should a proposer offer to a male responder than to a female responder, in order to create the same propensity for the offer to be accepted? We see that male responders' acceptance of propensity is lower by 0.598. This difference may be restored by increasing the offer by an amount 0.598/0.157. This computation may of course be carried out using the `nlcom` command:

```
. nlcom more_to_male: -_b[male_j]/_b[y]

more_to_male:  -_b[male_j]/_b[y]

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
more_to_male |   3.811882   1.612015     2.36   0.018     .6523915    6.971373
------------------------------------------------------------------------------
```

We see that a rational proposer should offer 3.81 more to a male than to a female, although the 95% confidence interval for this quantity is fairly wide.

It is interesting that this estimate of 3.81 for the additional amount that a rational proposer should give to a male in order to restore the acceptance probability is very close to the estimate of 3.74 obtained from the above regress command, of the additional amount that proposers actually do offer to males. It appears that proposers in this (simulated) data are indeed rational.

### 6.7.2 The proposer's decision as a risky choice problem

Let us remind ourselves of the relationship between the proposer's offer ($y$) and the responder's decision on whether to accept the offer ($d$). In Figure 6.5 we demonstrated that the probability of acceptance rises steeply with the amount of the offer, apparently reaching one when the offer reaches 50% of the endowment (recall that the total endowment was 100 tokens).

So, we could view the proposer's decision like this. If the proposer offers 50, he will keep 50 for himself with probability one; he will have a risk-free pay-off of 50. If he offers only 40, his pay-off will rise to 60, but he will receive this with probability less than 1, and otherwise he will receive zero. The lower the offer, the higher the possible pay-off, but the lower the probability of receiving this pay-off. Hence we see that the proposer's decision can be analysed as a risky choice problem. This is the approach taken by Roth et al. (1991), and others.

In Section 6.2.5, the probit model was used to obtain the following formula for the probability of any offer $y$ being accepted:

$$\hat{P}\,(d=1) = \Phi\,(-3.855 + 0.144y)$$

Let us assume that the proposer knows this probability formula. Note that this amounts to the assumption of rational expectations.

For the present purpose, let us treat the total endowment as one unit. So, if the proposer offers 50% of the endowment, their risk-free pay-off will be 0.5. If they offer 40%, their uncertain payoff will be 0.6, and so on.

The Excel sheet **proposer decision** contains the calculations necessary for the following analysis. If we assume a particular risk aversion parameter, say $r = 0.4$, then we have an expression for the proposer's expected utility from offering $y$:

$$EU\,(y) = \Phi\,(-3.855 + 0.144y) \times \left(\frac{100-y}{100}\right)^{1-0.4} \Big/ (1-0.4) \qquad (6.53)$$

Using (6.53), we can plot EU against each possible offer. This is done in Figure 6.9, from which we can see that the optimal offer for a proposer with $r = 0.4$ is 40. By

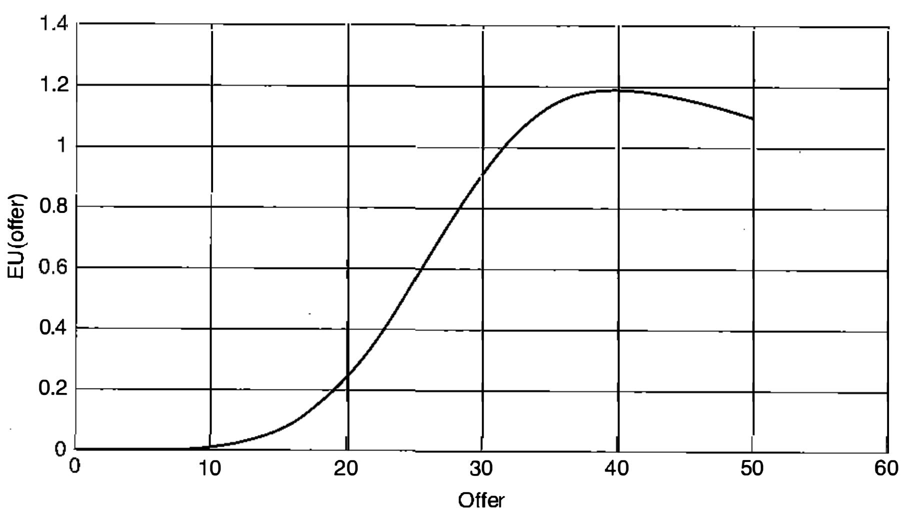



Figure 6.9: EU against offer with r = 0.4



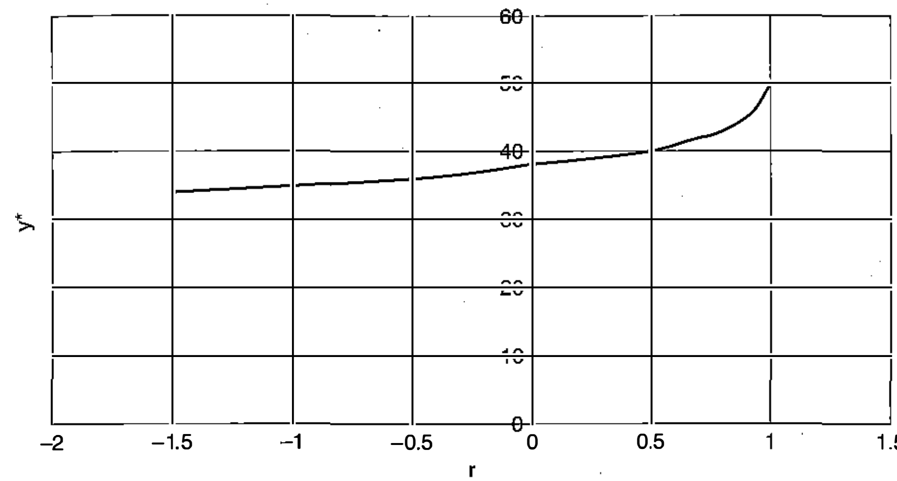



Figure 6.10: Optimal offer against r

repeating this exercise for different values of $r$, we can find the optimal offer for each risk attitude. The result is plotted in Figure 6.10.

We can then use Figure 6.10 to deduce a proposer's risk-aversion parameter from knowledge of their offer. For example, if they offer 47, their risk-aversion parameter must be 0.95 (they are quite risk averse).

Recall that 36 (18%) of the 200 proposers offered exactly 50% of their endowment. Should we attribute this behaviour to extreme risk aversion? Perhaps not. Individuals who offer 50% are likely to be doing so out of fairness considerations. They want to give 50% because they think it is the fair allocation, not because they are worried about having an offer rejected.

This sort of consideration leads us to a *mixture model*. Mixture models will play an important role in later chapters. One group from the population (around 18% it seems) are motivated by fairness and wish to share the endowment equally. The other 82% are motivated by self-interest and their degree of risk aversion dictates how much they offer, in accordance with the analysis above.

Extending this idea, it is interesting to take a close look at the subjects who are motivated by fairness (let us label them "egalitarians"; they are sometimes also referred to as "equal-splitters").

```
. gen egal=y==50

. tab egal

       egal |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        164       82.00       82.00
          1 |         36       18.00      100.00
------------+-----------------------------------
      Total |        200      100.00
```

As remarked above, 36 proposers offer 50%. Let us now investigate how these 36 divide by gender. To investigate whether egalitarianism is related to gender, we require a chi-squared test (see Chapter 3).

```
. tab male_i egal , chi2

           |         egal
    male_i |         0          1 |     Total
-----------+----------------------+----------
         0 |        66         25 |        91
         1 |        98         11 |       109
-----------+----------------------+----------
     Total |       164         36 |       200

          Pearson chi2(1) =  10.1506   Pr = 0.001
```

As we see in the table, 25 of the 91 females are egalitarian, while only 11 of the 109 males are so. The significance of this difference is summarised with the chi-squared test, and the accompanying $p$-value indicates a strongly significant relationship between gender and egalitarianism: females are significantly more likely than males to be egalitarian (according to this simulated data set).

## 6.8 Summary and Further Reading

This chapter has covered a number of data types commonly arising in experimental economics, and the ways in which they can be modelled. While the modelling of discreteness in experimental data has been the principal theme of this chapter, there have been a number of other themes. One has been maximum likelihood estimation and how to perform it in STATA. Another has been structural modelling, that is, the estimation of the parameters of the utility function. This has been done in the context of risky choice data, and then in the context of the proposer's decision in the ultimatum game.

A useful reference for binary data models is Cox (1970), and for censored and interval data, Maddala (1983). For more general aspects of ML and asymptotic theory, the reader is referred to Greene (2008). For information on how to use the ML command in STATA, the reader is referred to Gould et al. (2010).

One popular method of eliciting risk attitudes is the multiple price list (MPL, Holt & Laury, 2002), and methods for analysing the resulting data have been proposed in this chapter. Advantages and disadvantages of various types of MPL are discussed by Andersen et al. (2006).

The model considered in Section 6.7, in which the decision of the proposer in the ultimatum game is treated as a risky choice problem, is similar to the model of Roth et al. (1991).

We considered gender effects in the context of decisions in the ultimatum game. Similar tests were carried out in the same context by Eckel & Grossman (2001). For a more general review of gender effects in experimental economics, see Croson & Gneezy (2009).

An important technique used a number of times in the chapter is the delta method, on which further information may be found in Oehlert (1992).



## Exercises

1. In Section 6.2, the probit model for binary data was introduced, and in Section 6.3, the procedure was described for writing a program to maximise the probit log-likelihood function. An alternative to the probit model is the *logit* model, defined as:
$$P\left(y_i=1\right)=\frac{exp\left(\beta_0+\beta_1 w_i\right)}{1+exp\left(\beta_0+\beta_1 w_i\right)}$$
Consider again the **house_money_sim** data.
   (a) Estimate the logit model using the command:
   `logit y w.`
   (b) Test for the presence of a wealth effect.
   (c) Use Excel to predict the probability of choosing $S$ over the range of wealth levels.
   (d) Write a program in STATA that computes the logit log-likelihood, and maximise it using ML. It should give the same answer as (a).
   (e) Do logit and probit lead to similar conclusions?
   (f) In Section 6.2.2, the initial endowment necessary to induce risk-neutrality (i.e. a probability of 0.5 for the safe choice being chosen), under the probit model, was estimated to be \$9.23. Find an estimate of the same quantity using the logit results.
   (g) Using the delta method, find a standard error, and confidence interval, for the "risk-neutral initial endowment" computed in (f). Do this for both probit and logit. Which model estimates this quantity with greater precision?
2. In Section 6.2.1, we considered an example in which the objective was to estimate the proportion of red balls in an urn containing red and white balls. We considered a situation in which ten balls were drawn, of which seven were red, and another in which 100 balls were drawn, of which 70 were red. The following table summarises the results obtained there:

| Sample size | 10 | 100 |
|---|---|---|
| Number of reds | 7 | 70 |
| $\hat{p}$ | 0.7 | 0.7 |
| a.s.e.$(\hat{p})$ | 0.145 | 0.046 |
| Maximised LogL | −6.1086 | −61.086 |

In Section 6.5.3, we explained how these results could also be obtained by estimating a probit model with no explanatory variables, and then applying the delta method. There, we obtained the first column of the above table for the purpose of demonstration.

   (a) Obtain the second column of results in the above table using the probit model.

(b) All of the results in the table can also be obtained by using a logit model with no explanatory variables. Try this.
(c) Usually, the logit and probit models give results that are similar but not the same. In the present situation, you will see that they give results that are exactly the same. Why is this?

3. Consider the various choice models that have been introduced in this chapter. Now assume that in any task, with probability $\omega$, the individual loses concentration, and chooses randomly between the two alternatives. The parameter $\omega$ has come to be known as the "tremble probability". It is usually found to be between 0.01 and 0.10.

(a) Consider the Fechner model outlined in Section 6.4. That is, individuals all have CRRA utility, with the same risk aversion parameter, $r$. For any individual, the safe choice is made if:

$$EU\,(S) - EU\,(R) + \epsilon > 0$$

where $\epsilon$ is a Fechner error term, $\epsilon \sim N\left(0, \sigma^2\right)$. In Section 6.4, we derived the probability of the safe choice being made, and then constructed the sample log-likelihood. What is the probability of the safe choice in the presence of the tremble probability? Amend the ML program in Section 6.4 to include estimation of this parameter.
(b) Do the same for the heterogeneous agent model introduced in Section 6.5.1.
(c) Consider the public goods game analysed in Section 6.6.3. There, the decision variable was the number of tokens to contribute to the public fund ($y,\ 0 \leq y \leq 10$). How would you introduce a tremble parameter into this sort of model?

By the way, methods for introducing tremble parameters to models will be covered in much more detail in later chapters.

4. (a) Show that, for a discrete data model (e.g. a binary data model), the sample log-likelihood is always a negative quantity.
(b) Show that, for a model with continuous outcomes (e.g. a linear regression model), it is possible for the sample log-likelihood to be positive.

# Chapter 7
# Ordinal Data in Experimetrics

## 7.1 Introduction

In this chapter we are concerned with the modelling of ordinal, or ordered, data. Traditionally, this sort of data has arisen from attitudinal surveys, where responses are scored using a Likert scale (Likert, 1932). Such scales arise inevitably as a consequence of there being no natural unit of measurement for attitude.

Experimental economists have several reasons for being interested in ordinal data. For example, it has become popular recently to elicit data on emotions (fear, anger, sadness, etc.) experienced by subjects during an experiment, and to investigate the extent to which these emotions are driven by the subjects' experiences within the experiment. A natural means of eliciting such data is to ask subjects directly to rate their feelings, on a Likert scale, at the appropriate time in the experiment. The econometric objective would then be to use the resulting data to estimate the impact on emotions of the subject of their experiences in the experiment, perhaps controlling for individual characteristics (such as gender) which might also be expected to influence emotions. This will be demonstrated in Section 7.5.

Another experimental setting in which ordinal data arises is when subjects are asked how "sure" they are about a particular choice they have made. For example, a subject participating in a risky choice experiment might be asked to choose between a safe and a risky lottery, and then, having made the choice, they might be asked to indicate the degree of sureness with which they made the choice, by choosing between the alternatives: "not really sure"; "fairly sure"; "completely sure". This results in a Likert scale with six categories. Clearly, information on degrees-of-sureness potentially allows more precision in the estimation of the risk-attitude parameters than do straightforward binary choices. However, there is also a reason for believing that the binary data is more reliable than the sureness data: assuming real payoffs, the binary choice problem is incentive compatible; the report of degree of sureness is not, since there is no incentive to reveal truthfully one's degree of sureness. This leads us to consider if there is a way of testing the consistency between the binary choices and the ordinal reports of sureness. If such consistency is found, we would be able to conclude that subjects are reporting their degree of sureness truthfully despite the absence of a monetary incentive to do so. This and other issues are discussed in Section 7.6.

For various reasons, data on ordinal responses are awkward to handle statistically. The principal theme of the chapter is the *ordered probit model* which is suitable for the analysis of ordinal data, and is used in all of the examples. The central idea is that underlying the ordered response is a latent, continuously distributed random variable representing propensity to respond positively. The distributional parameters of this underlying latent variable are estimated using maximum likelihood, and these parameters have interpretations which can be useful to the investigator.

A common error is the failure to distinguish ordinal data from interval data. Interval data (as covered in Section 6.6.1) is data for which there is a natural unit of measurement, but the values are not fully observed; all that is known is the interval in which a value lies. As explained in Section 6.6.1, the appropriate method for the analysis of interval data is the interval regression model. The main practical difference from ordered probit is that, in interval regression, the "cut-point" parameters are known a priori, whereas in ordered probit, they need to be estimated.

Section 7.2 motivates the use of the ordered probit model by summarising the reasons why ordered outcomes require special treatment in statistical modelling. Section 7.3 provides a theoretical analysis of the model, including the construction of its log-likelihood function. Section 7.4 focuses on the so-called cut-points, and suggests a number of reasons why we might be particularly interested in the values taken by these parameters, which are usually treated as nuisance parameters. Section 7.5 covers an application to emotions data from an experiment. Section 7.6 covers an application in which experimental data is available on degree of sureness in risky choice.

## 7.2 Ordered Outcomes: The Case for Special Treatment

The attractions of the ordered probit model are most easily appreciated by considering the consequences of analysing ordered outcomes using linear regression techniques.

The first undesirable consequence of applying linear regression is that it implicitly assumes that, for example, the difference between a "very angry" response and a "moderately angry" response is the same as that between a "moderately angry" and a "slightly angry" response. There is no logical reason for expecting these differences to be the same, since the categories only reflect ordinality. The same assumption is not implicit in ordered probit.

This point is highly relevant because the interpretation that is given to a linear regression coefficient is in terms of the number of units by which we expect the dependent variable to change in response to a one-unit increase in an explanatory variable. Clearly, this interpretation is inappropriate if the dependent variable is ordinal.

Secondly, use of linear regression implicitly assumes that two subjects who give the same response have exactly the same attitude. This is not the case; a particular response is consistent with a range of attitudes. While differences in attitude for a given response are clearly unobservable, a model should allow for the fact that such differences exist.



A closely related issue is what are known as "floor" and "ceiling" distortions. For an example of the latter, if a respondent currently responds with the highest possible value, and their circumstances change in a way generally expected to increase response, their response will not increase; it will remain unchanged. The result of this is likely to be a bias towards zero in each regression coefficient.

Thirdly, since the responses to a question depend partly on its wording, and since in linear regression the responses are modelled directly, results cannot be invariant to the wording of the question. However, the distribution over the population of the underlying attitude, which is the focus of analysis, should be invariant to the wording of the question. Because the ordered probit model estimates the parameters of this underlying distribution, rather than the response itself, any such "framing effects" are likely to be avoided. As we discuss in Section 7.4, it is the cut-points, rather than the distribution itself, which are expected to adjust when the wording of the question changes.

## 7.3 The Ordered Probit Model: Theory

This section is concerned solely with theoretical aspects of the model. Let $i$ index respondent $i$, $i = 1, \cdots, n$, where $n$ is the sample size. Let $y_i$ be individual $i$'s ordinal response, and assume that this can take one of the integer values $1, 2, 3, \cdots, J$. Let $y_i^*$ $(-\infty < y_i^* < +\infty)$ be the underlying latent variable representing respondent $i$'s propensity to respond positively. Let $x_i$ be a vector of characteristics relevant in explaining the attitude of a respondent. The ordered probit model is based on the assumption that $y_i^*$ depends linearly on $x_i$, according to:

$$\begin{aligned} y_i^* &= x_i'\beta + u_i \qquad i = 1, \cdots, n \\ u_i &\sim N(0, 1) \end{aligned} \tag{7.1}$$

$\beta$ is a vector of parameters, not containing an intercept. These parameters will ultimately be interpretable in the same way as slope parameters in linear regression.

$y^*$ is unobserved, but the relationship between $y^*$ and the observed variable $y$ is:

$$\begin{aligned} y &= 1 \quad \text{if } -\infty < y^* < \kappa_1 \\ y &= 2 \quad \text{if } \kappa_1 < y^* < \kappa_2 \\ y &= 3 \quad \text{if } \kappa_2 < y^* < \kappa_3 \\ &\vdots \\ y &= J \quad \text{if } \kappa_{J-1} < y^* < \infty \end{aligned} \tag{7.2}$$

The parameters $\kappa_j$, $j = 1, \cdots, J-1$, are known as "cut-points".

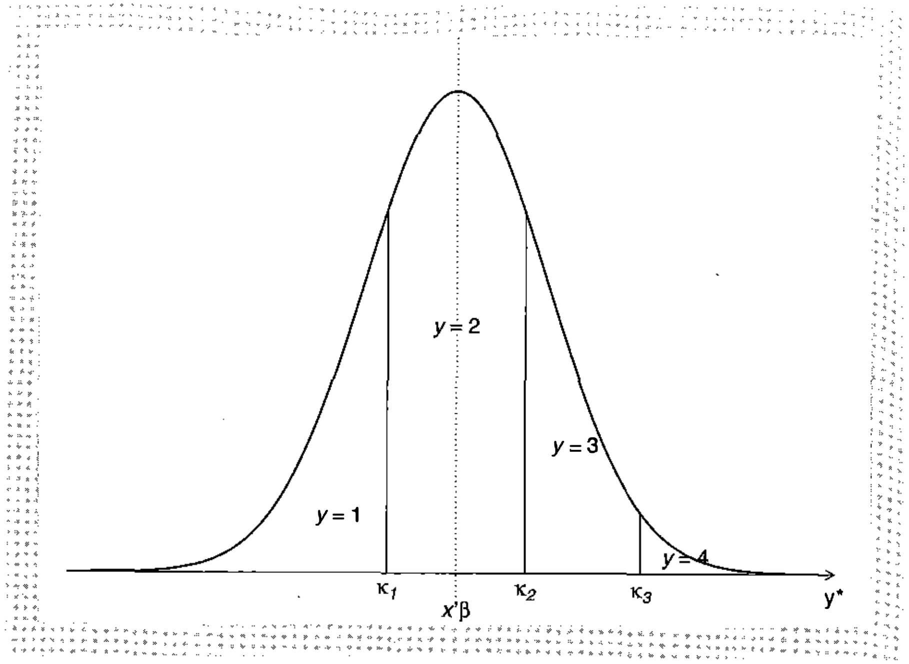



Figure 7.1: The density function of $y^*$, and its relationship to $y$

Figure 7.1 shows the density function of $y^*$. A set of cut-points for the case $J = 4$ is superimposed. Note that the mean $(x'\beta)$ of $y^*$ depends on the explanatory variables contained in the vector $x$, and therefore the whole distribution shifts when the value of one such variable changes, in a direction dictated by the sign of the corresponding $\beta$-coefficient. It is obvious from the diagram that such a shift causes a change in the distribution of responses, since the cut-points are fixed.

The absence of an intercept in the model specified here is a consequence of the $J-1$ cut-points all being free parameters. The need to normalise either the intercept or one of the cut-points, and in addition to set $Var(u_i) = 1$ in (7.1), is in order to set the otherwise arbitrary scale of the latent variable $y^*$.

The log-likelihood function will now be constructed. Let $P_i(y)$ be the probability that the $i$-th respondent's response is $y$. This probability is:

$$P_i(y) = P\left(\kappa_{y-1} < y_i^* < \kappa_y\right) = \Phi\left(\kappa_y - x_i'\beta\right) - \Phi\left(\kappa_{y-1} - x_i'\beta\right) \qquad (7.3)$$

where $\Phi(.)$ is the standard normal cumulative distribution function. So, based on a sample $(y_i, x_i, i = 1, \cdots, n)$, the log-likelihood function is:

$$LogL = \sum_{i=1}^{n} \ln\left[P_i(y_i)\right] = \sum_{i=1}^{n} \ln\left[\Phi\left(\kappa_{y_i} - x_i'\beta\right) - \Phi\left(\kappa_{y_i-1} - x_i'\beta\right)\right] \qquad (7.4)$$



The log-likelihood (7.4) is maximised with respect to the elements of $\beta$ along with the cut-points $\kappa_1, \kappa_2, \cdots, \kappa_{J-1}$, to give maximum likelihood estimates (MLEs) of both sets of parameters.

As a final point in this theoretical section of the chapter, note that when there are only two possible outcomes on the Likert scale, for example satisfied/unsatisfied, ordered probit simplifies to the more familiar binary probit model (covered in Chapter 6), with the only difference that the single cut-point, although equal in magnitude to the intercept arising from binary probit, is opposite in sign (see Exercise 1).

## 7.4 Interpreting the Cut-point Parameters

Studies making use of the ordered probit model typically focus on the interpretation of the elements of the vector $\beta$, since these parameters represent the impact of each variable on attitude. The cut-points $\kappa_1, \kappa_2, ..., \kappa_{J-1}$ are usually treated as nuisance parameters, and their values are rarely given interpretation. In this section, it is suggested that the cut-point estimates may be informative in important ways.

Firstly, consider a situation in which subjects are asked to score on a Likert scale their affinity for a particular item or product. If the item or product is one for which people have a tendency either to love or to hate,[1] then we might expect the cut-points to be tightly bunched in the middle of the distribution. If, in contrast, the item or product is one to which the majority of people are indifferent, we would expect the cut-points to be more widely dispersed. Thus the dispersion of the cut-points may be perceived as a measure of uniformity of opinion.

Secondly, the cut-points are a means of incorporating "focal points" in responses. A regular feature of Likert scale data is an apparent over-incidence of observations at the middle value of the scale.[2] When the ordered probit model is estimated in this situation, the cut-points that demarcate the middle value will be relatively wide apart. Hence the structure of the model is able to incorporate this rather awkward feature of the data.

Finally, it must be the case that the cut-points adjust according to the wording of the survey question. For example, if the wording of a question is obscure and hard to understand, we might expect the middle cut-points to be far apart, reflecting the fact that respondents who fail to understand the statement tend to report indifference; in Section 7.5, we shall see that this is particularly relevant when individuals are asked to reveal intensities of certain emotions.

[1] The standard example among British people is *marmite*, a sticky, dark brown paste with a distinctive, powerful flavour. In fact, the word "marmite" has become a metaphor for anything that tends to polarise opinions.

[2] A rather striking example is where individuals are asked to report their willingness to take risk on a Likert scale; the mode at the midpoint is particularly pronounced in this situation. See Dohmen et al. (2011).

## 7.5 Application to Data on Emotions

It is widely agreed that emotions play an important role in decision making. However, few experiments have focused on the role of emotions, perhaps because of the difficulties of measuring them. One exception is Bosman & van Winden (2002) who succeed in measuring subjects' emotions in the course of a "power-to-take" game. Here we use data from their experiment, which is available in the file **emotions**.

As a preliminary stage to the game, subjects earn an income ($Y$) by performing a real effort task. Subjects are divided between "takers" and "responders", and each taker is matched to a responder. Assume that, for a given pair, the taker has earned an income $Y_{take}$, and the responder has earned $Y_{resp}$. In the first main stage of the game, the taker decides how much of the responder's income to take for herself; the proportion $t \in [0, 1]$ of $Y_{resp}$ that she takes is known as the "take-rate". In the second stage, the responder is free to "destroy" a part of his own income (*before* the transfer to the taker); $d \in [0, 1]$ is the proportion of $Y_{resp}$ that he destroys.

For the taker, the total earnings in the experiment are $Y_{take} + t(1-d)Y_{resp}$. For the responder, total earnings are $(1-t)(1-d)Y_{resp}$.

The essence of this game is that the taker is free to take as much of the responder's income as she desires, while the responder can, if he so desires, retaliate by reducing the amount which will be taken, at the cost of reducing his own earnings. Real world analogies of this experimental set-up are: a government (taker) setting a tax rate, and workers (responders) protesting to too high a tax rate by supplying less labour; a monopolist (taker) setting a price, and the buyer (responder) deliberately deviating from their own optimal quantity if they consider the price to be outrageously high; a principal (taker) enforcing an incentive scheme for an agent (responder), with the agent retaliating by deviating from the optimal effort level dictated by the scheme, if they are not happy with it. Note that each of these cases is a well-established economic setting that is now being extended to allow for an efficiency cost which may be labelled "emotional hazard", whereby the emotions of the responder cause him or her to deviate from pay-off maximising behaviour.

Bosman & van Winden (2002) are primarily interested in the ways in which emotions are stirred in the responder, and in how these emotions influence responder behaviour. The power-to-take game is (at least in theory) particularly well-suited to the latter, since the responders' punishment decision is a continuous variable, allowing identification of the trade-off between the desire to punish and the desire for monetary gain. This would not be possible in the context of the ultimatum game (see Section 2.5) in which the responder's decision is simply whether or not to punish.

Emotions were elicited immediately following the completion of the two rounds of the game. Eleven emotions are considered: irritation, anger, contempt, envy, jealousy, sadness, joy, happiness, shame, fear, and surprise. Each is measured on a seven-point Likert scale, with one representing "no emotion" and seven representing "high intensity of emotion".



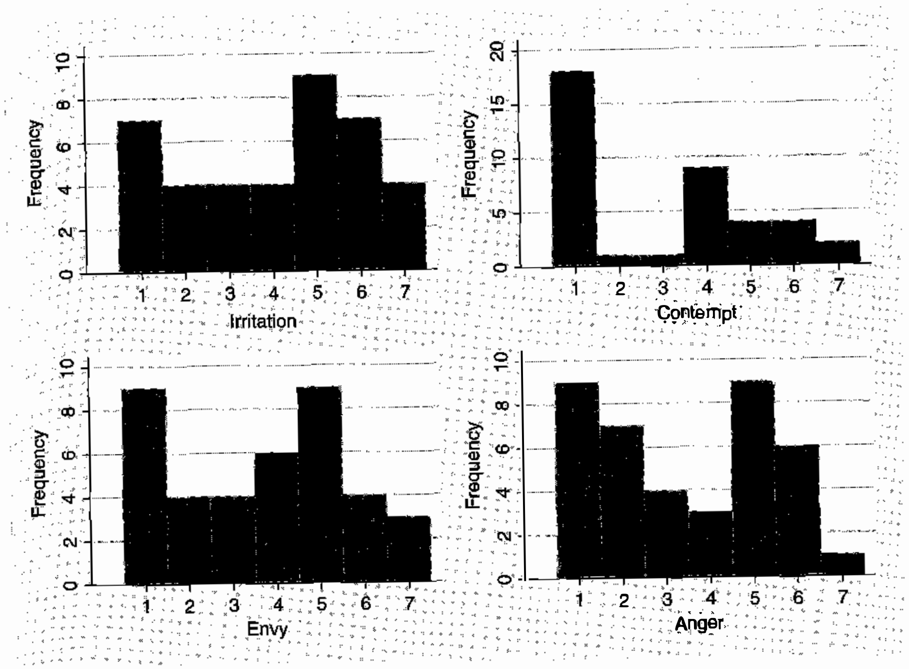



Figure 7.2: Histograms over the 39 responders of: irritation; contempt; envy; anger

Note that some of these emotions are positive emotions. Their presence is for the purpose of ensuring that subjects are not pushed in a particular direction. We shall focus here on the negative emotions.

In the data set **emotions**, there are 39 rows; each row represents a taker-responder pair. The emotion variables apply to the responder. In Figure 7.2, we present histograms of four of the emotion variables for the 39 responders. Note that the distributions have a tendency to bi-modality. Note also that, contrary to the discussion in the last section, there does not appear to be a concentration at the mid-point (4 in this case). It seems that subjects have a good idea of the state of their own emotions. One possible exception is "contempt", and the peak at the mid-point of this distribution may conceivably reflect the fact that some subjects do not fully understand the meaning of the word "contempt".

These ordinal variables may each be used as the dependent variable in an ordered probit model in order to investigate how emotions are determined within the experiment. The key explanatory variable is the take-rate. As Figure 7.3 shows, the spread of the take-rate over the sample of takers is quite wide, and this is useful for identifying its effect on a given emotion. Since we would like to consider the possibility that gender has an influence on experienced emotion, we also include in each model a dummy variable female_resp which represents the gender of the responder (1=female; 0=male). Of the 39 responders, 11 are female and 28 are male.

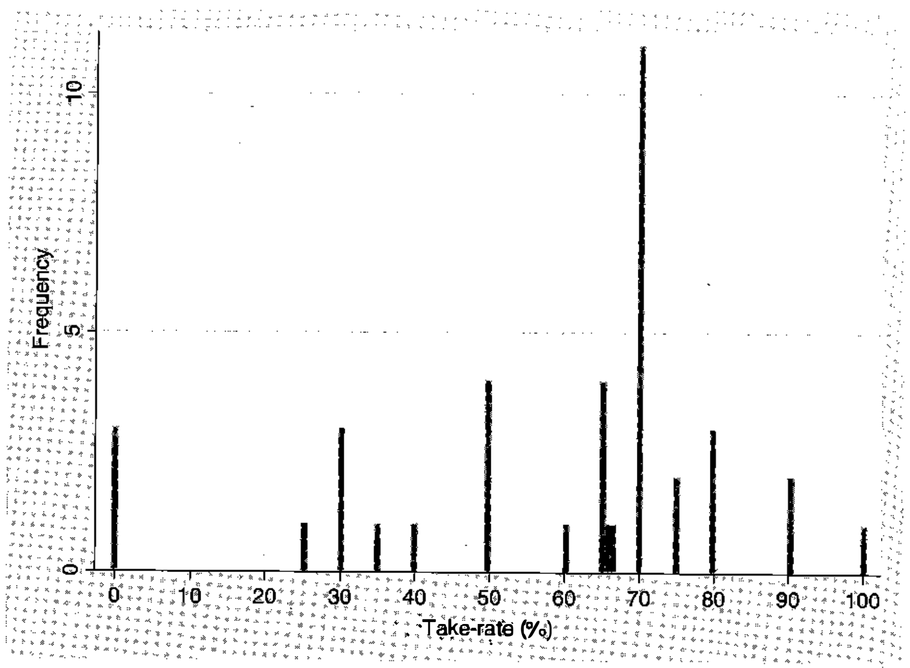



Figure 7.3: Histogram over the 39 takers of the take-rate (measured as a percentage)

Results from the four models are presented below:

```
oprobit irritation take_rate female_resp

Iteration 0:   log likelihood = -73.680431
Iteration 1:   log likelihood = -63.462026
Iteration 2:   log likelihood = -63.434641
Iteration 3:   log likelihood = -63.434636

Ordered probit regression                         Number of obs   =         39
                                                  LR chi2(2)      =      20.49
                                                  Prob > chi2     =     0.0000
Log likelihood = -63.434636                       Pseudo R2       =     0.1391

------------------------------------------------------------------------------
  irritation |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   take_rate |   .0343512    .008703     3.95   0.000     .0172935    .0514088
 female_resp |   -.475202   .4163888    -1.14   0.254    -1.291309    .3409051
-------------+----------------------------------------------------------------
       /cut1 |   .7125538   .5689905                     -.4026471    1.827755
       /cut2 |   1.199427   .5602036                      .1014481    2.297406
       /cut3 |   1.565227   .5681472                      .4516789    2.678775
       /cut4 |    1.89492   .5794498                      .7592191    3.030621
       /cut5 |   2.659484   .6192515                      1.445773    3.873194
       /cut6 |    3.54009   .7033593                      2.161531    4.918649
------------------------------------------------------------------------------

. oprobit contempt take_rate female_resp

Iteration 0:   log likelihood = -58.600542
Iteration 1:   log likelihood =  -56.64032
Iteration 2:   log likelihood = -56.633314
Iteration 3:   log likelihood = -56.633314
```



```
Ordered probit regression                         Number of obs   =         39
                                                  LR chi2(2)      =       3.93
                                                  Prob > chi2     =     0.1398
Log likelihood = -56.633314                       Pseudo R2       =     0.0336

------------------------------------------------------------------------------
    contempt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   take_rate |   .0162229   .0083824     1.94   0.053    -.0002063    .0326521
 female_resp |   .2449483   .4132229     0.59   0.553    -.5649536     1.05485
-------------+----------------------------------------------------------------
       /cut1 |   .9302437   .5825743                      -.211581    2.072068
       /cut2 |   1.000123   .5846536                     -.1457775    2.146023
       /cut3 |   1.069731   .5867442                     -.0802664    2.219729
       /cut4 |   1.721338   .6041431                      .5372394    2.905437
       /cut5 |   2.089303   .6157647                      .8824268     3.29618
       /cut6 |   2.737878   .6807572                      1.403618    4.072137
------------------------------------------------------------------------------

. oprobit envy take_rate female_resp

Iteration 0:   log likelihood = -72.646936
Iteration 1:   log likelihood = -69.420723
Iteration 2:   log likelihood = -69.414188
Iteration 3:   log likelihood = -69.414188

Ordered probit regression                         Number of obs   =         39
                                                  LR chi2(2)      =       6.47
                                                  Prob > chi2     =     0.0394
Log likelihood = -69.414188                       Pseudo R2       =     0.0445

------------------------------------------------------------------------------
        envy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   take_rate |   .0127636   .0079182     1.61   0.107    -.0027557     .028283
 female_resp |  -.5942395   .3989633    -1.49   0.136    -1.376193    .1877143
-------------+----------------------------------------------------------------
       /cut1 |  -.2143735   .5473093                      -1.28708     .858333
       /cut2 |    .124971   .5413873                     -.9361285    1.186071
       /cut3 |   .4149783   .5391188                     -.6416752    1.471632
       /cut4 |    .843965   .5452858                     -.2247755    1.912705
       /cut5 |   1.584586   .5782487                      .4512399    2.717933
       /cut6 |   2.117638   .6170117                      .9083169    3.326958
------------------------------------------------------------------------------

. oprobit anger take_rate female_resp

Iteration 0:   log likelihood =  -70.11592
Iteration 1:   log likelihood = -66.647307
Iteration 2:   log likelihood = -66.640516
Iteration 3:   log likelihood = -66.640516

Ordered probit regression                         Number of obs   =         39
                                                  LR chi2(2)      =       6.95
                                                  Prob > chi2     =     0.0309
Log likelihood = -66.640516                       Pseudo R2       =     0.0496

------------------------------------------------------------------------------
       anger |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   take_rate |   .0038118   .0077701     0.49   0.624    -.0114172    .0190409
 female_resp |  -.9790953   .4076395    -2.40   0.016    -1.778054   -.1801366
-------------+----------------------------------------------------------------
       /cut1 |  -.8344634   .5418195                      -1.89641    .2274833
       /cut2 |    -.26548   .5443219                     -1.332331    .8013712
       /cut3 |  -.0013279   .5510509                     -1.081368    1.078712
       /cut4 |   .2041105   .5558082                     -.8852535    1.293474
       /cut5 |   .9630893   .5622515                     -.1389033    2.065082
       /cut6 |   2.037269   .6084707                      .8446886     3.22985
------------------------------------------------------------------------------
```

Since these are all negative emotions, our prior belief is that a higher take-rate will bring about higher intensities of these emotions. We therefore expect the coefficient of take-rate to be positive, and we may conduct one-tailed (>) tests in order to assess the strength of each effect. Recall that this means that we are able to divide each p-value by 2. When we do this, we see that the take-rate has a strongly significant positive effect on irritation (p=0.000), a significant effect on contempt (p=0.026), a mildly significant effect on envy (0.050), and no effect on anger (0.312).

The effect of gender is, in most cases, insignificant. This may be simply a consequence of the low sample size. An exception is anger: it seems that female subjects experience significantly less anger (p=0.016) than males.

We are also interested in how the emotions impact on the responder's decision – that is, on the decision of how much of their own income do destroy. In the experiment, although responders are free to choose to destroy any fraction of their income, nearly all responders who choose to destroy, actually destroy all of their income. To be precise, of the 39 responders, eight chose to destroy, of whom seven destroyed everything. Given this, the responder decision is best modelled using a binary data model applied to the binary variable destroyyesno. The results from binary probit are presented below.

```
. probit destroyyesno irritation contempt envy anger female_resp

Iteration 0:   log likelihood = -19.789769
Iteration 1:   log likelihood = -10.215857
Iteration 2:   log likelihood = -8.3959292
Iteration 3:   log likelihood = -8.0645951
Iteration 4:   log likelihood = -8.0467872
Iteration 5:   log likelihood = -8.0467517
Iteration 6:   log likelihood = -8.0467517

Probit regression                               Number of obs   =         39
                                                LR chi2(5)      =      23.49
                                                Prob > chi2     =     0.0003
Log likelihood = -8.0467517                     Pseudo R2       =     0.5934

------------------------------------------------------------------------------
destroyyesno |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  irritation |   1.112865   .5828563     1.91   0.056    -.0295122    2.255243
    contempt |   .7724358   .3618324     2.13   0.033     .0632572    1.481614
        envy |  -.0277646   .1785066    -0.16   0.876    -.3776311    .3221019
       anger |  -.5782199   .2996307    -1.93   0.054    -1.165485    .0090455
 female_resp |  -.6617453    1.11229    -0.59   0.552    -2.841794    1.518303
       _cons |  -7.233187    4.00229    -1.81   0.071    -15.07753    .6111574
------------------------------------------------------------------------------
Note: 7 failures and 0 successes completely determined.
```

The list of explanatory variables consists of all of the negative emotions considered above, as well as gender. We see that, as expected, irritation and contempt both have a significant positive effect on the destroy probability. Envy does not seem to be important. Anger appears to have a negative effect. At first sight this is surprising: why would an angrier respondent be less likely to destroy? Further thought leads to a possible explanation for the negative coefficient. The emotions are elicited from the responder *after* his or her decision has been made, and it is conceivable that the anger is assuaged by the act of destroying. It is then interesting to note that the other emotions of irritation and contempt are not similarly assuaged. These tentative conclusions present interesting questions for further research.



Finally, there appears to be no evidence in this data that gender has an effect on the destroy probability.

It should be mentioned that Bosman & van Winden (2002) go further with their analysis by considering the impact on the responder of the take-rate relative to the responder's expectations (which also appear in the data set as the variable "expectation"). The idea here is that negative emotions are expected to arise when the take-rate is higher than expected. This avenue has not been followed here.

As a final point, we may consider the question of whether emotions are being correctly measured. After all, there is an absence of incentive compatibility: there is no logical reason for a self-interested subject to reveal his or her emotions truthfully. However, the econometric results presented above give clues to whether responses are nevertheless truthful. The two negative emotions which are stirred by high take-rates (irritation and contempt), are the same two emotions that bring about an increase in the destroy rate. This agreement is, to an extent, consistent with the validity of the emotion data.

## 7.6 Application to Data on Strength of Preference

In Chapter 6, an example relating to the house money effect was used to demonstrate the use of binary data models. The binary decision was a choice between a safe ($y = 1$) and a risky ($y = 0$) lottery, and the single explanatory variable was initial endowment ($w$). The principal hypothesis under test was the "house money effect" – that the level of the initial endowment has a negative effect on the propensity to choose the safe lottery. Evidence of the house money effect was found using the binary probit model. The data was contained in the file **house_money_sim**. Recall that this data set is simulated.

Let us now further assume that having made the choice, subjects were asked how sure they were about their choice, and were required to select one of: not sure (1); fairly sure (2); completely sure (3). The information is contained in the variable $s$ in our simulated data. Combining the self-reported level of sureness with the choice, we obtain a Likert scale with six ordered outcomes. Let this ordinal variable be $ys$, values of which are defined in the table below.

| $ys = 1$ | $ys = 2$ | $ys = 3$ | $ys = 4$ | $ys = 5$ | $ys = 6$ |
|---|---|---|---|---|---|
| Risky | Risky | Risky | Safe | Safe | Safe |
| Completely sure | Fairly sure | Not sure | Not sure | Fairly sure | Completely sure |

In order to generate $ys$ in STATA, one possible sequence of commands would be:

```
gen ys=.
replace ys=3+s if y==1
replace ys=4-s if y==0
```

If we apply the ordered probit model to the ordinal variable $ys$, we obtain a different estimate of the house money effect from the one obtained using the probit model. It is likely to be a superior estimate, since it is using more information: data distributed over six ordered outcomes clearly convey more information than over two outcomes. We therefore expect the estimate obtained using ordered probit to be more precise, reflected in a lower standard error. The results obtained are:

```
. oprobit ys w
Iteration 0:   log likelihood = -1747.2032
Iteration 1:   log likelihood = -1679.1718
Iteration 2:   log likelihood = -1679.1516
Iteration 3:   log likelihood = -1679.1516

Ordered probit regression                         Number of obs   =       1050
                                                  LR chi2(1)      =     136.10
                                                  Prob > chi2     =     0.0000
Log likelihood = -1679.1516                       Pseudo R2       =     0.0389

------------------------------------------------------------------------------
          ys |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1306568   .0112546   -11.61   0.000    -.1527153   -.1085982
-------------+----------------------------------------------------------------
       /cut1 |  -2.214993   .0901003                     -2.391586   -2.038399
       /cut2 |  -1.719572   .0794996                     -1.875389   -1.563756
       /cut3 |  -1.236927   .0731405                      -1.38028   -1.093575
       /cut4 |   -.690914   .0688578                     -.8258728   -.5559552
       /cut5 |  -.1926409   .0674608                     -.3248616   -.0604202
------------------------------------------------------------------------------
```

And this is to be compared with the following results obtained from binary probit, as reported in Section 6.2.2:

```
. probit y w

Iteration 0:   log likelihood =  -634.4833
Iteration 1:   log likelihood = -584.91375
Iteration 2:   log likelihood =  -584.5851
Iteration 3:   log likelihood = -584.58503
Iteration 4:   log likelihood = -584.58503

Probit regression                                 Number of obs   =       1050
                                                  LR chi2(1)      =      99.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -584.58503                       Pseudo R2       =     0.0786

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |  -.1409882   .0145377    -9.70   0.000    -.1694816   -.1124948
       _cons |   1.301654   .0911155    14.29   0.000     1.123071    1.480237
------------------------------------------------------------------------------
```

As expected, the point estimate of the house money effect from oprobit ($-0.131$) is not too far from the one from binary probit ($-0.141$). However, also as expected, the former is more precise than the latter: the standard error from ordered probit is 0.011 compared with 0.015 for binary probit. Hence the efficiency gain from using oprobit is around 25%.



However, there is one reason to be wary when using the ordinal data. According to many economists,[3] task-related incentives are the bedrock of the theories under test, and results from experiments to test those theories cannot be taken seriously in the absence of such task-related incentives. In the elicitation of strength of preference, it is not possible to offer an incentive in such a way as to induce truthful revelation. In short, asking for strength of preference is not incentive compatible. Of course, the straightforward binary choice between the two lotteries is, in contrast, incentive compatible, assuming that the prize will be paid out.

These considerations lead to the question of whether there is a way of testing consistency between the ordinal data and the binary data, in order to judge whether the ordinal data is reliable. One test which may be used for this purpose is the Hausman specification test (Hausman, 1978).

The basis of the test is as follows. Since the binary choice is incentive compatible, the binary data may be assumed to be reliable, and the estimate obtained using binary data may be assumed to be consistent. Since the task of self-reporting the strength of preference is not incentive compatible, the resulting ordinal data may or may not be reliable, and therefore the estimate obtained using the ordinal data may or may not be consistent. We therefore construct a test on the basis of the difference between the two estimates. Roughly speaking, if this difference is small, we may conclude that both estimates are correct, and that the ordinal data is reliable; if the difference is large, we may conclude that the estimate obtained from the ordinal data is biased as a consequence of this data being unreliable.

The null and alternative hypotheses are:

$H_0$: subjects respond truthfully when asked for their strength of preference.
$H_1$: subjects do not respond truthfully when asked for their strength of preference.

The Hausman test statistic will now be derived for this problem. This requires a small amount of theory. The latent model underlying both probit and ordered probit is:

$$y_i^* = x_i'\beta + u_i \quad i = 1, \cdots, n$$
$$u_i \sim N(0, 1) \tag{7.5}$$

We are primarily interested in the estimation of the parameter vector $\beta$ (which in this case is in fact a scalar, since there is only one explanatory variable, $w$). Let $\hat{\beta}$ be the estimate obtained using binary probit, and let $\tilde{\beta}$ be that obtained using ordered probit. $\hat{\beta}$ is a consistent estimator under both $H_0$ and $H_1$, but inefficient under $H_0$. $\tilde{\beta}$ is consistent and efficient under $H_0$, but inconsistent under $H_1$.

The test is therefore based on the difference between the two estimators:

$$\hat{\beta} - \tilde{\beta} \tag{7.6}$$

3 See, for example, Grether & Plott (1979). Consider also the well-known "precepts" of Smith (1982).

If this difference is large, there is evidence against $H_0$. To form a test statistic, we need the variance matrix of this difference. This is:

$$V(\hat{\beta} - \tilde{\beta}) = V(\hat{\beta}) + V(\tilde{\beta}) - 2cov(\hat{\beta}, \tilde{\beta}) \tag{7.7}$$

The key result in the construction of Hausman's test is that the covariance of an efficient estimator with its difference from an inefficient estimator is zero. This implies that:

$$cov\left[(\hat{\beta} - \tilde{\beta}), \tilde{\beta}\right] = cov(\hat{\beta}, \tilde{\beta}) - V(\tilde{\beta}) = 0 \tag{7.8}$$

or that:

$$cov(\hat{\beta}, \tilde{\beta}) = V(\tilde{\beta}) \tag{7.9}$$

Inserting (7.9) into (7.7) gives the variance matrix for the test:

$$V(\hat{\beta} - \tilde{\beta}) = V(\hat{\beta}) - V(\tilde{\beta}) \tag{7.10}$$

whence the Hausman test statistic may be written as:

$$H = (\hat{\beta} - \tilde{\beta})' \left[V(\hat{\beta}) - V(\tilde{\beta})\right]^{-1} (\hat{\beta} - \tilde{\beta}) \tag{7.11}$$

$H$ defined in (7.11) is asymptotically distributed as $\chi^2(K)$ under $H_0$, where $K$ is the dimensionality of $\beta$.

If $H > \chi^2_{K,0.05}$, we would reject $H_0$ in favour of $H_1$, and conclude that subjects do not respond truthfully when asked for their level of confidence in the choice they have made.

There is actually a `hausman` command in STATA.[4] However, this command cannot be used for the present purpose because the two models (probit and ordered probit) contain different numbers of parameters. We therefore need to program the test ourselves. The required code is as follows.

```
probit y w

mat b_p=e(b)
mat V_p=e(V)

oprobit ys w

mat b_op=e(b)
mat V_op=e(V)

scalar hausman=(b_p[1,1]-b_op[1,1])^2/(V_p[1,1]-V_op[1,1])

scalar list hausman
```

For this data set, the Hausman test returns a statistic of 1.26. Since there is only one element in $\beta$, the null distribution is $\chi^2(1)$. Since 1.26 falls well below the critical value, $\chi^2_{1,0.05} = 3.84$, we do not reject the null hypothesis, and we may conclude that the strength of preference data is consistent with the binary choice data, and therefore that the former has been revealed truthfully.

[4] See Section 5.5 for a demonstration of the `hausman` command.



In Section 9.4, we introduce the Monte Carlo method, and we use the Hausman test just described for the purposes of illustration. There, we simulate a large number of data sets (under both the null and the alternative) like the one described above, and for each we compute the Hausman test statistic. We then use a simple graphical technique to verify that the statistic does indeed follow the $\chi^2(1)$ distribution under the null hypothesis. Using the set of results obtained under the alternative hypothesis, we may obtain an accurate estimate of the power of the test. The power turns out to be impressive: with a sample size of 1,050, under an alternative hypothesis in which 30% of the population choose randomly when asked for their strength of preference, the probability of rejecting the null hypothesis (i.e. the power of the test) is 0.714.

## 7.7 Summary and Further Reading

It was explained early in the chapter why the use of linear regression techniques in the modelling of ordinal data is inappropriate. The ordered probit model deals consistently with the various features of such data, and is therefore invaluable as a statistical device, especially given the growing popularity of the use of ordinal data in experimental economics.

It has been demonstrated that the ordered probit model is easy to estimate using STATA. Interpretation of the results is also straightforward, since the key parameters are interpreted in a similar way to coefficients in linear regression analysis. The cut-point parameters are harder to interpret, but we have seen that they have an important role in explaining certain features of the data such as the focality of mid-point responses.

In both of the settings considered as examples, there are problems associated with incentive compatibility: because of the absence of a task-related incentive, there can be no guarantee that the ordinal responses are truthful. However, we have proposed ways of addressing this issue. This usually involves checking the consistency of the ordinal response against some other measure that has been elicited in an incentive-compatible way. The Hausman test outlined in Section 7.6 is a particularly powerful means of making this comparison.

Turning to recommendations for further reading, use of the ordered probit model has been evident in the biometrics and econometrics literatures for a number of years, the earliest contribution being traced to Aitchison & Silvey (1957). Textbook accounts have been provided by Maddala (1983) and Greene (2008). Similar models appear in the psychometrics literature (for example Masters, 1982), although the ordered probit model itself is rarely used there. A practical summary of the ordered probit model is provided by Daykin & Moffatt (2002).

Two applications of the ordered probit model were covered in this chapter. The first was on emotions, and readers wishing for further detail are referred to Bosman & van Winden (2002) whose data was used in the application. The second application was to the analysis of strength of preference. Data on strength of preference has previously been collected by Connolly & Butler (2006), although we are unaware of previous studies that apply ordered probit to such data.

The Hausman test was recommended as a means of testing the truthfulness of ordinal responses, in a situation in which an incentive-compatible binary response is also available. Readers interested in the Hausman test should consult Hausman (1978) or an econometric theory text such as Greene (2008).

## Exercises

1. Verify that when the ordered probit model is applied to binary data (i.e. data with only two possible outcomes) it becomes equivalent to the binary probit model.
2. The logit model for binary data was introduced in Exercise 1 of Chapter 6. A way of representing the logit model is:

$$\begin{aligned} y_i^* &= x_i'\beta + u_i \quad i = 1, \ldots, n \\ u_i &\sim logistic(mean\ zero) \end{aligned} \tag{7.12}$$

where the logistic (mean zero) distribution for u is defined by the pdf

$$f(u) = \frac{\exp(u)}{[1+\exp(u)]^2} \qquad -\infty < u < \infty \tag{7.13}$$

Consider the ordered logit model, defined in exactly the same way as the ordered probit model, except for the assumption concerning the distribution of $u$. Derive the likelihood function for the ordered logit model. Note that the ordered logit model can be estimated in STATA with the command `ologit`.
3. Use the `ologit` command in STATA to reproduce the results in Table 3 of Bosman & van Winden (2002).
4. Derive the log-likelihood function for the panel ordered probit model. You might find it useful to consult Fréchette (2001).
5. In Section 7.6, there was a variable $s$ representing strength of preference from which an ordinal variable was constructed. The Hausman test was then performed which led to the conclusion that subjects were truthfully reporting "strength of preference". This was not, in fact, surprising since the variable $s$ was generated as the actual strength of preference in the simulation. There are two different measures of strength of preference in the data: s_poor and s_bad (respectively partially and completely false representations of the truth). Carry out the Hausman test with each of these measures. Draw a conclusion regarding the power of the test.

# Chapter 8
# Dealing with Heterogeneity: Finite Mixture Models

## 8.1 Introduction

Finite mixture models, or just mixture models, are a class of model that offer a means of separating subjects into different types. Different types do not only exhibit different behaviour; *the processes giving rise to behaviour* also differ between types. These models are labelled as "finite" mixture models because a finite number of types is being assumed. An "infinite" mixture model, if such a label were used, would correspond to a random coefficient model, or random effects model, in which it is assumed that there is a continuous variation in some parameter indexing behavioural type.

Finite mixture models are very important in Experimetrics. This is because it is becoming ever more widely accepted that different subjects are motivated in different ways, and to assume that all subjects operate according to one model is to disrespect these differences. Often average behaviour is tracked and interpreted in terms of the behaviour of a typical subject. However, if there are different types of subject operating according to different decision processes, it is quite possible that average behaviour will not be a close representation of the actual behaviour of any of the subjects under study.

In Chapter 2 we developed a framework for treatment testing, in which it was recognised that each subject has his or her own "subject-specific treatment effect", and we were interested in finding the "average treatment effect" (ATE). There, we pointed out that the appropriateness of reporting treatment effects in terms of ATEs depends on the distribution of the subject-specific treatment effect. If this distribution is bell-shaped and symmetric, the ATE provides a sensible measure of the effect of the treatment. However, we also considered the plausible situation in which half of the population are responsive to a treatment (with a treatment effect of +1.0 say), while the remaining half are irresponsive to the treatment and therefore have a treatment effect of zero. In this situation, the ATE would be +0.5 but this would be a misleading measure of the effect of the treatment, since it would not be close to the actual treatment effect of any individual subject. As pointed out in Chapter 2, the best way to deal with such discreteness in the distribution of treatment effects is to apply the mixture modelling framework, with "subject types" differing in the way in which they respond to the treatment.

There is more than one possible approach to the estimation of finite mixture models. The approach adopted here is as follows. Firstly, on the basis of economic theory, the total number of types in the population is decided, and a label is assigned to each. Then a parametric model is specified for the behaviour of each type. The parameters of these various models are estimated altogether, along with the "mixing proportions" – parameters revealing the proportion of the population which is of each type. Once the model has been estimated, we can return to the data in order to determine the posterior probability of each individual subject being of each type. Note that there is no claim to be able to identify any individual subject as belonging to any particular type with certainty, although, in situations where data is informative, posterior type probabilities can be very close to one.

This chapter commences with a simple, somewhat contrived, example: a mixture of two normal distributions. Then we shall progress to a more realistic example of offers in the "acquiring a company game". Finally, we shall consider the more complex example of giving in a public goods experiment.

## 8.2 Mixture of Two Normal Distributions

We start with a simple example.

### 8.2.1 Data and model

Consider the variable $y$ contained in the file **mixture_sim**. There are 1,000 observations. A histogram of the variable $y$ is shown in Figure 8.1. The distribution appears to be a mixture of two bell-shaped (i.e. probably normal) distributions, one with a mean around 3, the other with a mean around 6.

If $y$ represents the decision made by subjects in an experiment (on this occasion, let us not concern ourselves about what decision is actually being made), we might say that subjects are of two types, and we would set about estimating the following mixture model:

$$\text{type 1:} \quad N\left(\mu_1, \sigma_1^2\right)$$

$$\text{type 2:} \quad N\left(\mu_2, \sigma_2^2\right)$$

$$\text{mixing proportions:} \quad p(\text{type1}) = p \quad p(\text{type2}) = 1 - p$$

The mixing proportions represent the proportions of the population which are of each type. Note that there are five parameters to be estimated: $\mu_1, \sigma_1, \mu_2, \sigma_2$, and $p$.

The density associated with a particular value of $y$, conditional on the subject being of type 1, is:

$$f\,(y|\text{type1}) = \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1}{\sigma_1}\right) \tag{8.1}$$



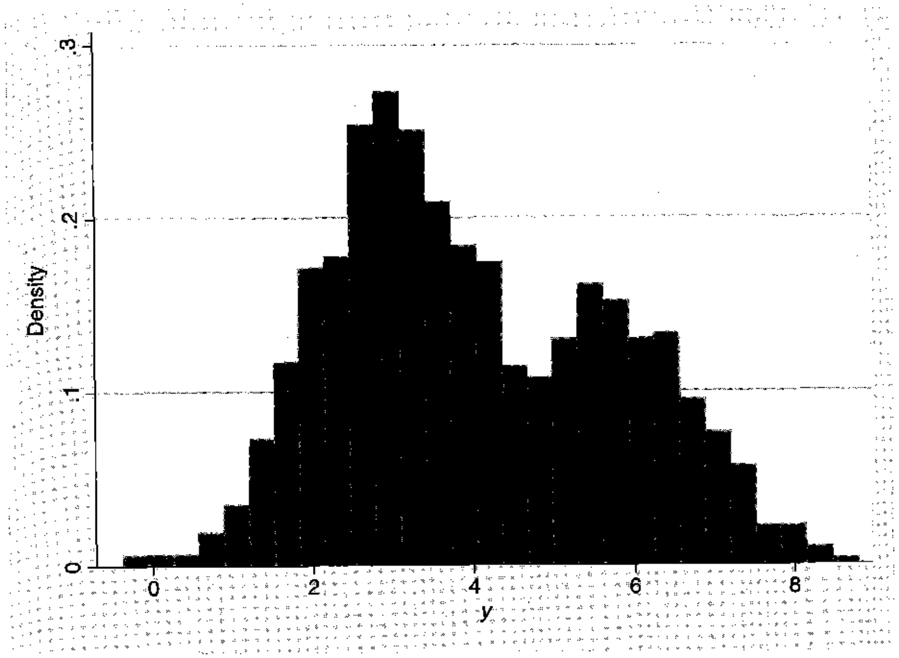



Figure 8.1: A histogram of the variable $y$ in the data set **mixture_sim**

and the density conditional on being type 2 is:

$$f\,(y|\text{type2}) = \frac{1}{\sigma_2}\phi\left(\frac{y-\mu_2}{\sigma_2}\right) \tag{8.2}$$

In (8.1) and (8.2) we are using a formula explained fully in Section 6.6.2 for the density of values drawn from a normal distribution. The *marginal* density associated with an observation is obtained by combining (8.1) and (8.2) with the mixing proportions, as follows:

$$f(y;\mu_1,\sigma_1,\mu_2,\sigma_2,p) = p \times \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1}{\sigma_1}\right) + (1-p) \times \frac{1}{\sigma_2}\phi\left(\frac{y-\mu_2}{\sigma_2}\right) \tag{8.3}$$

Equation (8.3) is used as the likelihood contribution for each observation.

The sample log-likelihood is then given by:

$$LogL = \sum_i \ln f\,(y_i;\mu_1,\sigma_1,\mu_2,\sigma_2,p) \tag{8.4}$$

Equation (8.4) is maximised with respect to the five parameters $\mu_1$, $\sigma_1$, $\mu_2$, $\sigma_2$, and $p$ in order to obtain their MLEs.

### 8.2.2 Posterior type probabilities

Having estimated a mixture model, one obvious thing to do is to compute the posterior probabilities of each subject being of each type. This involves Bayes' rule. For example, the posterior type 1 probability, given an observation $y$, is:

$$P(\text{type1}|y) = \frac{f(y|\text{type1})P(\text{type1})}{f(y|\text{type1})P(\text{type1}) + f(y|\text{type2})P(\text{type2})} = \frac{p \times \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1}{\sigma_1}\right)}{p \times \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1}{\sigma_1}\right) + (1-p) \times \frac{1}{\sigma_2}\phi\left(\frac{y-\mu_2}{\sigma_2}\right)} \tag{8.5}$$

### 8.2.3 The estimation program

A STATA program which estimates the model and then computes and plots posterior probabilities is shown below. Table 8.1 provides a correspondence between components of the likelihood function (8.4) and the names used in the program.

| Component of $LogL$ | STATA name |
|---|---|
| $\mu_1,\mu_2$ | mu1 , mu2 |
| $\sigma_1, \sigma_2$ | sig1, sig2 |
| $p$ | p |
| $f(y\|\text{type1}) = \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1}{\sigma_1}\right)$ | f1 |
| $f(y\|\text{type2}) = \frac{1}{\sigma_2}\phi\left(\frac{y-\mu_2}{\sigma_2}\right)$ | f2 |
| $\ln[f(y)] = \ln\left[p \times \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1}{\sigma_1}\right) + (1-p) \times \frac{1}{\sigma_2}\phi\left(\frac{y-\mu_2}{\sigma_2}\right)\right]$ | logl |
| $P(\text{type1}\|y)$ | postp1 |
| $P(\text{type2}\|y)$ | postp2 |

Table 8.1: Components of $LogL$ and corresponding STATA names

The annotated code is as follows. One important point is that the variable $y$ is a "global" variable, because it exists both inside and outside the likelihood-evaluation program. This is why quotation marks are not used when $y$ is used within the program.

```
* LIKELIHOOD EVALUATION PROGRAM STARTS HERE:

program define mixture
args logl mu1 sig1 mu2 sig2 p
tempvar f1 f2
```



```
* GENERATE TYPE-CONDITIONAL DENSITIES:

quietly gen double `f1'=(1/`sig1')*normalden((y-`mu1')/`sig1')
quietly gen double `f2'=(1/`sig2')*normalden((y-`mu2')/`sig2')

* COMBINE TYPE-CONDITIONAL DENSITIES WITH MIXING PROPORTIONS
* TO GENERATE MARGINAL DENSITY, AND TAKE LOG.  THIS IS THE FUNCTION THAT
* NEEDS TO BE MAXIMISED WHEN SUMMED OVER THE SAMPLE:

quietly replace `logl'=ln(`p'*`f1'+(1-`p')*`f2')

* GENERATE THE POSTERIOR TYPE PROBABILITIES, AND MAKE THEM
* AVAILABLE OUTSIDE THE PROGRAM:

quietly replace postp1=`p'*`f1'/(`p'*`f1'+(1-`p')*`f2')
quietly replace postp2=(1-`p')*`f2'/(`p'*`f1'+(1-`p')*`f2')

quietly putmata postp1, replace
quietly putmata postp2, replace

end

* END OF LIKELIHOOD EVALUATION PROGRAM

* READ DATA:

use mixture_sim, clear

* INITIALISE TWO POSTERIOR PROBABILITY VARIABLES:

gen postp1=.
gen postp2=.

* SPECIFY STARTING VALUES, AND APPLY ML:

mat start=(3, 1.5, 6, 1.5, .5)
ml model lf mixture /mu1 /sig1 /mu2 /sig2 /p
ml init start, copy
ml maximize

* EXTRACT POSTERIOR TYPE PROBABILITY, AND PLOT THEM AGAINST y:

drop postp1 postp2
getmata postp1
getmata postp2

sort y
line  postp1 postp2 y , lpattern(l -)
```

As usual, it is necessary to specify starting values for the parameters being estimated. These are stored in the vector "start". In this case, the starting values have been obtained by examination of the histogram of $y$ (see Figure 8.1). In other situations, starting values are obtained using simple estimation methods such as linear regression.

### 8.2.4 Results

The results from executing the code presented in Section 8.2.3 above are as follows:

```
                                                  Number of obs    =       1000
                                                  Wald chi2(0)     =          .
Log likelihood = -1908.2805                       Prob > chi2      =          .
```

| | | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|---|
| mu1 | | | | | | | |
| | _cons | 2.981757 | .0743116 | 40.13 | 0.000 | 2.836109 | 3.127405 |
| sig1 | | | | | | | |
| | _cons | 1.014725 | .0499721 | 20.31 | 0.000 | .9167818 | 1.112669 |
| mu2 | | | | | | | |
| | _cons | 5.950353 | .1158028 | 51.38 | 0.000 | 5.723384 | 6.177322 |
| sig2 | | | | | | | |
| | _cons | .9768525 | .0721166 | 13.55 | 0.000 | .8355064 | 1.118198 |
| p | | | | | | | |
| | _cons | .6494311 | .0296983 | 21.87 | 0.000 | .5912235 | .7076387 |

We see that the estimates of the five parameters (with standard errors) are:

$$\begin{aligned}
\hat{\mu}_1 &= 2.982(0.074)\\
\hat{\sigma}_1 &= 1.015(0.050)\\
\hat{\mu}_2 &= 5.950(0.116)\\
\hat{\sigma}_2 &= 0.977(0.072)\\
\hat{p} &= 0.649(0.030)
\end{aligned}$$

Hence we see that 64.9% of the population comes from the distribution $N(2.982, 1.015^2)$, while the remaining 35.1% comes from $N(5.950, 0.977^2)$.

However, when considering any particular observation, we cannot be certain which one of the two distributions it comes from. This is why the posterior probabilities are useful. Note that the variables containing the posterior probabilities (postp1 and postp2) are generated inside the likelihood evaluation program. In order to extract these variables, mata commands are required. The `putmata` command is used from within the program, and the `getmata` command is used outside it.

In Figure 8.2 we show a plot of the posterior probabilities against $y$, obtained using the following commands.

```
sort y
line  postp1 postp2 y , lpattern(l -)
```

This graph tells us that: observations below 3 are almost certain to be from the first distribution; observations greater than 6 are almost certain to come from the second distribution. For observations between 3 and 6, we cannot know with confidence which distribution applies. For an observation with $y = 4.70$, both distributions are equally likely to apply.

## 8.3 The fmm Command in STATA

fmm (standing for "finite mixture model") is a user-written STATA command (Deb, 2012) that directly estimates mixture models of the type considered in Section 8.2.



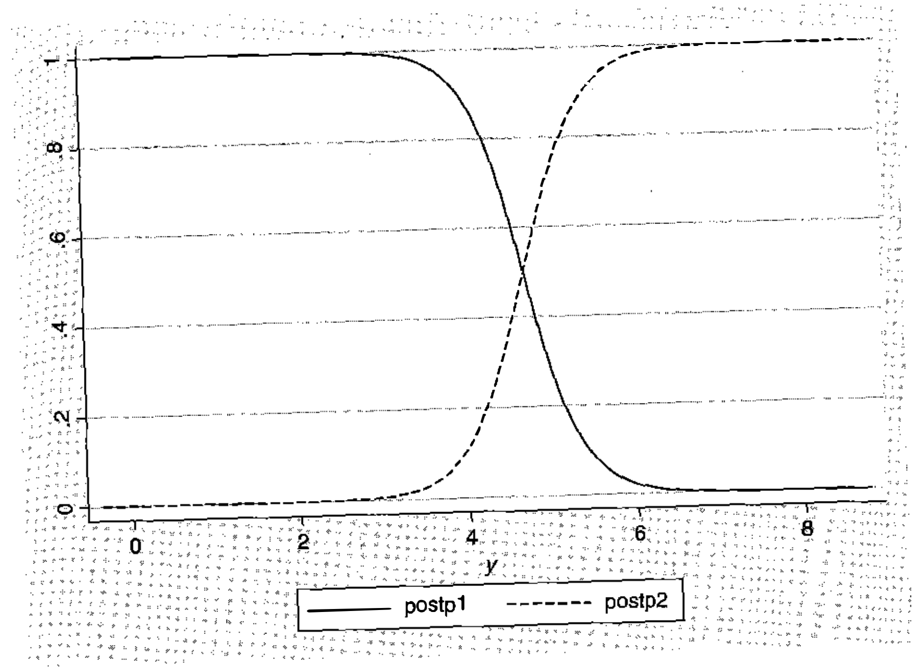



Figure 8.2: Posterior probabilities of type 1 and type 2 in the mixture data

The required syntax is:

```
fmm y, mix(normal) comp(2)
```

The main argument is the variable under analysis ($y$). The first option specifies that a mixture of normals is assumed, and the second specifies that the mixture contains two types (or components). The results from this command are as follows. Note that these results are equivalent to the results obtained using `ml` in Section 8.2, except that some parameters are named differently.

```
. fmm y, mix(normal) comp(2)

2 component Normal regression                     Number of obs   =       1000
                                                  Wald chi2(0)    =          .
                                                  Prob > chi2     =          .
Log likelihood = -1908.2805
```

| y | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| component1 | | | | | | |
| _cons | 2.981758 | .0743115 | 40.13 | 0.000 | 2.83611 | 3.127406 |
| component2 | | | | | | |
| _cons | 5.950353 | .1158024 | 51.38 | 0.000 | 5.723385 | 6.177322 |
| /imlogitpi1 | .6165402 | .130444 | 4.73 | 0.000 | .3608746 | .8722058 |
| /lnsigma1 | .0146181 | .0492469 | 0.30 | 0.767 | -.081904 | .1111401 |
| /lnsigma2 | -.0234201 | .0738254 | -0.32 | 0.751 | -.1681152 | .1212749 |

```
------------------------------------------------------------------------------
      sigma1 |   1.014725    .049972                      .9213604    1.117551
      sigma2 |    .976852   .0721165                      .8452565    1.128935
         pi1 |   .6494313   .0296982                      .5892521    .7052045
         pi2 |   .3505687   .0296982                      .2947955    .4107479
------------------------------------------------------------------------------
```

Following the `fmm` command, the post-estimation command `predict` may be used to obtain the posterior type probabilities, as follows:

```
predict post1 , pos eq(component1)
predict post2 , pos eq(component2)
```

The posterior probabilities (post1 and post2) thus obtained are identical to those obtained in Section 8.2.

## 8.4 A Mixture Model for the "Acquiring a Company" Task

The "acquiring a company" task (Bazerman & Samuelson, 1983) is described as follows. The subject is considering making an offer to buy a (hypothetical) company. The current value of the company is unknown to the subject, but known to the current owner. From the subject's point of view, this value is uniformly distributed between 0 and 100 (dollars per share). If the subject's offer is accepted, he or she "takes over" the company, and its value automatically increases by 50%. The task is simply to decide how much to offer for the company.

In this situation, a subject might well apply the following reasoning. The "expected" value of the company, to the current owner if he keeps it, is 50. The "expected value" to the acquirer (i.e. the subject) is therefore 75. Hence, if the subject offers any amount between 50 and 75, both parties will benefit.

The flaw in the reasoning just given is that it neglects the informational content of the seller's acceptance of an offer. The correct reasoning is as follows. If any offer is accepted, the acceptance is a signal that the current value is *lower* than the offer, and hence has a conditional expectation of 50% of that offer. Hence, conditional on acceptance of an offer, the acquirer's value will only be 75% of that offer. Therefore any positive offer must result in an expected loss. Of course this leads to the result that the only offer that does not result in loss is zero, that is, no offer at all.

Buyers who adopt the first style of reasoning, without considering the negative information conveyed in the acceptance of an offer, are said to be falling prey to a form of "winner's curse", a concept analysed in the context of auctions in Chapter 4.

When analysing data on offers made in this task, it seems reasonable to assume that the population of subjects is divided into those who fall prey to winner's curse, whose offers are expected to be in the range 50–75, and those who have successfully "solved" the problem, whose offers are expected to be close to zero. A mixture model, similar to the type analysed in Section 8.2, may then be used to estimate the proportion of the population who correctly solve the game.



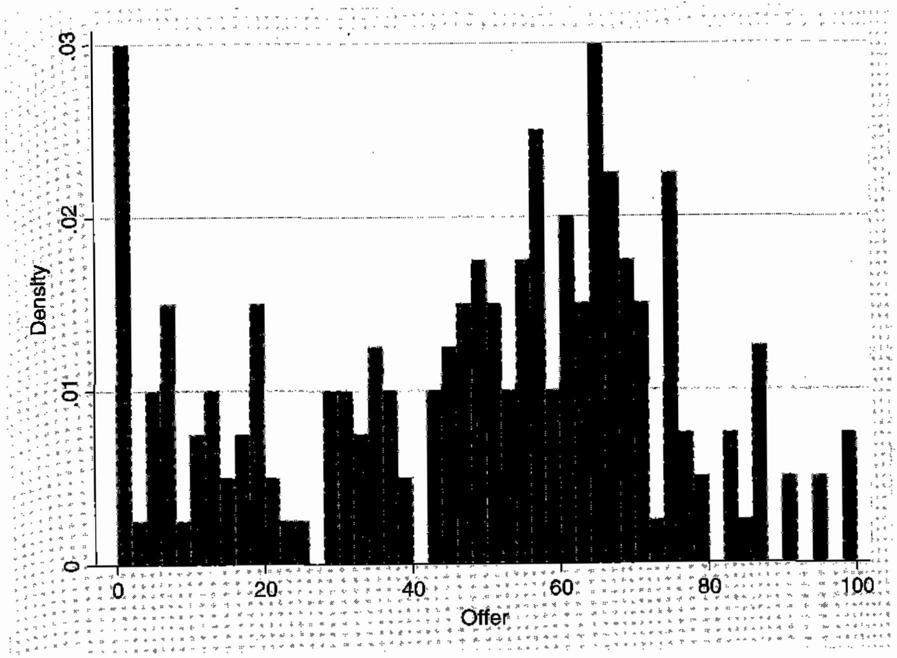



Figure 8.3: Histogram of the offers of 200 acquirers in the "acquiring a company" game

When this game is played by experimental subjects, only a small proportion offer exactly zero. But let us assume that subjects with small positive offers have understood that the best offer is zero, but give nevertheless, perhaps because they feel an obligation to do so, or perhaps in an attempt to make the game more exciting.

Let us imagine that an experiment has been carried out, with 200 subjects each of whom has made an offer. Simulated data is contained in the file **acquire_sim**. A frequency histogram of offers (y) is shown in Figure 8.3.

We see that the distribution of offers is, as expected, bi-modal, with one mode around 60, and the other at zero. Note, however, that only 11 of the 200 subjects offer exactly zero. A number of subjects make small positive offers, and as explained above, we are allowing these subjects to be of the same "type" as those at exactly zero.

When the `fmm` command (Section 8.3) is applied to this data set, assuming a straightforward mixture of two normals, the log-likelihood fails to converge. This is because the assumption of two normals does not provide an adequate fit to the data, as a consequence of the accumulation of values at zero. The obvious way of dealing with this problem is to assume that zero observations are censored zeros. Note also that there is a small accumulation of observations at the upper limit of 100, and, for good measure, we shall treat these observations as right-censored. The methods for allowing for lower and upper censoring in the construction of log-likelihood functions was explained fully in Section 6.6.3.

The STATA code required to estimate this mixture model and then obtain the posterior type probabilities is:

```
program define acquire1
args lnf mu1 sig1 mu2 sig2 p
tempvar y f1 f2

quietly gen double 'y'=$ML_y1
quietly gen double 'f1'=(1/'sig1')*normalden(('y'-'mu1')/'sig1')
quietly replace 'f1'=normal(-'mu1'/'sig1') if y==0
quietly replace 'f1'=1-normal((100-'mu1')/'sig1') if y==100
quietly gen double 'f2'=(1/'sig2')*normalden(('y'-'mu2')/'sig2')
quietly replace 'f2'=normal(-'mu2'/'sig2') if y==0
quietly replace 'f2'=1-normal((100-'mu2')/'sig2') if y==100
quietly replace 'lnf'=ln('p'*'f1'+(1-'p')*'f2')

end

mat start=(0, 20, 60, 20, .8)
ml model lf acquire1 (y= ) () () () ()
ml init start, copy

ml maximize
```

The output is as follows:

```
                                                  Number of obs   =        200
                                                  Wald chi2(0)    =          .
Log likelihood =  -885.3141                       Prob > chi2     =          .

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
       _cons |   10.02066   3.645138     2.75   0.006      2.87632      17.165
-------------+----------------------------------------------------------------
eq2          |
       _cons |   12.57707   3.266815     3.85   0.000     6.174229    18.97991
-------------+----------------------------------------------------------------
eq3          |
       _cons |   60.07094   2.209104    27.19   0.000     55.74118    64.40071
-------------+----------------------------------------------------------------
eq4          |
       _cons |   17.00645   1.659162    10.25   0.000     13.75455    20.25835
-------------+----------------------------------------------------------------
eq5          |
       _cons |     .24227   .0521544     4.65   0.000     .1400492    .3444908
------------------------------------------------------------------------------
```

The results are summarised as follows:

$$Type\ \ 1 : y^{\star} \sim N(10.0, 12.6^2)$$
$$Type\ \ 2 : y^{\star} \sim N(60.1, 17.0^2)$$
$$y = 0\ \ if\ \ y^{\star} < 0; \quad y = 100\ \ if\ \ y^{\star} > 100$$
$$P(Type\ \ 1) = 0.24$$

We thus see that the proportion of the population who correctly solve the game is estimated to be 24%, with the remaining 76% falling prey to winner's curse. The mean offer for the first type is 10, while that for the second type is 60.



## 8.5 A Mixture Model for Giving in a Public Goods Experiment

### *8.5.1 Background*

Public goods experiments were introduced in Section 2.5.4. Each subject has to divide an endowment between a public account and a private account. In Section 2.5.4 it was explained that the game has a unique Nash equilibrium consisting of zero contributions by every subject. In experiments, a substantial proportion of subjects do indeed contribute zero. However, a sizeable proportion of subjects also contribute positively, with much variation in contributions both between and within subjects, and the objective of public goods experiments is typically to investigate the motivations behind such positive contributions.

In this section of the book, we develop a model which allows for a variety of motivations. It uses data from a real experiment (Bardsley, 2000), whose design is tailored towards separate identification of these various motivations.[1] The key feature of the experimental design is that subjects take turns to contribute, and each observes the contributions made by subjects placed earlier in the sequence. Also, the task is repeated, so the resulting data are panel data.

A number of econometric issues need to be addressed. Most importantly, it is clear from previous literature that there are different types of agent in the population, each with different contribution motives. Hence, a mixture model, of the type used earlier in this chapter in other contexts, is appropriate for separating the various subject types.

The mixture model that we develop assumes three types. A "reciprocator" is a subject who contributes more when contributions by others placed earlier in the sequence are higher. We capture reciprocity by allowing the (reciprocator) subject's contribution to depend on the median of previous contributions within the sequence. A "strategist" is a subject who is selfish, but is willing to make positive contributions in anticipation of reciprocity by others placed later in the sequence. Since as the sequence progresses there are less subjects left to play, a strategist has less incentive to contribute when placed later in the sequence. For example, for a strategist in last position, there is definitely no contribution motive. We therefore define "strategists" as agents whose contribution depends negatively on their position in the sequence, with a zero contribution when in last position. Finally, a "free-rider" is a subject who displays a tendency to contribute zero regardless of the behaviour of other subjects or of their position in the sequence.

Application of the mixture framework in this setting is more complicated than in previous examples in this chapter, because of the panel structure of the data. This is a further econometric issue to be addressed.

[1] The Bardsley (2000) data set has been used previously to demonstrate the use of the Tobit model in Section 6.6.3.

Another potential influence on a subject's contribution is task number. One almost universal finding in public goods games is a downward trend in contributions as the game is repeated. Standard explanations are in terms of a learning process. Subjects are learning, either about the game's incentive structure (i.e. learning to be rational), or about others' behaviour (*social* learning). A novel feature of this experimental design is that, for reasons to be explained in the next sub-section, it removes the effect of social learning; any decay in contributions over the course of the experiment will be attributable exclusively to learning about the incentive structure.

A final econometric issue is censoring. Contributions constitute a doubly censored dependent variable, since the lowest possible contribution is zero, and the highest possible contribution is the amount of the endowment. As in Section 6.6.3, a two-limit Tobit model (Nelson, 1976) is therefore required in order to obtain consistent estimates of the effects of experimental variables.

### 8.5.2 Experiment

In the data set of Bardsley (2000), there are 98 subjects, divided into groups of size 7. Each subject performs 20 tasks.

The experimental design has a number of distinctive features. Firstly, within a single game, subjects take turns to contribute, and each subject observes the sequence of previous contributions. There are two reasons why this is important. Firstly, because it is possible for a subject to observe previous contributions of others, it is possible for us to be able to assess the extent to which their contribution is driven by the previous contributions of others. That is, it becomes possible to test for reciprocity. Secondly, given that a subject is aware of her own position in the ordering, she will clearly know how many subjects are contributing after her, and she will therefore be in a position to assess the benefits from "strategic" contribution. By investigating the effect of position in the ordering on contribution, it therefore becomes possible for us to test for strategic behaviour.

Another distinctive feature of this experiment is the use of the "conditional information lottery" (CIL)[2] as an incentive mechanism. In a CIL, the game played by the subject is camouflaged amongst a set of 19 (in this case) controlled fictional tasks. Conditional on a task's being the real one, the task information describes the real game (so "others' behaviour" is as shown). Subjects are told beforehand that only one task is a real game, that in the remainder "others' behaviour" is simply an artefact of the design, and that only the real task is to be paid out. Subjects do not know which task is the real task, and it is therefore reasonable to suppose that they treat each task as if it is the real task.

The CIL is similar to the random lottery incentive (RLI) mechanism described in Section 2.6, in the sense that only one from a set of tasks is for real and subjects

[2] See Bardsley (2000) for a full explanation of the CIL.



do not know which is the real task until the end of the experiment. Unlike the RLI, however, the experimenter knows which task is for real from the start.

The main benefit from using the CIL is that it removes the effect of social learning. This is because subjects are aware that only one game is actually being played with the other subjects in the group. Given that a subject is temporarily assuming that the game currently being played is the real game, it is logical for that subject also to assume that all previous tasks were fictional, so anything learnt from those previous tasks cannot logically be about the behaviour of other subjects in the group. For this reason any decay in contributions over the course of the experiment will be attributable exclusively to learning about the incentive structure.

### 8.5.3 The data

The data set of Bardsley (2000) is contained in the file **bardsley**. The same data set was used in Section 6.6.3 to demonstrate the use of the two-limit tobit model.

As noted previously, 98 subjects were observed over 20 tasks. It is revealing to examine the pooled distribution of contributions. A histogram of this variable is shown in Figure 8.4. The histogram clearly reveals censoring at zero, and to a lesser extent, censoring at the upper limit, 10. The overall mean contribution is 2.711, compared with a median of 1.0, this difference confirming the clear positive skew evident in the histogram.

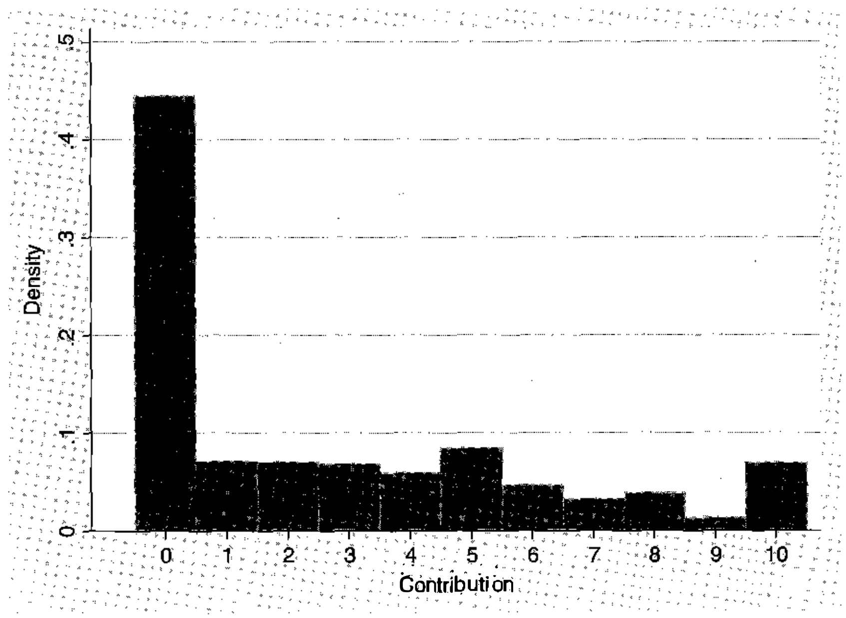



Figure 8.4: Distribution of contributions in Bardsley's experiment

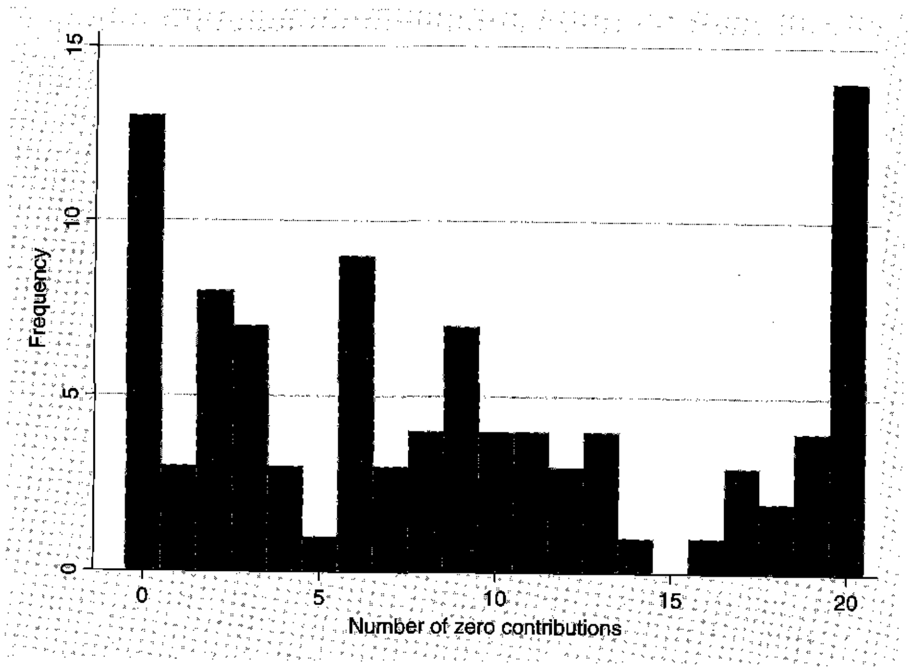



Figure 8.5: A histogram of the number of zero contributions by each subject

In order to give a feel for the extent of between-subject variation, Figure 8.5 shows the distribution of the number of zero contributions made by each subject. Apart from establishing the wide variation in behaviour, Figure 8.5 is useful in providing a rough estimate of the number of free-riders in the sample. Remembering the (strict) definition of a free-rider as a subject contributing zero on *every* occasion, we see that 14 of the 98 subjects (14.3%) satisfy this definition. However, the definition of free-riders which we incorporate in estimation is less rigid than this. We include a tremble parameter which allows a small probability of subjects setting their contribution completely at random on any task. This means that a genuine free-rider may be observed making positive contributions on a small number of occasions. Figure 8.5 is useful because the clearly discernible cluster of 24 subjects (24.5%) who contribute zero on at least 16 out of 20 occasions, may reasonably be tentatively identified as free-riders who are subject to a tremble.

It is also revealing to use graphical methods to investigate the nature of the effects of the various determinants of contribution. For this purpose, we present three scatter plots with Lowess smoothers in Figure 8.6. Since the effects of these variables cannot apply to the behaviour of free-riders, contributions of the 24 subjects identified loosely as free-riders in the context of Figure 8.5 are excluded from Figure 8.6. The scatters themselves are clearly uninformative since the vast majority of the possible combinations of MED and contribution are represented in the plots.



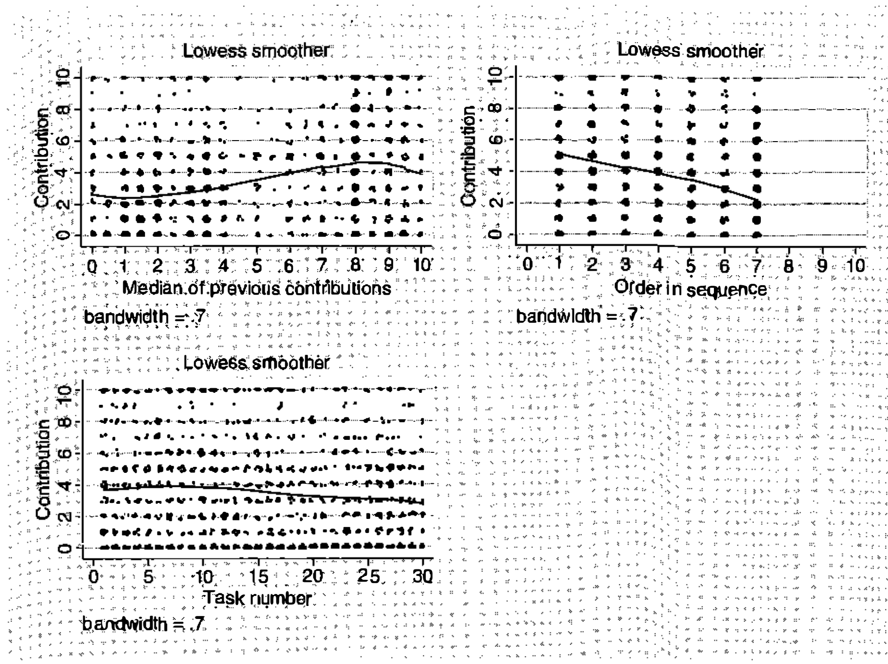



Figure 8.6: Jittered scatter plots and Lowess smoothers of contribution against (i) median of previous contributions, (ii) order in sequence, and (iii) task number. Free-riders excluded

For this reason, we include "jitter"[3] in the scatterplots to make it possible at least to see which locations contain the most points. We also include Lowess smoothers, which plot the estimated conditional mean of contribution.

The Lowess smoothers in Figure 8.6 reveal that all three variables appear to have an impact on contribution. Moreover, the direction of each effect is in accordance with theoretical predictions: the median of previous contributions has a positive effect, as predicted by reciprocity theory; order in the sequence has a negative effect, implying strategic behaviour; task number has a negative effect, implying a process of learning about the game. We also see that the three effects are monotonic and roughly linear. These observations are useful in guiding the specification of the parametric model developed in the next section.

### *8.5.4 The finite mixture two-limit Tobit model with tremble*

The econometric analysis used here is similar to that of Bardsley & Moffatt (2007).

[3] The concept of "jitter" in the context of scatterplots was explained in footnote 6 of Chapter 6.

We assume that there are $n$ subjects, each of whom has been observed over $T$ tasks. Let $y_{it}$ be the observed contribution by subject $i$ in task $t$. The variable $y_{it}$ has a lower limit of 0 and an upper limit of 10. The two-limit tobit model (Nelson, 1976), with limits 0 and 10, is therefore appropriate. To adopt conventional terminology in limited dependent variables modelling, we refer to zero contributions as being in "regime 1", contributions greater than 0 but less than 10 in "regime 2", and contributions of 10 in "regime 3".

The underlying desired contribution is $y_{it}^*$ and this will be assumed to depend linearly on a set of explanatory variables. However, as explained in Section 8.5.1, we are assuming that each subject is one of three types: reciprocator (*rec*), strategist (*str*) and free-rider (*fr*), and the determination of $y_{it}^*$ for a given subject depends crucially on which type that subject is. An important feature of the model (and finite mixture models in general) is that there is no switching between types: given that a particular subject is of a given type, the subject is of that type for every task undertaken.

For reciprocators and strategists, we specify the following latent model for the desired contribution (free-riders are treated separately since no explanatory variables are relevant to their contribution):

$$\begin{aligned}
&\text{reciprocator:} && y_{it}^* = \beta_{10} + \beta_{11}MED_{it} + \beta_{13}(TSK_{it} - 1) + \epsilon_{it,rec} \\
&\text{strategist:} && y_{it}^* = \beta_{20} + \beta_{22}(ORD_{it} - 1) + \beta_{23}(TSK_{it} - 1) + \epsilon_{it,str} \\
&&& i = 1, ..., n \quad t = 1, ..., T \\
&&& \epsilon_{it,rec} \sim N(0, \sigma_1^2) \qquad \epsilon_{it,str} \sim N(0, \sigma_2^2)
\end{aligned} \tag{8.6}$$

where $ORD_{it}$ is subject $i$'s position in the group for the $t$th task solved, $MED_{it}$ is the median of previous contributions by other subjects in the group, and $TSK_{it}$ is the task number.[4] Reciprocity implies $\beta_{11} > 0$, while strategic behaviour implies $\beta_{22} < 0$. The parameters $\beta_{13}$ and $\beta_{23}$ represent learning by reciprocators and strategists respectively, and are expected to be negative. The reason for subtracting one from $TSK$ and $ORD$ is to ensure that the intercept in each equation has a convenient interpretation: expected contribution for a subject in first position in the first task.

When $ORD = 1$, $MED$ is clearly not defined, but will be set to 8.00 for the purpose of estimation; this value is obtained by a trial and error process, being that which maximises the log-likelihood.[5] It could usefully be interpreted as subjects' *a priori* expectation of others' contributions, formed before the start of the experiment. The relatively high value embodies the idea that reciprocators start the game with an optimistic outlook regarding the generosity of other players.

The relationship between *desired* contribution $y_{it}^*$ and *actual* contribution $y_{it}$ is specified by the following censoring rules:

[4] $TSK$ is not the same as $t$, since some of the tasks are part of a separate experiment. While $t$ goes from 1 to 20, the range of $TSK$ is from 1 to 30.

[5] The same procedure was used by Bardsley & Moffatt (2007). See their Table 4 note 2.



For reciprocators and strategists:

$$y_{it} = \begin{cases} 0 & \text{if } y_{it}^* \le 0 \\ y_{it}^* & \text{if } 0 < y_{it}^* < 10 \\ 10 & \text{if } y_{it}^* \ge 10 \end{cases} \tag{8.7a}$$

For free-riders:

$$y_{it} = 0 \quad \forall t \tag{8.7b}$$

As mentioned in Section 8.5.3, we also introduce a "tremble parameter", $\omega$ (see Moffatt & Peters, 2001). On any single response, with probability $\omega$ a subject loses concentration and chooses randomly between the 11 possible contributions. One purpose of this parameter is to relax the rigid segregation rules between the three subject types. If, for example, a subject contributes zero on every occasion except one, we wish to assign a positive probability to this subject being a free-rider who lost concentration on one occasion. The presence of the tremble parameter allows this.

Loomes et al. (2002), in their econometric models of risky choice, include a tremble parameter which decays in magnitude in the course of the experiment, to allow for a learning process: subjects are less likely to make "random" choices when they have more experience. A similar strategy is adopted here. We specify:

$$\omega_{it} = \omega_0 \exp\left[\omega_1 (TSK_{it} - 1)\right] \tag{8.8}$$

There are now two parameters associated with the tremble: $\omega_0$ represents the tremble probability at the start of the experiment, while $\omega_1$ represents the rate of decay. We expect $\omega_1$ to take a negative value, and the larger it is in magnitude, the faster the implied decay.

For each regime and each subject type, we have the following likelihood contributions for a single response, where $\Phi(.)$ and $\phi(.)$ are the standard normal cdf and pdf respectively:

**Regime 1** ($y = 0$):

$$\begin{aligned}
P(y_{it} = 0 | i = \text{rec}) &= (1 - \omega_{it})\Phi\left(-\frac{\beta_{10} + \beta_{11}MED_{it} + \beta_{13}(TSK_{it} - 1)}{\sigma_1}\right) \\
&\quad + \frac{\omega_{it}}{11} \\
P(y_{it} = 0 | i = \text{str}) &= (1 - \omega_{it})\Phi\left(-\frac{\beta_{20} + \beta_{22}(ORD_{it} - 1) + \beta_{23}(TSK_{it} - 1)}{\sigma_2}\right) \\
&\quad + \frac{\omega_{it}}{11} \\
P(y_{it} = 0 | i = \text{fr}) &= 1 - \frac{10\omega_{it}}{11}
\end{aligned} \tag{8.9a}$$

**Regime 2** $(0 < y < 10)$:

$$f(y_{it}|i=\text{rec}) = (1-\omega_{it})\frac{1}{\sigma_1}\phi\left(\frac{y_{it}-\beta_{10}-\beta_{11}MED_{it}-\beta_{13}(TSK_{it}-1)}{\sigma_1}\right) + \frac{\omega_{it}}{11}$$

$$f(y_{it}|i=\text{str}) = (1-\omega_{it})\frac{1}{\sigma_2}\phi\left(\frac{y_{it}-\beta_{20}-\beta_{22}(ORD_{it}-1)-\beta_{23}(TSK_{it}-1)}{\sigma_2}\right) + \frac{\omega_{it}}{11}$$

$$f(y_{it}|i=\text{fr}) = \frac{\omega_{it}}{11} \tag{8.9b}$$

**Regime 3** $(y = 10)$:

$$P(y_{it}=10|i=\text{rec}) = (1-\omega_{it})\left[1-\Phi\left(\frac{10-\beta_{10}-\beta_{11}MED_{it}-\beta_{13}(TSK_{it}-1)}{\sigma_1}\right)\right] + \frac{\omega_{it}}{11}$$

$$P(y_{it}=10|i=\text{str}) = (1-\omega_{it})\left[1-\Phi\left(\frac{10-\beta_{20}-\beta_{22}(ORD_{it}-1)-\beta_{23}(TSK_{it}-1)}{\sigma_2}\right)\right] + \frac{\omega_{it}}{11}$$

$$P(y_{it}=10|i=\text{fr}) = \frac{\omega_{it}}{11} \tag{8.9c}$$

The manner in which the tremble parameter appears in (8.9) may require explanation. When a tremble occurs, each of the 11 outcomes, 0–10, is each equally likely, hence the term $\omega_{it}/11$ appearing in nearly every equation. In regime 2, what is required is a density, not a probability, so we imagine that when a tremble occurs contributions are realisations from a continuous uniform distribution on $(-0.5, 10.5)$, whence the density associated with any particular realisation is $\omega_{it}/11$.

It is the existence of the three distinct types of subject that leads to a finite mixture model. We introduce three "mixing proportions", $p_{\text{rec}}$, $p_{\text{str}}$, and $p_{\text{fr}}$, which represent the proportion of the population who are reciprocators, strategists, and free-riders respectively. Since these three parameters sum to unity, only two are estimated.

The likelihood contribution for subject $i$ is:

$$\begin{aligned} L_i = {} & p_{\text{rec}}\prod_{t=1}^{T} P(y_{it}=0|\text{rec})^{I_{y_{it}=0}} f(y_{it}|\text{rec})^{I_{0<y_{it}<10}} P(y_{it}=10|\text{rec})^{I_{y_{it}=10}} \\ & + p_{\text{str}}\prod_{t=1}^{T} P(y_{it}=0|\text{str})^{I_{y_{it}=0}} f(y_{it}|\text{str})^{I_{0<y_{it}<10}} P(y_{it}=10|\text{str})^{I_{y_{it}=10}} \\ & + p_{\text{fr}}\prod_{t=1}^{T} P(y_{it}=0|\text{fr})^{I_{y_{it}=0}} f(y_{it}|\text{fr})^{I_{0<y_{it}<10}} P(y_{it}=10|\text{fr})^{I_{y_{it}=10}} \end{aligned} \tag{8.10}$$



where $I_{(.)}$ is the indicator function (taking the value 1 if the subscripted expression is true, 0 otherwise), and the nine conditional probabilities/densities are specified in (8.9) above.

The sample log-likelihood is then:

$$LogL = \sum_{i=1}^{n} \log(L_i) \tag{8.11}$$

$LogL$ is maximised to obtain MLEs of the eight parameters appearing in (8.9), and in addition the two tremble parameters and two of the three mixing proportions. The total number of estimated parameters in the full model is 12. This model may be described as the "finite-mixture two-limit Tobit model with tremble".

### 8.5.5 Program

As mentioned, the panel structure of this data set is a complicating feature. Each subject is observed 20 times. When computing the likelihood contribution for a given subject, we require the joint probability of all 20 of the decisions made by that subject. In practical terms, this means that we require a different likelihood evaluator in the ML program from that used in previous examples.

There are a number of different likelihood evaluators in STATA. The one that was used in the various examples in Chapter 6, and in the examples considered earlier in this chapter, was `lf` (linear form). A feature of the log-likelihood function defined in (8.10) and (8.11) is that it does not satisfy the linear form restrictions, and therefore `lf` cannot be used. This is because the likelihood contributions that need to be summed in order to obtain the sample log-likelihood are not each derived from the information in a single row of the data, but are instead derived from the entire block of rows corresponding to a given subject. There is only one likelihood contribution for each such block of rows. Because of this, the *d-family* evaluators are required in place of the `lf` evaluator. The simplest of these is the `d0` evaluator, which simply requires the log-likelihood contributions to be evaluated. This is the one that will be used here. The `d1` and `d2` evaluators require analytical derivatives of the log-likelihood to be programmed as well as the function evaluation.

The STATA code is presented below. Table 8.2 gives the names in the code corresponding to each of the components in the construction of $LogL$ above. One section of the code which may require explanation is:

```
by i: replace `pp1'=exp(sum(ln(max(`p1',1e-12))))
by i: replace `pp2'=exp(sum(ln(max(`p2',1e-12))))
by i: replace `pp3'=exp(sum(ln(max(`p3',1e-12))))
```

This essentially takes the product of the probabilities contained in (in the first case) p1, over all $T$ observations for subject $i$. The reason why we apply the three functions exp(sum(ln(.))) is simply because, although STATA has a `sum` function

| Component of $LogL$ | STATA name |
|---|---|
| $\beta_{10}, \beta_{11}, \beta_{13}$ | xb1 |
| $\beta_{20}, \beta_{22}, \beta_{23}$ | xb2 |
| $\sigma_1, \sigma_2$ | sig1, sig2 |
| $\omega_0, \omega_1, \omega$ | w0, w1, w |
| $p_{rec}, p_{str}, p_{fr}$ | p_rec, p_str, p_fr |
| $P(y=0\|\text{rec}), P(y=0\|\text{str}), P(y=0\|\text{fr})$ | p1_1, p2_1, p3_1 |
| $f(y\|\text{rec}), f(y\|\text{str}), f(y\|\text{fr}); 0<y<10$ | p1_2, p2_2, p3_3 |
| $P(y=10\|\text{rec}), P(y=10\|\text{str}), P(y=10\|\text{fr})$ | p1_3, p2_3, p3_3 |
| $P(y_{it}=0\|\text{rec})^{I_{y_{it}=0}} f(y_{it}\|\text{rec})^{I_{0<y_{it}<10}} P(y_{it}=10\|\text{rec})^{I_{y_{it}=10}}$ | p1 |
| $P(y_{it}=0\|\text{str})^{I_{y_{it}=0}} f(y_{it}\|\text{str})^{I_{0<y_{it}<10}} P(y_{it}=10\|\text{str})^{I_{y_{it}=10}}$ | p2 |
| $P(y_{it}=0\|\text{fr})^{I_{y_{it}=0}} f(y_{it}\|\text{fr})^{I_{0<y_{it}<10}} P(y_{it}=10\|\text{fr})^{I_{y_{it}=10}}$ | p3 |
| $\prod_{t=1}^{T} P(y_{it}=0\|\text{rec})^{I_{y_{it}=0}} f(y_{it}\|\text{rec})^{I_{0<y_{it}<10}} P(y_{it}=10\|\text{rec})^{I_{y_{it}=10}}$ | pp1 |
| $\prod_{t=1}^{T} P(y_{it}=0\|\text{str})^{I_{y_{it}=0}} f(y_{it}\|\text{str})^{I_{0<y_{it}<10}} P(y_{it}=10\|\text{str})^{I_{y_{it}=10}}$ | pp2 |
| $\prod_{t=1}^{T} P(y_{it}=0\|\text{fr})^{I_{y_{it}=0}} f(y_{it}\|\text{fr})^{I_{0<y_{it}<10}} P(y_{it}=10\|\text{fr})^{I_{y_{it}=10}}$ | pp3 |
| $L_i$ | pp |
| $\log(L_i)$ | logl |
| $P(i=\text{rec}\|y_{i1},...,y_{iT}); P(i=\text{str}\|y_{i1},...,y_{iT}); P(i=\text{fr}\|y_{i1},...,y_{iT})$ | postp1; postp2; postp3 |

Table 8.2: Components of $LogL$ and corresponding STATA names

(which takes the cumulative sum of a variable over observations), it does not have a `product` function. Hence we evaluate the required product by exploiting the identity:

$$\prod_t p_t \equiv \exp\left(\sum_t \ln p_t\right) \tag{8.12}$$

The reason why we take the log of $max(p1, 1e-12)$, rather than simply $p1$, is to prevent the numerical problems that would arise if the probabilities were ever extremely close to zero.

Note that the mixing proportion for the third type ($p_3$) is deduced from $p_1$ and $p_2$ using the delta method. Note also that the final section of the code generates posterior type probabilities. This will be discussed in Section 8.5.7.



The annotated code is as follows.

```
* ESTIMATION OF MIXTURE MODEL FOR BARDSLEY DATA

prog drop _all

* LIKELIHOOD EVALUATION PROGRAM STARTS HERE:

program define pg_mixture
args todo b logl
tempvar p1_1 p2_1 p3_1 p1_2 p2_2 p3_2 p1_3 p2_3 p3_3 p1 p2 p3 pp1 pp2 pp3 pp w

tempname xb1 xb2 sig1 sig2 w0 w1 p_rec p_str

* ASSIGN PARAMETER NAMES TO THE ELEMENTS OF THE PARAMETER VECTOR b:

mleval `xb1' = `b', eq(1)
mleval `xb2' = `b', eq(2)
mleval `sig1' = `b', eq(3) scalar
mleval `sig2'=`b', eq(4) scalar
mleval `w0'=`b', eq(5) scalar
mleval `w1'=`b', eq(6) scalar
mleval `p_rec'=`b', eq(7) scalar
mleval `p_str'=`b', eq(8) scalar

quietly{

* INITIALISE THE p* VARIABLES WITH MISSING VALUES:

gen double `p1_1'=.
gen double `p2_1'=.
gen double `p3_1'=.
gen double `p1_2'=.
gen double `p2_2'=.
gen double `p3_2'=.
gen double `p1_3'=.
gen double `p2_3'=.
gen double `p3_3'=.

gen double `p1'=.
gen double `p2'=.
gen double `p3'=.

gen double `pp1'=.
gen double `pp2'=.
gen double `pp3'=.
gen double `pp'=.

* GENERATE THE TREMBLE PROBABILITY:

gen double `w'=`w0'*exp(`w1'*tsk_1)

* COMPUTE TYPE-CONDITIONAL DENSITIES UNDER REGIME 1:

replace `p1_1'=(1-`w')*normal(-`xb1'/`sig1')+`w'/11
replace `p2_1'=(1-`w')*normal(-`xb2'/`sig2')+`w'/11
replace `p3_1'=1-(10/11)*`w'

* COMPUTE TYPE-CONDITIONAL DENSITIES UNDER REGIME 2:

replace `p1_2'=(1-`w')*(1/`sig1')*normalden((y-`xb1')/`sig1')+`w'/11
replace `p2_2'=(1-`w')*(1/`sig2')*normalden((y-`xb2')/`sig2')+`w'/11
replace `p3_2'=`w'/11

* COMPUTE TYPE-CONDITIONAL DENSITIES UNDER REGIME 3:

replace `p1_3'=(1-`w')*(1-normal((10-`xb1')/`sig1'))+`w'/11
replace `p2_3'=(1-`w')*(1-normal((10-`xb2')/`sig2'))+`w'/11
replace `p3_3'=`w'/11

* MATCH TYPE-CONDITIONAL DENSITIES TO ACTUAL REGIMES (d IS REGIME):

replace `p1' = (d==1)*`p1_1'+(d==2)*`p1_2'+(d==3)*`p1_3'
```

```
replace `p2' = (d==1)*`p2_1'+(d==2)*`p2_2'+(d==3)*`p2_3'
replace `p3' = (d==1)*`p3_1'+(d==2)*`p3_2'+(d==3)*`p3_3'

* FIND PRODUCT OF TYPE-CONDITIONAL DENSITIES FOR EACH SUBJECT:

by i: replace `pp1'=exp(sum(ln(max(`p1',1e-12))))
by i: replace `pp2'=exp(sum(ln(max(`p2',1e-12))))
by i: replace `pp3'=exp(sum(ln(max(`p3',1e-12))))

* COMBINE TYPE-CONDITIONAL DENSITIES TO OBTAIN MARGINAL DENSITY FOR EACH SUBJECT
* (ONLY REQUIRED IN FINAL ROW FOR EACH SUBJECT):

replace `pp'=`p_rec'*`pp1'+`p_str'*`pp2'+(1-`p_rec'-`p_str')*`pp3'
replace `pp'=. if last~=1

* SPECIFY (LOG-LIKELIHOOD) FUNCTION WHOSE SUM OVER SUBJECTS IS TO BE MAXIMISED

mlsum `logl'=ln(`pp') if last==1

* GENERATE POSTERIOR TYPE PROBABILITIES, AND MAKE THESE AVAILABLE OUTSIDE THE
  PROGRAM

replace postp1=`p_rec'*`pp1'/`pp'
replace postp2=`p_str'*`pp2'/`pp'
replace postp3=(1-`p_rec'-`p_str')*`pp3'/`pp'

putmata postp1, replace
putmata postp2, replace
putmata postp3, replace

}

end

* END OF LOG-LIKELIHOOD EVALUATION PROGRAM

clear
set more off

* READ DATA

use "bardsley"

by i: gen last=_n==_N

gen int d=1
replace d=2 if y>0
replace d=3 if y==10

gen double ord_1=ord-1
gen double tsk_1=tsk-1

* SET MEDIAN OF PREVIOUS CONTRIBUTIONS TO 8 FOR SUBJECTS IN FIRST POSITION:

replace med=8 if ord==1

*  SPECIFY EXPLANATORY-VARIABLE LISTS FOR RECIPROCATOR (LIST1)
*  AND STRATEGIST (LIST2) EQUATIONS:

local list1 "med tsk_1"
local list2 "ord_1 tsk_1"

* INITIALISE VARIABLES TO BE USED FOR POSTERIOR TYPE PROBABILITIES:

gen postp1=.
gen postp2=.
gen postp3=.

* SPECIFY STARTING VALUES:

mat start=(0.57,-0.10,6.1,-0.93,-0.05,5.2,3.3,3.7,0.11,-0.05,0.26,0.49)

* SPECIFY LIKELIHOOD EVALUATOR, PROGRAM, AND PARAMETER NAMES:
```



```
ml model d0 pg_mixture (=`list1') (=`list2') /sig1 /sig2 /w0  /w1  /p_rec /p_str
ml init start, copy

* USE ML COMMAND TO MAXIMISE LOG-LIKELIHOOD, AND STORE RESULTS AS "WITH_TREMBLE":

ml max, trace search(norescale)
est store with_tremble

* COMPUTE THIRD MIXING PROPORTION USING DELTA METHOD:

nlcom p_fr: 1-[p_rec]_b[_cons]-[p_str]_b[_cons]

* EXTRACT POSTERIOR TYPE PROBABILITIES AND PLOT THEM AGAINST
*  NUMBER OF ZERO CONTRIBUTIONS:

drop postp1 postp2 postp3

getmata postp1
getmata postp2
getmata postp3

label variable postp1 "rec"
label variable postp2 "str"
label variable postp3 "fr"

by i: gen n_zero=sum(y==0)

scatter postp1 postp2 postp3 n_zero if last==1, title("with tremble") ///
ytitle("posterior probability") msymbol(x Dh Sh) jitter(3) saving(with, replace)

* ESTIMATE MODEL WITHOUT TREMBLE, AND STORE RESULTS AS "WITHOUT_TREMBLE":

constraint 1 [w0]_b[_cons]=0.00
constraint 2 [w1]_b[_cons]=0.00

ml model d0 pg_mixture (=`list1') (=`list2') ///
/sig1 /sig2 /w0  /w1  /p_rec /p_str, constraints(1 2)

ml init start, copy
ml max, trace search(norescale)
est store without_tremble

nlcom p_fr: 1-[p_rec]_b[_cons]-[p_str]_b[_cons]

* EXTRACT AND PLOT POSTERIOR TYPE PROBABILITIES FOR MODEL WITHOUT TREMBLE:

drop postp1 postp2 postp3

getmata postp1
getmata postp2
getmata postp3

label variable postp1 "rec"
label variable postp2 "str"
label variable postp3 "fr"

scatter postp1 postp2 postp3 n_zero if last==1, title("without tremble") ///
ytitle("posterior probability") msymbol(x Dh Sh) jitter(3) saving(without, replace)

* CARRY OUT LIKELIHOOD RATIO TEST FOR PRESENCE OF TREMBLE:

lrtest with_tremble without_tremble

* COMBINE THE TWO POSTERIOR PROBABILITY PLOTS

gr combine with.gph without.gph
```

The model is estimated twice, first with all parameters unconstrained, and second with the two tremble parameters constrained to zero. Note that all that is required for this is to define the two constraints using the `constraint` command, and then to include the `constraints(.)` option with the `ml` command.

The STATA output from the first estimation (the model with tremble) is as follows:

```
                                                  Number of obs   =      1960
                                                  Wald chi2(2)    =    108.07
Log likelihood = -3267.6884                       Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
         med |   .5986768   .0611812     9.79   0.000      .478764    .7185897
       tsk_1 |  -.0961738   .0202228    -4.76   0.000    -.1358099   -.0565378
       _cons |   4.004373   .4541826     8.82   0.000     3.114191    4.894555
-------------+----------------------------------------------------------------
eq2          |
       ord_1 |  -.9644647   .0823747   -11.71   0.000    -1.125916   -.8030133
       tsk_1 |  -.0516767   .0171891    -3.01   0.003    -.0853668   -.0179866
       _cons |   5.299356   .3828538    13.84   0.000     4.548976    6.049736
-------------+----------------------------------------------------------------
sig1         |
       _cons |   3.442242   .1674648    20.56   0.000     3.114017    3.770467
-------------+----------------------------------------------------------------
sig2         |
       _cons |   3.705603   .1611296    23.00   0.000     3.389795    4.021411
-------------+----------------------------------------------------------------
w0           |
       _cons |   .1041737   .0321191     3.24   0.001     .0412213    .1671261
-------------+----------------------------------------------------------------
w1           |
       _cons |  -.0492261   .0218192    -2.26   0.024    -.0919909   -.0064612
-------------+----------------------------------------------------------------
p_rec        |
       _cons |   .2710854   .0484671     5.59   0.000     .1760916    .3660791
-------------+----------------------------------------------------------------
p_str        |
       _cons |   .4832813   .0538023     8.98   0.000     .3778307    .5887319
------------------------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        p_fr |   .2456333   .0436144     5.63   0.000     .1601508    .3311159
------------------------------------------------------------------------------
```

As recommended by Moffatt & Peters (2001), the likelihood ratio test has been used to test for the presence of a tremble. Results from the likelihood ratio test comparing the above model with the tremble-free model are as follows. The p-value of 0.0000 represents overwhelming evidence of the presence of a tremble.

```
. lrtest with_tremble without_tremble

Likelihood-ratio test                                 LR chi2(2)  =    149.89
(Assumption: without_trem~e nested in with_tremble)   Prob > chi2 =    0.0000
```

Results from both models are presented and discussed in the next section.

### 8.5.6 Results

The parameter estimates from the finite mixture two-limit Tobit model, with and without tremble, are shown in Table 8.3. The first column contains results from the



full model. The second column shows the results of the model with no tremble. As noted at the end of the last section, the likelihood ratio test provided overwhelming evidence of the presence of a tremble in this data set, implying that the full model is superior. The importance of including the tremble is also clearly seen by observing how different the estimates are, in particular the mixing proportions, when the tremble is absent. For example, in the absence of a tremble, the proportion of free riders is estimated to be 0.143 which is exactly equal to the proportion of subjects who contributed zero on every occasion (see Section 8.5.3); as mentioned previously, the presence of the tremble allows the set of free-riders to include subjects who contributed zero on *nearly* all occasions, and accordingly, the estimate of the proportion of free-riders rises to 0.246, which is remarkably close to the proportion of subjects non-parametrically identified as free riders in Section 8.5.3, on the grounds that they contributed zero in at least 16 out of 20 tasks. This estimate of 24.6% is also in close agreement with previous estimates appearing elsewhere in the literature (see e.g. Gächter & Fehr, 2000).

The two equations set out in (8.6) are estimated as:

$$\text{Reciprocators}: \quad E(y^*|MED, TSK) = 4.004 + 0.599MED - 0.096(TSK-1) \qquad (8.13a)$$

$$\text{Strategists}: \quad E(y^*|ORD, TSK) = 5.299 - 0.964(ORD-1) - 0.052(TSK-1) \qquad (8.13b)$$

As seen in Table 8.3, all coefficients are strongly significant. For reciprocators, as expected, the median of previous contributions has a significantly positive effect on the current contribution: if all previous contributions were raised by one unit, we would expect the current contribution to rise by around three-fifths of one unit, but by significantly less than one whole unit. This result is consistent with the biased reciprocity observed in Fischbacher et al. (2001) (biased in the sense that subjects, although influenced positively by the contributions of others, tend to donate less than the levels contributed by others).

For strategists, again as expected, the effect of the subject's order in the sequence is negative. In particular, the "expected" contribution of a strategist first-mover (in task 1) is 5.3, while the same strategist in last position ($ORD = 7$) would be expected to contribute zero – a highly reassuring result, since as we noted earlier there is no selfish contribution motive for a subject in last position.

The effect of TSK is significantly negative for both types, simply implying a diminution of contributions with experience. If this is interpreted as the effect of learning about the incentive structure of the game, it seems that reciprocators learn about such matters somewhat faster than strategists.

The tremble probability is 0.104 at the start of the experiment (task 1), but, in accordance with the significant negative estimate of $\omega_1$, decays to 0.041 by the end (task 20). This dramatic decay of the tremble amounts to further evidence of learning (Moffatt & Peters, 2001; Loomes et al., 2002).

Turning to the estimates of the mixing proportions, we see that very close to 25% of the population are free-riders; around 25% are reciprocators; the remaining 50% are strategists.

| | Full model | No tremble |
|---|---|---|
| **Reciprocators** | | |
| constant | 4.004(0.454) | 3.166(0.358) |
| MED | 0.599(0.061) | 0.490(0.045) |
| TSK-1 | −0.096(0.020) | −0.061(0.015) |
| $\sigma_1$ | 3.442(0.167) | 3.577(0.126) |
| **Strategists** | | |
| constant | 5.299(0.382) | 4.493(0.518) |
| ORD-1 | −0.964(0.082) | −1.128(0.102) |
| TSK-1 | −0.052(0.017) | −0.080(0.023) |
| $\sigma_2$ | 3.706(0.161) | 5.104(0.253) |
| **Tremble** | | |
| $\omega_0$ | 0.104(0.032) | — |
| $\omega_1$ | −0.049(0.022) | — |
| **Mixing proportions** | | |
| $p_{\text{rec}}$ | 0.271(0.048) | 0.382(0.051) |
| $p_{\text{str}}$ | 0.483(0.054) | 0.472(0.053) |
| $p_{\text{fr}}$ | 0.246(0.044) | 0.143(0.035) |
| n | 98 | 98 |
| T | 20 | 20 |
| k | 12 | 10 |
| LogL | −3267.69 | −3342.63 |
| AIC | 3.35 | 3.42 |

Table 8.3: Maximum likelihood estimates from mixture model applied to Bardsley's (2000) data, with and without tremble

Notes: Asymptotic standard errors in parentheses. The estimate and standard error of $p_{fr}$ is deduced from the estimates of $p_{rec}$ and $p_{str}$ using the delta method. When ORD=1, MED is set to 8 for the purpose of estimation. AIC is Akaike's Information Criterion, defined as $2(-LogL+k)/(nT)$, where $k$ is the number of parameters in the model. The preferred model is the one with the lower AIC.

### 8.5.7 Posterior type probabilities

The three posterior type probabilities are given by:

$$P(i=\text{rec}|y_{i1},\ldots,y_{iT}) = \frac{p_{\text{rec}}\prod_{t=1}^{T}P(y_{it}=0|\text{rec})^{I_{y_{it}=0}}f(y_{it}|\text{rec})^{I_{0<y_{it}<10}}P(y_{it}=10|\text{rec})^{I_{y_{it}=10}}}{L_i}$$

$$P(i=\text{str}|y_{i1},\ldots,y_{iT}) = \frac{p_{\text{str}}\prod_{t=1}^{T}P(y_{it}=0|\text{str})^{I_{y_{it}=0}}f(y_{it}|\text{str})^{I_{0<y_{it}<10}}P(y_{it}=10|\text{str})^{I_{y_{it}=10}}}{L_i}$$



$$P(i=\text{fr}|y_{i1},\ldots,y_{iT}) = \frac{p_{\text{fr}}\prod_{t=1}^{T}P(y_{it}=0|\text{fr})^{I_{y_{it}=0}}f(y_{it}|\text{fr})^{I_{0<y_{it}<10}}P(y_{it}=10|\text{fr})^{I_{y_{it}=10}}}{L_i}$$

where $L_i$ is the likelihood contribution for subject $i$, defined in (8.10). These posterior probabilities are computed (as postp1-postp3) at the end of the program.

In Figure 8.7, we plot the three posterior probabilities, obtained from estimation of both models, against the number of zero contributions made by the subject. Comparing the two plots, we see once again that the main difference between the two models is in the assignation of subjects to the "free-rider" type. For the tremble-free model (right-hand plot), only subjects contributing zero in all 20 tasks are assigned the status of "free-rider". For the model with tremble however (left-hand plot), all subjects who contributed zero in 16 or more tasks are seen to be very likely to be free-riders. Further inspection of the left-hand plot reveals that subjects who contribute zero in a moderate number of tasks (6–14) tend to be strategists, while subjects who rarely contribute zero appear to be a mixture of strategists and reciprocators. Note finally that very few points in the scatter that are far from zero or one on the vertical axis, indicating that it is only for a small number of subjects that the model is incapable of detecting type with confidence.

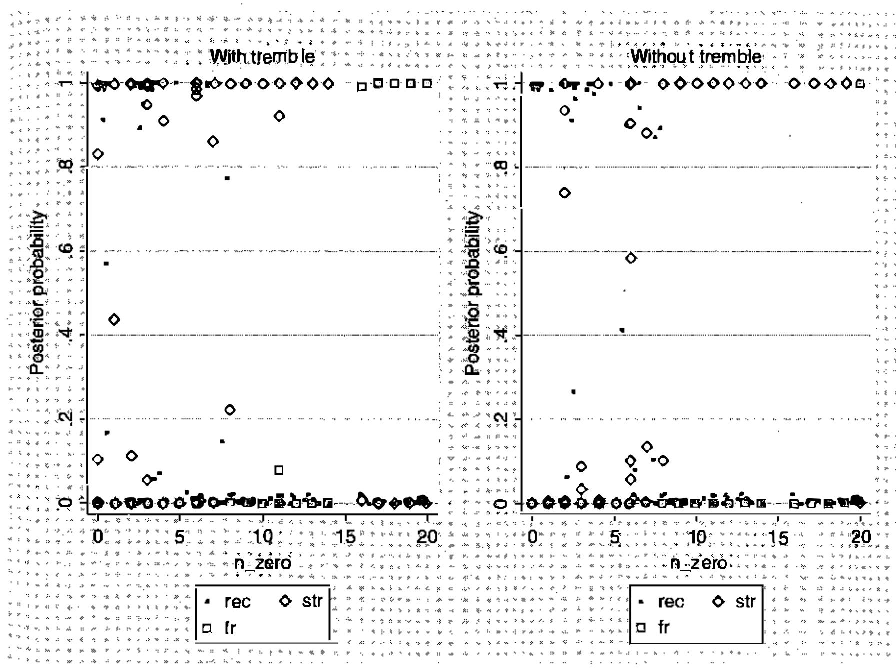



Figure 8.7: Jittered scatter of posterior type probabilities against number of zero contributions from model with tremble (left-hand graph) and from model without tremble (right-hand graph)

## 8.6 Summary and Further Reading

The theme of this chapter has been finite mixture modelling. A popular textbook covering this topic is McLachlan & Peel (2000). An early application of finite mixture modelling in experimental economics was El-Gamal & Grether (1995). More recently, similar models have been used by Bardsley & Moffatt (2007), Cappelen et al. (2007), Conte et al. (2011), and Conte & Moffatt (2014).

It has been stressed that the recommended approach to mixture modelling is to set the number of types and to specify a parametric model to define each type, before estimation commences. Perhaps readers should be aware that an alternative approach to estimation, avoided in this book, is to allow the data to determine the number of types, as well as the equations defining them. The resulting models come under the heading of "latent class models" which have been surveyed recently by Collins & Lanza (2010).

The first (real) application in the chapter was to the "acquiring a company" game, and interested readers are referred to Bazerman & Samuelson (1983) and Ball et al. (1991).

The second application was to a public goods game. An overview of data from such experiments is provided by Davis & Holt (1993, ch. 5) and Ledyard (1995). Several authors have made distinctions between different types of agent, including free-riders and reciprocators (Gächter & Fehr, 2000). The mixture model estimated here uses data from Bardsley (2000) and the model itself is similar to that of Bardsley & Moffatt (2007).

# Chapter 9
# Simulating Experimental Data, and the Monte Carlo Method

## 9.1 Introduction

Simulated data is an important ingredient of this book. The purpose of this chapter is to demonstrate the methods used to simulate data. The examples will start at a very simple level, and then progress to data sets with more complicated features.

Simulation is important for many reasons. Firstly, it is useful to be able to simulate a data set in a situation in which a real data set is unavailable, if only in order to demonstrate techniques for analysing the data to others. Of course, it must not be forgotten that the data is simulated – conclusions relating to behaviour should never be drawn! It is nevertheless important to simulate the data in such a way as to resemble real data as closely as possible, for example by incorporating any anomalies that are known to exist. Many of the data sets used in this book have been simulated with this objective in mind.

A second reason why simulation is useful is in trying out a new estimation program. When writing an estimation program, it is perfectly normal to make many errors. Sometimes the errors are obvious because the program does not run at all. Other times, the program runs and gives estimates, but the estimates are incorrect as a result of errors in the code. This type of error is harder to detect. Simulation is a useful method of establishing that a program is doing exactly what it was designed to do, before setting it loose on real data. This is because, when data is simulated, the "true" parameter values are obviously known, and the correctness of the code may be judged, roughly speaking, by how close the estimates are to the true values.

A third major use of simulation is in Monte Carlo investigations. Two different simulations give two different data sets, and therefore two different sets of parameter estimates. It is actually desirable to conduct a large number of simulations, each giving a different set of estimates, and then to observe the distribution of the resulting estimates. We are mainly interested in whether the distributions are centred on the true values (implying that the estimates are unbiased), and in the variance of the estimates around the true values (this representing estimation precision). Such a procedure is known as a Monte Carlo study. The procedure is also commonly applied to investigate the performance of test statistics, where the central objective is to see how closely the distribution of the test statistic corresponds to the theoretical distribution assumed under the null hypothesis.

When conducting a Monte Carlo study, a central concept is the data generating process (DGP) which must be decided at the outset. This normally consists of deterministic and random components, with the distributions of the latter being fully specified, along with the sample size. Repetitions of the simulation are referred to as *replications*. Within each replication, the DGP is simulated, and a model is estimated.

The estimated model is not necessarily the same as the DGP. In fact, a common use of Monte Carlo studies is the investigation of the impact on the properties of estimators from misspecifying the model (i.e. using an estimated model that is deliberately different from the DGP). The approach is particularly useful in situations in which theory is not informative (e.g. at low sample sizes).

## 9.2 Random Number Generation in STATA

An important part of any simulation is the generation of random numbers from a chosen distribution. In STATA, the available distributions include:[1]

```
gen double u=runiform() //  [Uniform(0,1)]
gen double z=rnormal()   //  [Normal(0,1)]
gen int n=rpoisson(3)  //  [Poisson(3)]
```

If a non-standard normal distribution is required, the required mean and standard deviation may be inserted as arguments of the rnormal function (e.g. `gen double x=rnormal(3,2)`). Another way of doing this is to transform the standard normal variable, by multiplying by the standard deviation and adding the mean.

As a first example, the following command sequence generates 1000 random numbers from a standard normal distribution, transforms them to a $N(3, 2^2)$ distribution, and then creates a histogram (Figure 9.1) that allows us to examine the resulting distribution.

```
set obs 1000
gen double z=rnormal()
gen double x=3+2*z
hist x, normal
```

The `normal` option on the `histogram` command causes a normal density function to be superimposed. We see that the random numbers are close to being normally distributed. Naturally, a larger sample size would be expected to yield a frequency distribution that is closer to the theoretical counterpart.

If the above sequence of commands were repeated, a different set of random numbers would be generated, resulting in a different frequency distribution to the one shown below. Sometimes, it is desirable to generate the same set of random

[1] The "//" used in these commands is a way of introducing comments within command lines in do-files; everything following the "//" will be treated as a comment.



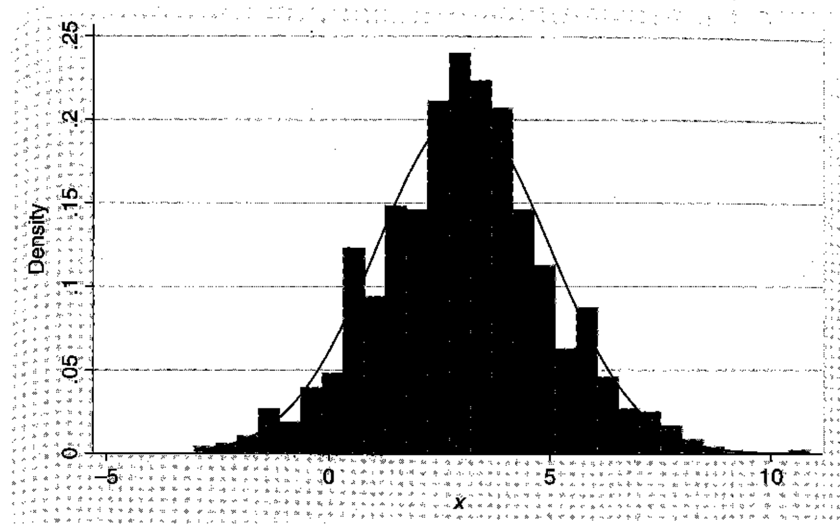



Figure 9.1: Histogram of 1000 random numbers from $N(3, 2^2)$

numbers over and over (if, for example, two different researchers need to generate exactly the same results). In order to ensure that each set of random numbers is the same, a random number seed needs to be set at the start of the program. The necessary command is:

```
set seed 7654321
```

The number appearing in this command (i.e. the "seed") can be any number between 0 and $2^{31}-1$ $(= 2, 147, 483, 647)$. The chosen seed exactly determines the sequence of random numbers generated by the command. This is actually because, strictly speaking, the random numbers are not "random" at all; they are pseudo-random numbers, meaning that they are determined by an algorithm whose only input is the random number seed. The algorithm is designed so as to produce a sequence of numbers that resembles a sequence of truly random numbers as closely as possible. Inasmuch as we can take it that the built-in STATA algorithm succeeds in doing this, we may take the liberty of referring to the generated sequences as "random".

The code shown above includes the use of a transformation which transforms a standard normal variable into a $N(3, 2^2)$ variable. Other types of transformation are useful. Most importantly, if the further transformation were applied to the variable $x$:

```
gen double y=exp(x)
```

then the new variable $y$ would be from a lognormal$(3, 2^2)$ distribution.

Another transformation that is sometimes required is that from a uniform to a normal. The following code generates a uniform(0, 1) variable $u$, and then converts it into a standard normal using the transformation $z = \Phi^{-1}(u)$, that is, by applying the inverse normal cdf:

```
gen double u=runiform()
gen double z=invnorm(u)
```

This transformation is very important when working with Halton draws (see Chapter 10). Halton draws are treated as draws from a uniform(0, 1) distribution, and typically the first thing that needs to be done to them is to transform them to normal draws using `invnorm()`.

## 9.3 Simulating Data Sets

### 9.3.1 *Simulating data from a linear model*

In this section, we present an example in which data is simulated from a linear regression model, and then the model is estimated using the simulated data. This is a very simple version of an approach that is used later in the book for more complex models.

The model we assume is as follows. As remarked in Section 9.1, this model is what is known as the DGP for the simulation.

$$\begin{aligned} y_i &= 2.0 + 0.5x_i + u_i \qquad i = 1, \ldots, 100 \\ x_i &\sim U(0, 1) \\ u_i &\sim N(0, 1) \end{aligned} \tag{9.1}$$

The sequence of commands for simulating (and then estimating) (9.1) is:

```
set obs 100
set seed 7654321
gen double x=runiform()
gen double u=rnormal()
gen y=2.0+1.0*x+0.5*u
scatter y x, ylabel(0(1)4)
regress y x
```

The scatter plot of the simulated y against x is shown in Figure 9.2. As expected, we see a data set with quite a lot of noise, although there is a clearly discernible positive relationship between the two variables.

The results from the regression are:

```
. regress y x

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =   28.65
       Model |  6.71489828     1  6.71489828           Prob > F      =  0.0000
    Residual |  22.9728286    98  .234416618           R-squared     =  0.2262
-------------+------------------------------           Adj R-squared =  0.2183
       Total |  29.6877269    99  .299876029           Root MSE      =  .48417

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .8213198    .153457     5.35   0.000     .5167893     1.12585
       _cons |   2.088775   .0905173    23.08   0.000     1.909147    2.268404
------------------------------------------------------------------------------
```

Both the intercept and the slope are estimated "correctly" in the sense that the estimates are close to the true values (2 and 1 respectively), and the true values



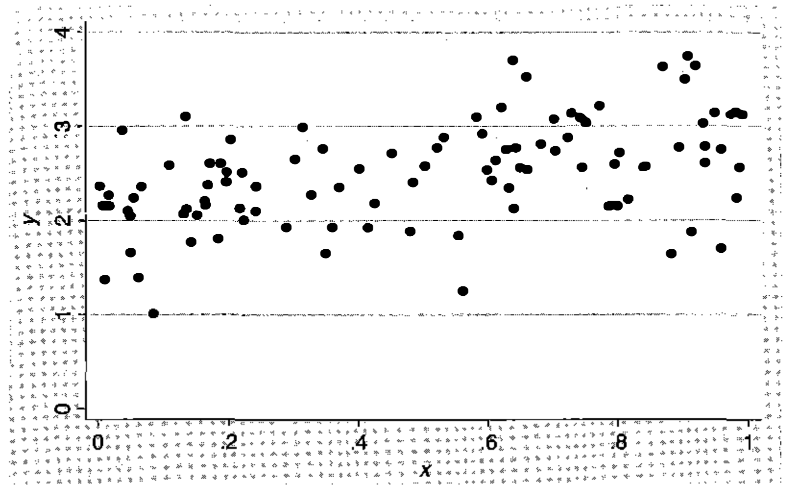



Figure 9.2: Scatter plot of simulated data from linear model

are within the respective confidence intervals. We also see that the estimate of the standard deviation of the error term ("Root MSE" in the results) is very close to the true value of 0.50.

To extend the single simulation to a Monte Carlo simulation, the sequence of commands (without the `scatter` command) needs to be contained in a program, and then the program called by the `simulate` command. Note that the $x$ variable is generated outside of the program because this variable should be fixed between replications (to ensure that $x$ is available inside the program, the command `keep x` is used). We name the program `ols_monte`. We will store the slope estimate, the intercept estimate, and the RMSE from each replication. The required sequence of commands is as follows.

```
capt prog drop ols_monte

  program define ols_monte, rclass
       syntax [, obs(integer 1)  ]
       set obs `obs'
keep x
tempvar u y

gen double u=rnormal()

gen y=2.0+1.0*x+0.5*u

regress y x

return scalar slope = _b[x]
return scalar intercept = _b[_cons]
return scalar rmse=e(rmse)

end

clear
set obs 100
set seed 7654321
gen double x=runiform()
simulate slope=r(slope) intercept=r(intercept) rmse=r(rmse) ///
, reps(1000): ols_monte, obs(100)
```

The syntax command allows us to alter the sample size when the simulate command is used. The simulate command first specifies the three quantities that we wish to extract from the simulations; then it specifies the number of replications to be performed; then it specifies the name of the program to be called; finally, it specifies the sample size to be simulated within each replication.

The results from the above simulate command are stored as three variables (named slope, intercept and rmse) each with 1000 observations. Summary statistics for the three variables are:

```
. summ

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       slope |      1000    1.000787    .1619344   .5037797   1.545957
   intercept |      1000    1.999025    .0962207    1.69315   2.257928
        rmse |      1000    .4992077    .0362079   .3740984   .6196097
```

We see that the means of the three estimates are very close to their respective true values (1, 2, and 0.5), confirming that all three estimators are unbiased. We also note that the standard deviations of the slope and intercept estimates are close to the standard errors estimated for these estimates when the regression was performed once (0.15 and 0.09 respectively). This confirms that these standard error estimates are also centred close to their true values (strictly speaking, it is incorrect to claim that ols standard errors are "unbiased").

### 9.3.2 Simulating panel data

As we well know, data resulting from economic experiments are usually panel data. It is therefore important to be able to incorporate the panel structure into a simulation.

Let us start by considering how to simulate data from the following random effects model:

$$y_{it} = 2.0 + 0.5x_{it} + u_i + \epsilon_{it} \qquad i = 1, ..., 50 \quad t = 1, ..., 20 \qquad (9.2)$$
$$x_{it} \sim U(0,1)$$
$$u_i \sim N\left(0, 0.5^2\right)$$
$$\epsilon_{it} \sim N(0,1)$$

The first thing we need to do is to generate the subject identifier, $i$, and the task identifier, $t$. This is done as follows:

```
set obs 1000
egen int i=seq(), f(1) t(50) b(20)
egen int t=seq(), f(1) t(20)
xtset i t
```

Recall that the purpose of the xtset command is to declare i and t as the variables defining the structure of the panel.



We then generate the $x$ variable and the two random terms:

```
set seed 7654321
gen x=runiform()
gen e=rnormal()

by i: generate double u=0.5*rnormal() if _n==1
by i: replace u=u[1] if u==.
```

The slightly tricky error term to generate is the "between-error", $u$, because it needs to take the same value in all of the rows for a given subject. This requires two lines (the last two lines of the above sequence). The first puts the random number from $N\left(0, 0.5^2\right)$ into the first row for each subject. The second command simply copies this value into all of the other rows for the subject.

Finally we generate the dependent variable:

```
gen double y=2+0.5*x+u+e
```

Now the panel data set is complete, and we may estimate the model. The correct estimator to use in this situation is the random effects estimator. In this case, the results are:

```
. xtreg y x, re

Random-effects GLS regression                   Number of obs      =      1000
Group variable: i                               Number of groups   =        50

R-sq:  within  = 0.0394                         Obs per group: min =        20
       between = 0.0305                                        avg =      20.0
       overall = 0.0357                                        max =        20

                                                Wald chi2(1)       =     40.06
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6939987   .1096485     6.33   0.000     .4790916    .9089058
       _cons |   1.951544   .0934667    20.88   0.000     1.768353    2.134735
-------------+----------------------------------------------------------------
     sigma_u |  .49402351
     sigma_e |  1.0085816
         rho |  .19349876   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

We see that the estimates of all four parameters (the slope, the intercept, and the two variance components) are, as expected, close to their true values.

### 9.3.3 Simulating dynamic panel data

In certain applications, the response for a given subject depends partly on that subject's response in the previous task. Models capturing this situation are known as dynamic panel data models. Consider the following DGP:

$$y_{it} = 0.0 + 0.3y_{i,t-1} + 0.5x_{it} + u_i + \epsilon_{it} \qquad i = 1, ..., 50 \quad t = 2, ..., 20 \quad (9.3)$$
$$x_{it} \sim U(0,1)$$
$$u_i \sim N\left(0, 0.5^2\right)$$
$$\epsilon_{it} \sim N(0,1)$$

The presence of $y_{i,t-1}$ in the first line of (9.3) is what makes it a dynamic panel data model. Note also that the task identifier $t$ can only take values from 2 to 20 because of the presence of the lag.

Data from the DGP (9.3) is simulated as follows:

```
set obs 1000

egen int i=seq(), f(1) t(50) b(20)
egen int t=seq(), f(1) t(20)
xtset i t

set seed 654321

gen x=runiform()
gen e=rnormal()

by i: generate double u=0.5*rnormal() if _n==1
by i: replace u=u[1] if u==.

by i: gen double y=0 if _n==1
by i: replace y=0.0+0.3*y[_n-1]+0.5*x+u+e  if y==.
```

The key part of this code is the pair of commands used to generate $y$. The first of these generates a variable $y$ with zero in the first row for each subject, and missing values in every other row. The second fills in the missing values sequentially, with `y[_n-1]` representing the $y$-value in the previous period.

The standard approach to estimation in the dynamic panel data model is the consistent generalised method of moments (GMM) estimator of Arellano & Bond (1991). The STATA command is `xtabond y x`. Note that the lagged dependent variable should not be included in the command. Applying this estimator to the data simulated above gives the following results:

```
. xtabond y x

Arellano-Bond dynamic panel-data estimation  Number of obs      =        900
Group variable: i                            Number of groups   =         50
Time variable: t
                                             Obs per group:   min =       18
                                                              avg =       18
                                                              max =       18

Number of instruments =    155               Wald chi2(2)       =      73.17
                                             Prob > chi2        =     0.0000
One-step results
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           y |
         L1. |   .3225384   .0393107     8.20   0.000     .2454908     .399586
             |
           x |   .3074247   .1298663     2.37   0.018     .0528914    .5619579
       _cons |   -.00245    .0735888    -0.03   0.973    -.1466814    .1417815
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).y
        Standard: D.x
Instruments for level equation
        Standard: _cons
```

We see that all three estimates, coefficient oflagged-$y$, coefficient of $x$, and constant, are close to their true values (0.3, 0.5 and 0 respectively) in the sense that each confidence interval contains the true value.



### 9.3.4 Simulating binary panel data

Binary panel data is perhaps even more relevant to Experimetrics than continuous panel data such as that considered in the last two sub-sections. For this reason, more attention will be paid to binary panel data. We will also use the Monte Carlo simulation as an opportunity to investigate the consequences of neglecting between-subject heterogeneity in this context.

Consider the following DGP:

$$
\begin{aligned}
y_{it}^* &= -1.0 + 2.0x_{it} + u_i + \epsilon_{it} \qquad i = 1, ..., 50 \quad t = 1, ..., 20 \qquad (9.4)\\
y_{it} &= I(y_{it}^* > 0)\\
x_{it} &\sim U(0, 1)\\
u_i &\sim N(0, 1)\\
\epsilon_{it} &\sim N(0, 1)
\end{aligned}
$$

$I(.)$ appearing in the second line of (9.4) is the indicator function, taking the value 1 if the statement in parentheses is true, and zero otherwise.

The following code is for a Monte Carlo simlulation of the DGP (9.4) with 1,000 replications. Again the $x$ variable is simulated outside the program to ensure that it is fixed over replications. The main differences between this piece of code and the one used in the last sections for fully observed data are the presence of the variable `ys`, which corresponds to the latent variable $y^*$ in the DGP, and the command `gen y=ys>0` which creates a binary variable y being 1 if $ys > 0$, and 0 otherwise. Within each replication, two different models are estimated: random effects probit (which is the correct model); and pooled probit (which neglects the between-subject heterogeneity embodied in $u$). The slope estimate (i.e. the coefficient of $x$) resulting from each model is stored (as "slope_pooled" and "slope_re").

```
program define panel_monte, rclass
        syntax [, obs(integer 1)  ]
        set obs `obs'
keep x
tempvar i t z1 z2 ee e u v ys y


egen i=seq(), f(1) t(50) b(20)
egen t=seq(), f(1) t(20)
tsset i t


gen double e=rnormal()

by i: generate u=rnormal() if _n==1
by i: replace u=u[1] if u==.

gen ys=-1.0+2.0*x+u+e
gen y=ys>0

probit y x
return scalar slope_pooled = _b[x]

xtprobit y x, re i(i)
return scalar slope_re = _b[x]
```

```
end

clear
set obs 1000
gen double x=runiform()
simulate slope_pooled=r(slope_pooled) slope_re=r(slope_re), ///
reps(1000): panel_monte, obs(1000)
```

The histograms in Figure 9.3 show the distributions of the two estimates. Remember that the true value of the slope parameter is 2.0. The left panel shows a distribution that is centred on 2.0, and therefore confirms that random effects probit model leads to unbiased estimation. The right panel shows the distribution of the slope estimate from the pooled probit model. This shows that there is a serious downward bias in the slope estimate, of around 30%, when pooled probit is used.

This example is very useful because it clearly demonstrates the serious consequences of assuming that all subjects are identical when there are differences between them, and it also demonstrates the benefits from estimating a model (in this case random effects probit) that allows for between-subject heterogeneity.

## 9.4 Monte Carlo Investigation of the Hausman Test

In Section 7.6, in the chapter on ordinal data, we presented an example in which the ordered probit model was used to analyse data on the strength of preference in a risky choice experiment. We assumed that, in addition to choosing between a risky

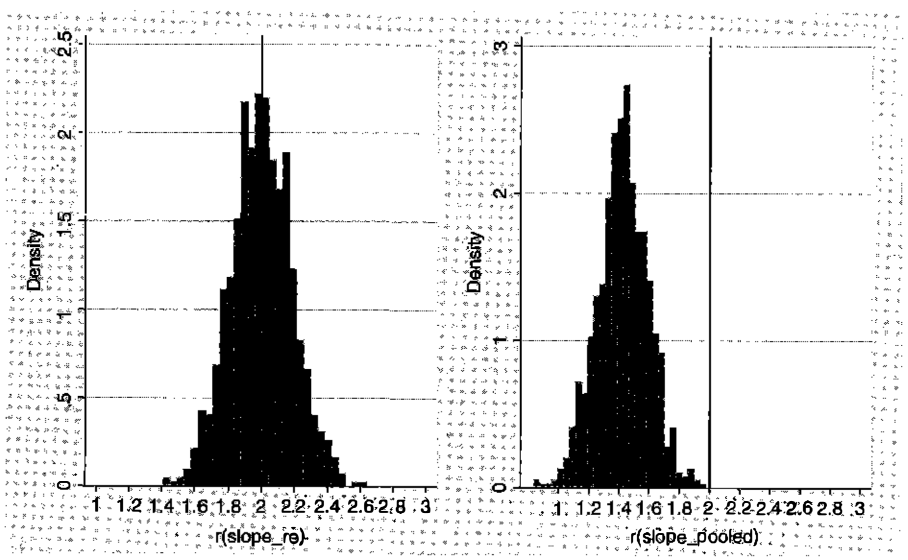



Figure 9.3: Distribution of slope estimate from random effects probit (left panel); distribution of slope estimate from pooled probit (right panel)



and a safe lottery, subjects were asked to indicate how sure they were about their choice: not sure; fairly sure; or completely sure. It was explained that this results in ordinal data which can be modelled using ordered probit. It was also explained that the results from ordered probit are superior to the results from binary probit provided subjects truthfully report their strength of preference. It was then explained how a Hausman test comparing the two sets of estimates may be used as a test for the truthfulness of strength-of-preference responses.

Here, we use the Monte Carlo method to investigate the properties of the Hausman test just described, and to verify that it does what is intended. We simulate a large number of data sets (under both the null and the alternative), and for each, we compute the Hausman test statistic. We then investigate the size and power performance of the test.

The null hypothesis for this test is that subjects respond truthfully when asked to state their strength of preference. The alternative hypothesis that we propose is that a proportion of subjects ($p$ say) respond completely at random when asked for strength of preference, although they continue to respond truthfully when initially asked to choose a lottery. We could then state that the null hypothesis under test is $p = 0$, and it is being tested against the alternative $p > 0$.

In the simulation, we set $p = 0.30$. We require two different outcomes for strength of preference. We first generate "y_good" as the truthful outcome, and "y_bad" as the completely random outcome, in which all subjects choose randomly. We then generate "y_poor" as the partially random outcome, in which only 30% choose randomly. In addition, we generate "y_bin" which is the binary outcome. The test statistics generated using the two variables ("y_good" and "y_poor") are named "hausman1" and "hausman2" respectively. The sample size, and the explanatory variable $w$, are exactly as used in Section 7.6.

The code is as follows. One line may require explanation. The variable named `three` is generated as a discrete uniform variable on $\{1, 2, 3\}$. Hence the line `gen y_bad= 3*y_bin+three` generates, as required, an ordinal variable (y_bad) taking values between 1 and 6. However, when y_bin=0 (i.e. when the risky lottery is chosen), y_bad is randomly distributed between 1 and 3, while when y_bin=1 (i.e. safe is chosen), y_bad is randomly distributed between 4 and 6.

```
  program define hausman, rclass
      syntax [, obs(integer 1)  ]
      drop _all
      set obs `obs'

* GENERATE SUBJECT ID AND WEALTH (w)

egen subj=fill(1/2)
egen w=seq(), f(0) t(20) b(50)
replace w=w/2

* GENERATE LATENT VARIABLE y*

gen ystar=1.5-0.2*w+rnormal()

* GENERATE BINARY VARIABLE y_bin

gen y_bin=ystar>0

* SET CUT-POINTS
```

```
scalar k1=-1
scalar k2=-0.5
scalar k3=0
scalar k4=.5
scalar k5=1

* GENERATE TRUTHFUL OUTCOME "y_good"

gen y_good=1+(ystar>k1)+(ystar>k2)+(ystar>k3)+(ystar>k4)+(ystar>k5)

* GENERATE "COMPLETELY RANDOM" OUTCOME "y_bad"

gen u=uniform()
gen three=1+(u>0.33)+(u>0.67)
gen y_bad=3*y_bin+three

* GENERATE PARTIALLY RANDOM OUTCOME "y_poor"

gen y_poor=y_good
replace y_poor=y_bad if uniform()<0.3

* ESTIMATE BINARY PROBIT

probit y_bin w
mat b_p=e(b)
mat V_p=e(V)

* ESTIMATE ORDERED PROBIT USING TRUTHFUL OUTCOME (I.E. UNDER NULL)

oprobit y_good w
mat b_op=e(b)
mat V_op=e(V)

* COMPUTE HAUSMAN TEST STATISTIC UNDER NULL

return scalar  hausman1=(b_p[1,1]-b_op[1,1])^2/(V_p[1,1]-V_op[1,1])

* ESTIMATE ORDERED PROBIT USING PARTIALLY RANDOM OUTCOME (I.E. UNDER ALTERNATIVE)

oprobit y_poor w
mat b_op=e(b)
mat V_op=e(V)

* COMPUTE HAUSMAN TEST STATISTIC UNDER ALTERNATIVE

return scalar hausman2=(b_p[1,1]-b_op[1,1])^2/(V_p[1,1]-V_op[1,1])

end

* RUN SIMULATION, WITH 1000 REPLICATIONS; SET SAMPLE SIZE TO 1050

simulate hausman1=r(hausman1) hausman2=r(hausman2) , reps(1000): hausman, obs(1050)
```

The size performance of the test is investigated by considering the distribution of hausman1.

```
. summ hausman1, detail

                         r(hausman1)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .0001179       1.94e-06
 5%     .0054242       4.30e-06
10%     .0199324       .0000106       Obs                1000
25%     .1131179       .0000199       Sum of Wgt.        1000

50%     .4849418                      Mean           1.031761
                       Largest        Std. Dev.      1.413984
75%     1.434592       8.040253
90%     2.811562       9.071658       Variance       1.999351
95%     3.698014       9.883641       Skewness       3.045492
99%     6.178411        16.1178       Kurtosis       20.28341
```



We see that the percentiles of the simulated hausman1 are close to the theoretical percentiles of the $\chi^2(1)$ distribution. In particular, we note that the 95th percentile of hausman1 is 3.70, reasonably close to the theoretical percentile of 3.84, while the 99th percentile of hausman1 (6.18) is not far from the theoretical value (6.63). These comparisons confirm that the test has close to optimal size, meaning that the actual probability of rejecting the null when the null is true is close to 0.05 when the nominal test size is 0.05, and close to 0.01 when the nominal test size is 0.01.

Another very useful way of comparing the empirical distribution of a test statistic to its null distribution is to use a quantile-quantile (qq) plot. This is a plot of the quantiles of the empirical distribution against the corresponding quantiles of the theoretical distribution. To obtain this plot, we do as follows:

```
qchi hausman1, df(1) msize(1)
```

The result is shown in Figure 9.4. The closeness of the qq-plot to the 45 degree line confirms that the empirical distribution of the test statistic is close to the $\chi^2(1)$ null distribution.

To investigate the power performance of the test, we consider the distribution of hausman2, that is, the test statistic simulated under the assumption that 30% of subjects do not truthfully report strength-of-preference. A histogram is shown in Figure 9.5. It shows that the majority of the distribution is to the right of the test's critical values, and therefore that the probability of rejecting the null-hypothesis (i.e. the power of the test) is high. More precisely, we find that 71.4% of the realisations of the test statistic are greater than the critical value of 3.84,

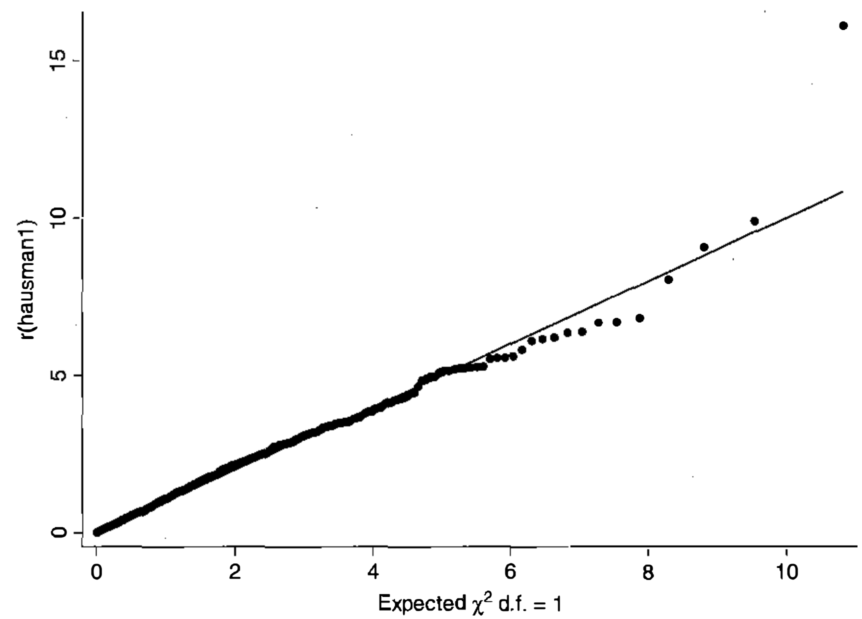



Figure 9.4: qq plot of hausman test statistic under null hypothesis

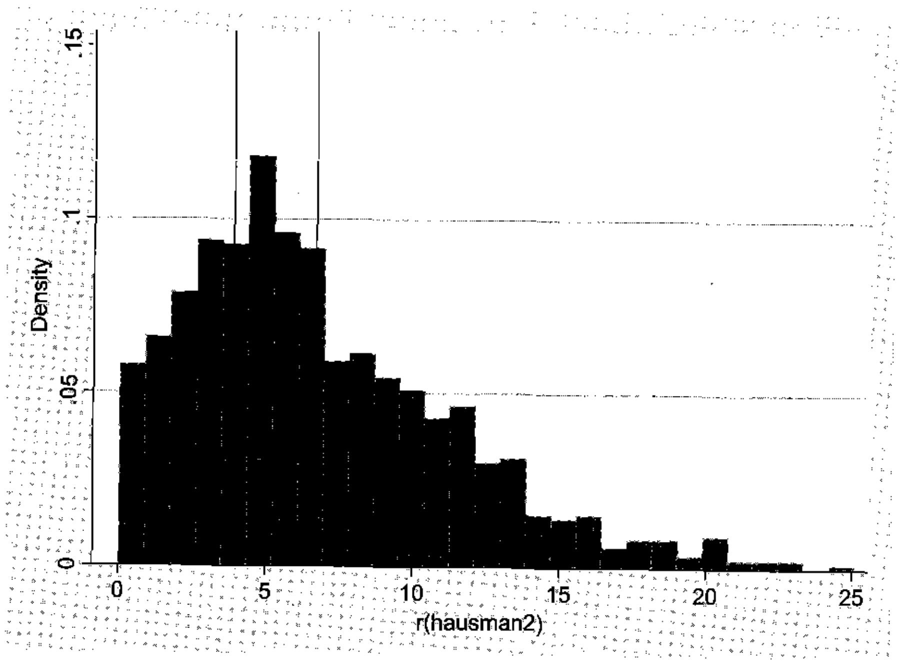



Figure 9.5: Histogram of Hausman test statistic under alternative hypothesis
Note: Vertical lines drawn at $\chi^2(1, 0.05)$ and $\chi^2(1, 0.01)$.

indicating that the power of the test (when testing against the specified alternative) is 0.714.

Of course, if the proportion of subjects choosing randomly ($p$) were set to a higher value than 0.30, we would expect to find a higher power. By considering a grid of values of $p$, we could consider how the probability of rejecting the null hypothesis depends on $p$. To do this would be to trace out the "power function" of the test (see Exercise 2).

## 9.5 Summary and Further Reading

There have been two broad themes to this chapter: simulation (meaning creation of a single data set from an assumed model, and analysis of that data set); and Monte Carlo (meaning repeated simulations from an assumed model, with analysis performed for each simulation, and finally investigation of the distributional properties of estimates and tests).

There is not much more to say about the simulation part. The models simulated were relatively straightforward, with the possible exception of the dynamic panel model simulated in Section 9.3. Readers interested in the estimation of dynamic panel models should read Roodman (2009).



Regarding Monte Carlo, it has been demonstrated that this is a very useful technique for investigating (or just verifying) the distributional properties of an econometric estimator or test. Another point is that it is potentially a very useful learning technique. Kennedy (1998) strongly recommends the use of Monte Carlo analysis in teaching as a complement to the teaching of econometric theory, stressing that it is particularly useful in understanding the concept of the sampling distribution.

## Exercises

1. In Section 9.3.4, a Monte Carlo study was conducted in which the impact of unobserved heterogeneity on the estimation of the binary probit model was assessed. Conduct a similar exercise for the two-limit Tobit model, of the type used in Sections 6.6 and 8.5 for the modelling of contributions in public goods experiments. How serious are the consequences of neglecting heterogeneity in this setting?
2. In Section 9.4, a Monte Carlo study was conducted to investigate the performance of the Hausman test when testing the null hypothesis that subjects truthfully report strength of preference. This null hypothesis was expressed as $p = 0$ where $p$ is the proportion of subjects who do not report truthfully. The alternative hypothesis was assumed to be $p = 0.3$, and the power of the test was found to be 0.714.

   a. By changing the number at the end of the command `replace y_poor=y_bad if uniform() < 0.3`, carry out the Monte Carlo separately for a grid of values of $p$ ranging from 0 to 1. For each run, store the proportion of values of hausman2 that exceed 3.84 (i.e. store the power of the test, for each value of $p$ assumed). Plot the power of test against $p$. Note that this is a plot of the power function of the test.
   b. The sample size assumed in the Monte Carlo is 1,050. By altering the number in the `obs` option in the `simulate` command, obtain the power function for a lower sample size (e.g. 100) and for a higher sample size (e.g. 5,000). Draw all three power functions on the same plot. What happens to the power function as the sample size increases?

# Chapter 10
# Introduction to the Method of Maximum Simulated Likelihood (MSL)

## 10.1 Introduction

The importance of between-subject heterogeneity in the analysis of experimental data has been established in previous chapters. In particular, in Chapters 4 and 5, we introduced between-subject heterogeneity in the context of linear models, by means of fixed-effects, random-effects, and multi-level models. In this chapter, these panel data techniques are extended to discrete data models. Key to this objective is the introduction of an estimation procedure that is used many times later in the book in the presence of between-subject heterogeneity. This procedure is the method of maximum simulated likelihood (MSL). This is a method which uses simulation in order to evaluate the integral over the heterogeneity term that appears in the likelihood function. The method will be introduced by applying it to two standard models: random effects probit which is used for binary panel data; and random effects Tobit which is used for censored panel data. One reason for using these examples is that these are models that can be estimated using built-in STATA commands (`xtprobit` and `xttobit` respectively), and it is useful to compare the results from our MSL routine with the results obtained using these commands, simply in order to verify that our MSL routine has been programmed correctly and is doing what we intend it to do.

By focusing on these relatively simple examples, this chapter will serve the important purpose of providing a detailed explanation of every step of the MSL procedure. This style of explanation is intended to provide the necessary preparation for an understanding of the more complicated applications of the routine that are encountered later in the book.

The next time the MSL technique will be encountered is Chapter 11, where we develop the panel hurdle model. Later, in Chapter 13, we shall use MSL to estimate risky choice models that incorporate continuous between-subject variation in risk attitude. Finally, in Section 15.3 we use MSL to estimate a model of altruism that allows between-subject heterogeneity in parameters of the utility function.

## 10.2 The Principle of Maximum Simulated Likelihood (MSL)

When modelling data consisting of multiple observations per subject, we tend to encounter log-likelihood functions with the following structure:

$$LogL = \sum_{i=1}^{n} \ln \int_u \left[ \prod_{t=1}^{T} g(y_{it}|x_{it}; u) \right] f(u) du \tag{10.1}$$

A log-likelihood of the form (10.1) would arise in a situation in which there are $n$ subjects, each of whom has engaged in $T$ tasks. $y_{it}$ is the decision variable for subject $i$ in task $t$. This decision is assumed to depend on a set of explanatory variables contained in $x_{it}$, and also on the univariate random variable $u$. The exact nature of this dependence is embodied in the conditional density function $g(.|\,.\,;\,.)$. Note that $u$ varies between subjects, but there is only one value of $u$ for each subject. $u$ is distributed over subjects according to the density function $f(u)$, and is known as the subject-specific random effect. It is the variation in $u$ which captures between-subject heterogeneity.

The main practical problem concerning the evaluation of (10.1) is the integral over the random effect term, $u$. There are two possible approaches to the evaluation of this integral: quadrature and simulation. Here, and throughout the text, we will use simulation. The estimation method using simulation is known as the method of maximum simulated likelihood (MSL, see Train, 2003).

The basic principle of MSL is explained as follows. Consider the following integral over the random variate $\epsilon$:

$$I = \int_{-\infty}^{\infty} t(\epsilon) f(\epsilon) d\epsilon \tag{10.2}$$

where $f(\epsilon)$ is the density function of $\epsilon$, and $t(\epsilon)$ is some other function of $\epsilon$. The quantity defined in (10.2) is, of course, $E[t(\epsilon)]$, so we could evaluate it by finding the "mean" of $t(\epsilon)$ over a number of values of $\epsilon$. That is:

$$\hat{I} = \frac{1}{R} \sum_{r=1}^{R} t(\epsilon_r) \tag{10.3}$$

If the values $\epsilon_1, \ldots, \epsilon_R$ are random numbers drawn from the density $f(\epsilon)$, and if the number of draws $R$ is sufficiently high, (10.3) will be an accurate approximation to the integral (10.2).

This is the procedure that is applied in order to evaluate the integral appearing in (10.1), except that the problem is made more complicated by the panel structure. Panel data is arranged in $n$ blocks, one block for each subject. Assuming a balanced panel, each block contains the same number $T$ of rows for each subject. The total number of rows in the data set is therefore $n \times T$. Only one set of random draws is required for each block of the data. Therefore a set of $R$ draws is appended (horizontally) to each row of the data set, but within a given block, these $T$ rows of draws are all the same. Consider a particular block. For each of the $R$ draws (i.e. for each column of the array of draws), the product (over $t$) appearing within the square brackets in (10.1) is found, and the resulting $R$ products are averaged, in the spirit of (10.3), in order to obtain an approximation to the integral appearing within the right-hand side of (10.1). An integral is obtained thus for each of the $n$ blocks in the data set. These $n$ integrals are finally logged and summed to give the sample log-likelihood (10.1).



When maximising log-likelihood functions from panel data in STATA, there is an important difference from the situation of independent observations encountered in Chapter 6. In the examples considered in Chapter 6, one contribution to the log-likelihood was generated for each row of the data set, and accordingly, "method lf" was used (lf standing for linear-form restrictions). With panel data, the linear-form restrictions are not met, since the log-likelihood is formed for groups of observations. For this reason, "method d0" is used instead of "method lf". "Method d0" has already been used for the analysis of public goods panel data in Section 8.5, although in that example, simulation was not required because there was no integral in the likelihood function.

## 10.3 Halton Draws

### 10.3.1 The case for using Halton draws

For the present purpose, there are better ways to draw numbers from a density $f(.)$ than simply to use a random number generator. As explained by Train (2003, ch. 9) there are two issues: coverage and covariance.

The coverage issue is as follows. With independent random draws, it is possible that the draws are clumped together in some areas, with a scarcity of draws in other areas. Procedures that guarantee an even coverage can be expected to provide a better approximation to the integral.

The covariance issue is as follows. With independent draws, the covariance between draws is zero. The variance of a simulator such as (10.3) will therefore be the variance based on a single draw divided by $R$. If consecutive draws were negatively correlated instead of independent, then the variance of the simulator would be lower, that is, it would be a more accurate approximation to the true integral. For example, consider $R = 2$. $\hat{I} = (t(\epsilon_1) + t(\epsilon_2))/2$, and its variance is $V(\hat{I}) = (V[t(\epsilon_1)] + V[t(\epsilon_2)] + 2\text{Cov}[t(\epsilon_1), t(\epsilon_2)])/4$. If the draws are independent, then the variance is $V(\hat{I}) = V[t(\epsilon_r)]/2$. If the two draws are negatively correlated with each other, the covariance term becomes negative, and the variance becomes less than $V[t(\epsilon_r)]/2$. More intuitively, when the draws are negatively correlated, a value of $t(\epsilon_r)$ that is above the true mean $I$, will tend to be followed by a value $t(\epsilon_{r+1})$ that is below $I$, such that their average will be close to $I$.

Halton sequences (Halton, 1960) provide both coverage and negative correlation. The sequence is defined in terms of a given prime number, *p*, which may be looked on as the "seed" for the sequence. For example, consider $p = 2$. The sequence is:

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, \frac{5}{16}, \frac{13}{16}, \frac{3}{16}, \frac{11}{16}, \frac{7}{16}, \frac{15}{16}, \frac{1}{32}, \ldots$$

For $p = 3$, the sequence is:

$$\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}, \frac{19}{27}, \frac{4}{27}, \frac{13}{27}, \frac{22}{27}, \frac{7}{27}, \frac{16}{27}, \frac{25}{27}, \frac{2}{27}, \frac{11}{27}, \frac{20}{27}, \frac{5}{27}, \frac{14}{27}, \frac{23}{27}, \frac{8}{27}, \frac{17}{27}, \frac{26}{27}, \frac{1}{81}, \ldots$$

In Section 10.3.2 it will be demonstrated how these sequences can be generated in STATA.

### *10.3.2 Generating Halton draws in STATA*

There are (at least) two ways of generating Halton sequences in STATA. The first is to use the `mdraws` command (Cappellari & Jenkins, 2006). This is a user-written routine that needs to be installed. Having installed `mdraws`, the following code will generate the first 1000 values of the above two sequences in the variables h1_1 and h2_1. The first 31 values of each are shown in the first two columns of Figure 10.1.

```
clear
set obs 1000
matrix p=(2,3)
mdraws, neq(2) dr(1) prefix(h) primes(p)
```

Four options have been used with the `mdraws` command: `neq(2)` specifies the number of different prime numbers being used to generate sequences; `dr(1)` specifies the number of columns over which to spread each of the sequences (this becomes useful later when we need to generate large numbers of columns of draws); `prefix(h)` specifies a prefix for the name of each variable; `primes(p)` specifies the prime numbers ("seeds") which have previously been specified in the vector p.

The second method for generating Halton draws is using the `halton` function which is available in mata. The commands required to generate exactly the same two sequences as obtained above are:

```
mata h=halton(100,2,1)
getmata (h*)=h
```

As a result of this code, the two sequences will be named h1 and h2.



| | h1_1 | h2_1 | z1 | z2 |
|---|---|---|---|---|
| 1 | .5 | .33333333 | 0 | -.4307273 |
| 2 | .25 | .66666667 | -.6744897 | .4307273 |
| 3 | .75 | .11111111 | .6744897 | -1.22064 |
| 4 | .125 | .44444444 | -1.150349 | -.1397103 |
| 5 | .625 | .77777778 | .3186394 | .7647097 |
| 6 | .375 | .22222222 | -.3186394 | -.7647097 |
| 7 | .875 | .55555556 | 1.150349 | .1397103 |
| 8 | .0625 | .88888889 | -1.534121 | 1.22064 |
| 9 | .5625 | .03703704 | .1573107 | -1.786156 |
| 10 | .3125 | .37037037 | -.4887764 | -.3308726 |
| 11 | .8125 | .7037037 | .8871465 | .5350828 |
| 12 | .1875 | .14814815 | -.8871465 | -1.044409 |
| 13 | .6875 | .48148148 | .4887764 | -.0464357 |
| 14 | .4375 | .81481481 | -.1573107 | .8957798 |
| 15 | .9375 | .25925926 | 1.534121 | -.6456308 |
| 16 | .03125 | .59259259 | -1.862732 | .2342192 |
| 17 | .53125 | .92592593 | .0784124 | 1.446104 |
| 18 | .28125 | .07407407 | -.5791321 | -1.446104 |
| 19 | .78125 | .40740741 | .7764218 | -.2342192 |
| 20 | .15625 | .74074074 | -1.00999 | .6456308 |
| 21 | .65625 | .18518519 | .4022501 | -.8957798 |
| 22 | .40625 | .51851852 | -.2372021 | .0464357 |
| 23 | .90625 | .85185185 | 1.318011 | 1.044409 |
| 24 | .09375 | .2962963 | -1.318011 | -.5350828 |
| 25 | .59375 | .62962963 | .2372021 | .3308726 |
| 26 | .34375 | .96296296 | -.4022501 | 1.786156 |
| 27 | .84375 | .01234568 | 1.00999 | -2.246197 |
| 28 | .21875 | .34567901 | -.7764218 | -.3970128 |
| 29 | .71875 | .67901235 | .5791321 | .4649388 |
| 30 | .46875 | .12345679 | -.0784124 | -1.157879 |
| 31 | .96875 | .45679012 | 1.862732 | -.1085237 |

Figure 10.1: Columns 1 and 2 contain Halton sequences with $p = 2$ and $p = 3$; columns 3 and 4 show same sequences transformed to normality

Whenever Halton draws are used in this book, they are generated using `mdraws`. This is because the user-written `mdraws` command offers more flexibility than the built-in `halton` function.

In Figure 10.2 (right-hand pane) the two Halton sequences generated above are plotted against each other. For comparison, in the left-hand pane we show a scatter plot of two sequences of random draws from a uniform distribution. Notice the superiority in coverage of the Halton draws over the random draws.

The draws have been generated from a uniform(0, 1) distribution. The distributions that we usually encounter are related to the normal distribution. We therefore need to transform the draws from uniform draws to normal draws. The transformation that is required is the inverse normal cdf:

$$z = \Phi^{-1}(u) \tag{10.4}$$

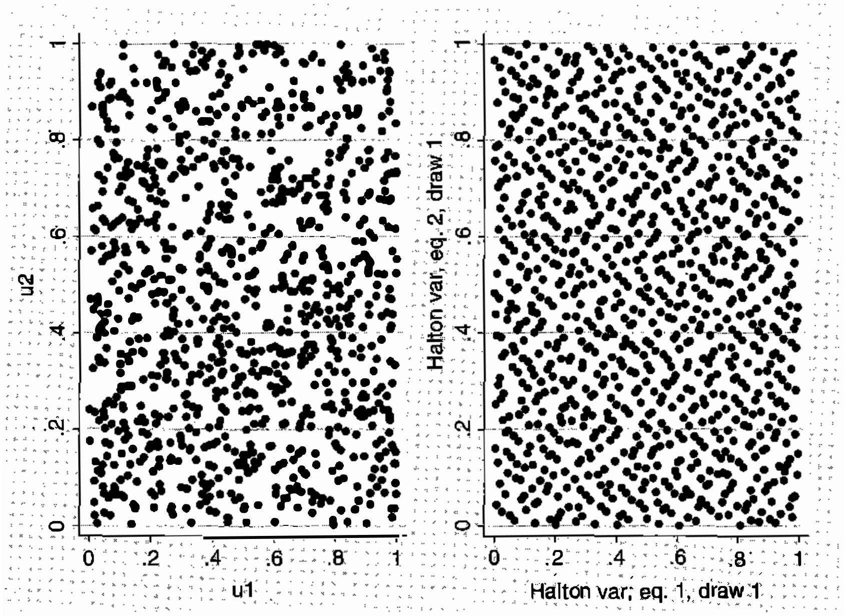



Figure 10.2: 1000 random draws from two uniforms (left pane); 1000 draws from two Halton sequences (right pane)

In STATA the function $\Phi^{-1}(.)$ is named `invnorm`. The third and fourth columns of Figure 10.1 contain the two simulated standard normals obtained from the two Halton sequences using this transformation.

Figure 10.3 (right pane) shows the histogram of one of the transformed sets of draws obtained using (10.4). Again for comparison, we show in the left pane a histogram of random numbers from a standard normal. Notice that the transformed Halton draws are much closer than the random draws to the true probability density function of a standard normal.

Having applied the transformation (10.4), we have standard normal draws. In order to create draws from a $N(\mu, \sigma^2)$, we would apply the further transformation:

```
gen x = mu + sig * z
```

And for a lognormal($\mu$, $\sigma^2$) we would apply the still further transformation:

```
gen y = exp(x)
```

While the parameters of the model (e.g. $\mu$ and $\sigma$ in the above examples) will be constantly changing throughout the likelihood maximisation procedure, the Halton draws are fixed. The draws are made at the start of the program, and the same draws are then used repeatedly throughout the procedure.



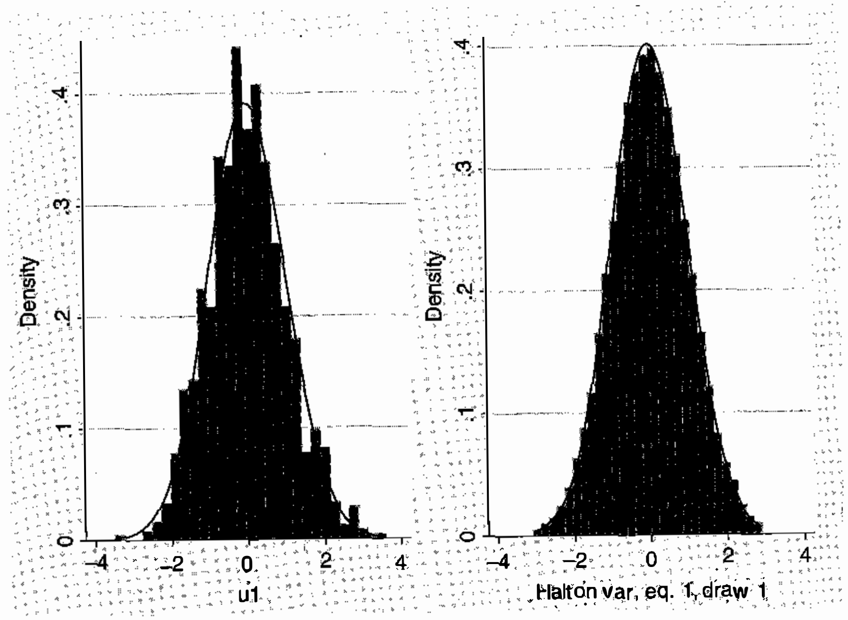



Figure 10.3: 1,000 realisations of standard normal obtained from random draws (left pane); 1,000 realisations of standard normal obtained from Halton draws (right pane)

### 10.3.3 Halton draws for panel data estimation

As mentioned in Section 10.2, use of Halton draws in panel models is complicated in view of the panel structure. Panel data (balanced) is arranged in blocks of $T$ rows, and only one set of random draws is required for each block. Therefore a set of $R$ draws is appended (horizontally) to each row of the data set, but within a given block, the $T$ rows of draws are all the same.

In order to generate Halton draws with this structure, we use the following procedure in STATA. We will assume that the subject is indexed by $i$ and the task is indexed by $t$, and these variables are already present in memory. It is first useful to generate dummy variables "first" and "last" indicating whether an observation is (respectively) the first or last for a given subject. Then, using `mdraws` we generate the required number of draws for each subject, but these are only placed in the *first* row of each block; missing values will appear in all other rows. We then perform a loop over the columns containing Halton draws. Within the loop, three tasks are performed: firstly, the `recast` command is applied to ensure that the variable is stored in double precision;[1] secondly, the missing values are filled in by copying the values appearing in the first row; thirdly, each Halton draw is transformed from a uniform distribution to a standard normal distribution using the `invnorm(.)` function.

[1] See Section 10.4.4 for an explanation of "double precision".

```
by i: generate int first=1 if _n==1
by i: generate int last=1 if _n==_N

mat p=[3]
mdraws if first==1, neq(1) dr(3) prefix(h) primes(p)

scalar draws=r(n_draws)

local hlist h1*

quietly{

foreach v of varlist `hlist' {
recast double `v'
by i: replace `v'=`v'[1] if `v'==.
replace `v'=invnorm(`v')
}
}
```

Having implemented the above procedure (with $draws = 3$) on a data set with $n = 10$ and $T = 3$, the data set appears as in Figure 10.4. Notice that the (transformed) draws are a mixture of positive and negative numbers, centred on zero, which is exactly what we expect of a standard normal variate. The most important feature of this structure of draws is that, for any given subject ($i$), the set of three draws is the same in each of the three rows ($t = 1, 2, 3$) for that subject.

## 10.4 The Random Effects Probit Model

### 10.4.1 Model

Consider the random effects probit model:

$$
\begin{aligned}
y_{it}^* &= \beta_1 + \beta_2 x_{it} + u_i + \epsilon_{it} \qquad i = 1, \ldots, n \quad t = 1, \ldots, T \\
&= x_{it}'\beta + u_i + \epsilon_{it} \\
u_i &\sim N(0, \sigma_u^2) \\
\epsilon_{it} &\sim N(0, 1)
\end{aligned}
\tag{10.5}
$$

where $y_{it}^*$ is a latent variable. This model differs from the simple probit model discussed extensively in Chapter 6 through the introduction of the term $u_i$, the subject-specific random effect. This is a key term because it captures between-subject differences, and its variance $\sigma_u^2$ may be interpreted as a measure of between-subject heterogeneity. In the second line of (10.5) we are adopting the convention of writing $x_{it}$ as the vector of explanatory variables, the first of which is a constant, and $\beta$ as the corresponding vector of parameters, the first of which is an intercept.

The relationship between the latent variable $y_{it}^*$ and the observed (binary) variable $y_{it}$ is:

$$
y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0 \\ 0 & \text{if } y_{it}^* \leq 0 \end{cases} \tag{10.6}
$$



| | i | t | first | last | h1_1 | h1_2 | h1_3 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | . | -.4307273 | .4307273 | -1.2206403 |
| 2 | 1 | 2 | . | . | -.4307273 | .4307273 | -1.2206403 |
| 3 | 1 | 3 | . | 1 | -.4307273 | .4307273 | -1.2206403 |
| 4 | 2 | 1 | 1 | . | -.1397103 | .76470967 | -.76470967 |
| 5 | 2 | 2 | . | . | -.1397103 | .76470967 | -.76470967 |
| 6 | 2 | 3 | . | 1 | -.1397103 | .76470967 | -.76470967 |
| 7 | 3 | 1 | 1 | . | .1397103 | 1.2206403 | -1.7861556 |
| 8 | 3 | 2 | . | . | .1397103 | 1.2206403 | -1.7861556 |
| 9 | 3 | 3 | . | 1 | .1397103 | 1.2206403 | -1.7861556 |
| 10 | 4 | 1 | 1 | . | -.33087257 | .53508282 | -1.0444088 |
| 11 | 4 | 2 | . | . | -.33087257 | .53508282 | -1.0444088 |
| 12 | 4 | 3 | . | 1 | -.33087257 | .53508282 | -1.0444088 |
| 13 | 5 | 1 | 1 | . | -.04643572 | .89577982 | -.64563075 |
| 14 | 5 | 2 | . | . | -.04643572 | .89577982 | -.64563075 |
| 15 | 5 | 3 | . | 1 | -.04643572 | .89577982 | -.64563075 |
| 16 | 6 | 1 | 1 | . | .23421919 | 1.4461036 | -1.4461036 |
| 17 | 6 | 2 | . | . | .23421919 | 1.4461036 | -1.4461036 |
| 18 | 6 | 3 | . | 1 | .23421919 | 1.4461036 | -1.4461036 |
| 19 | 7 | 1 | 1 | . | -.23421919 | .64563075 | -.89577982 |
| 20 | 7 | 2 | . | . | -.23421919 | .64563075 | -.89577982 |
| 21 | 7 | 3 | . | 1 | -.23421919 | .64563075 | -.89577982 |
| 22 | 8 | 1 | 1 | . | .04643572 | 1.0444088 | -.53508282 |
| 23 | 8 | 2 | . | . | .04643572 | 1.0444088 | -.53508282 |
| 24 | 8 | 3 | . | 1 | .04643572 | 1.0444088 | -.53508282 |
| 25 | 9 | 1 | 1 | . | .33087257 | 1.7861556 | -2.2461975 |
| 26 | 9 | 2 | . | . | .33087257 | 1.7861556 | -2.2461975 |
| 27 | 9 | 3 | . | 1 | .33087257 | 1.7861556 | -2.2461975 |
| 28 | 10 | 1 | 1 | . | -.3970128 | .46493877 | -1.1578786 |
| 29 | 10 | 2 | . | . | -.3970128 | .46493877 | -1.1578786 |
| 30 | 10 | 3 | . | 1 | -.3970128 | .46493877 | -1.1578786 |

Figure 10.4: An array of Halton draws (transformed to normality) suitable for use with panel data with $n = 10$; $T = 3$; $R = 3$ draws

From (10.5) and (10.6) we may obtain the probabilities of the two binary outcomes, conditional on the value of $u_i$:

$$
P(y_{it} = 1|u_i) = P(y_{it}^* > 0|u_i) = P\left(\epsilon_{it} > -(x_{it}'\beta + u_i)|u_i\right) = \Phi(x_{it}'\beta + u_i) \tag{10.7}
$$

$$
P(y_{it} = 0|u_i) = 1 - P(y_{it} = 1|u_i) = 1 - \Phi(x_{it}'\beta + u_i) = \Phi\left(-(x_{it}'\beta + u_i)\right) \tag{10.8}
$$

Combining (10.7) and (10.8), we may obtain the likelihood contribution for subject $i$ conditional on the value of $u_i$:

$$
L_i|u_i = \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(x_{it}'\beta + u_i\right)\right] \tag{10.9}
$$

Notice that in (10.9) we are continuing with the convention introduced in Chapter 6 of defining the binary variable $yy$ to be $+1$ when $y=1$, and $-1$ when $y=0$. We obtain the marginal likelihood from the conditional likelihood by integrating (10.9) over the normally distributed variate $u$:

$$L_i = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(x_{it}'\beta + u\right)\right] \frac{1}{\sigma_u}\phi\left(\frac{u}{\sigma_u}\right) \mathrm{d}u \qquad (10.10)$$

Finally we take the log of (10.10) and sum over subjects to obtain the sample log-likelihood function for model (10.5):

$$LogL = \sum_{i=1}^{n} \ln\left[\int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(x_{it}'\beta + u\right)\right] \frac{1}{\sigma_u}\phi\left(\frac{u}{\sigma_u}\right) \mathrm{d}u\right] \qquad (10.11)$$

Having estimated the parameters of Model (10.5), we will obtain *posterior random effects* for each subject. These are given by:

$$\hat{u}_i = E\left(u_i | y_{i1} \ldots y_{iT}\right) = \frac{\int_{-\infty}^{\infty} u \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(x_{it}'\hat{\beta} + u\right)\right] \frac{1}{\hat{\sigma}_u}\phi\left(\frac{u}{\hat{\sigma}_u}\right) \mathrm{d}u}{\int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(x_{it}'\hat{\beta} + u\right)\right] \frac{1}{\hat{\sigma}_u}\phi\left(\frac{u}{\hat{\sigma}_u}\right) \mathrm{d}u} \qquad (10.12)$$

where hats indicate parameter estimates.

The easiest way to estimate Model (10.5) is using the following STATA commands:

```
xtset i t
xtprobit y x, re
```

However, we also wish to estimate Model (10.5) using MSL. In Section 10.4.2, we shall outline our method for simulating data from Model (10.5). Then in Section 10.4.3, we shall explain how to apply MSL to estimate the model on the simulated data. The STATA command `xtprobit` does not appear to have a post-estimation routine for computing the posterior random effects (10.12). The MSL estimation routine that we develop will include this feature.

## *10.4.2 Simulation*

The data on which we will demonstrate estimation of the random effects probit model by MSL will be simulated data. In this section we outline the simulation.

The following code simulates data from model (10.5) with $n=60$ subjects, and $T=20$ tasks. The parameter values used in the simulation are:

$$\beta_1 = -1.0; \quad \beta_2 = 2.0; \quad \sigma_u = 0.5$$



```
clear

* SET SAMPLE SIZE AND RANDOM NUMBER SEED

set obs 1200
set seed 7654321

*  GENERATE SUBJECT IDENTIFIER (i) AND TASK IDENTIFIER (t);
* ENSURE THAT THESE ARE STORED AS INTEGERS
egen i=seq(), f(1) b(20)
egen t=seq(), f(1) t(20)

recast int i t

*  DECLARE DATA TO BE PANEL DATA

xtset i t

* GENERATE x (FROM UNIFORM) AND e (FROM NORMAL)

gen double x=runiform()
gen double e=rnormal()

* GENERATE u (SUBJECT-SPECIFIC EFFECT)

by i: generate double u=0.5*(invnorm(uniform())) if _n==1
by i: replace u=u[1] if u==.

* GENERATE LATENT VARIABLE y*, AND BINARY VARIABLE y

gen double ystar=-1.0+2.0*x+u+e
gen int y=ystar>0

* ESTIMATE RANDOM EFFECTS PROBIT MODEL USING xtprobit COMMAND

xtprobit y x
```

Figure 10.5 shows the first 31 rows of the simulated data (with the unobserved variables removed). The data is said to be in "long form", with each row representing only one of the decisions made by a particular subject.[2] Since each subject makes $T=20$ decisions, only the decisions of the first two subjects appear in this screenshot. The final line of the above code applies the `xtprobit` command to this data set. The results are as follows:

```
. xtprobit y x, re

Random-effects probit regression                Number of obs      =      1200
Group variable: i                               Number of groups   =        60

Random effects u_i ~ Gaussian                   Obs per group: min =        20
                                                               avg =      20.0
                                                               max =        20

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(1)       =    157.25
Log likelihood  = -710.41547                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

[2] An alternative means of presenting a panel data set is "wide form", in which there is only one row for each subject, and all decisions made by a given subject are coded within that row.

```
-------------+----------------------------------------------------------------
           x |   1.847581    .147337    12.54   0.000     1.558806    2.136356
       _cons |  -.9181971   .1090388    -8.42   0.000    -1.131909   -.704485
-------------+----------------------------------------------------------------
    /lnsig2u |  -1.178868    .253014                     -1.674766   -.6829695
-------------+----------------------------------------------------------------
     sigma_u |   .5546412    .070166                      .4328418    .7107143
         rho |   .2352558   .0455199                      .1577898    .3355989
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =    95.05 Prob >= chibar2 = 0.000
```

We see that the estimates are all close to the true values, with all three of the true values inside their respective confidence intervals. The purpose of the next few subsections is to obtain the same set of results using a self-written MSL program.

### 10.4.3 Estimation by MSL

The log-likelihood function for the random effects probit model was presented in (10.11) above, and is given by:

| | i | t | x | y |
|---|---|---|---|---|
| 1 | 1 | 1 | .6588475 | 1 |
| 2 | 1 | 2 | .98049626 | 1 |
| 3 | 1 | 3 | .0373436 | 0 |
| 4 | 1 | 4 | .10857736 | 0 |
| 5 | 1 | 5 | .90152806 | 1 |
| 6 | 1 | 6 | .35898743 | 1 |
| 7 | 1 | 7 | .91839248 | 0 |
| 8 | 1 | 8 | .1706897 | 0 |
| 9 | 1 | 9 | .18685892 | 1 |
| 10 | 1 | 10 | .90669361 | 1 |
| 11 | 1 | 11 | .58999881 | 0 |
| 12 | 1 | 12 | .93117496 | 1 |
| 13 | 1 | 13 | .41433274 | 1 |
| 14 | 1 | 14 | .74338348 | 1 |
| 15 | 1 | 15 | .13376886 | 1 |
| 16 | 1 | 16 | .648391 | 1 |
| 17 | 1 | 17 | .47783562 | 1 |
| 18 | 1 | 18 | .01708559 | 1 |
| 19 | 1 | 19 | .55203962 | 1 |
| 20 | 1 | 20 | .24069986 | 1 |
| 21 | 2 | 1 | .58143611 | 1 |
| 22 | 2 | 2 | .81475681 | 0 |
| 23 | 2 | 3 | .91146142 | 1 |
| 24 | 2 | 4 | .94796911 | 1 |
| 25 | 2 | 5 | .97362812 | 1 |
| 26 | 2 | 6 | .72510801 | 1 |
| 27 | 2 | 7 | .95791281 | 1 |
| 28 | 2 | 8 | .24174224 | 0 |
| 29 | 2 | 9 | .18369073 | 0 |
| 30 | 2 | 10 | .72085298 | 1 |
| 31 | 2 | 11 | .13015481 | 0 |

Figure 10.5: First 31 rows of simulated binary panel data set



$$LogL = \sum_{i=1}^{n} \ln\left[\int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(x_{it}'\beta + u\right)\right] \frac{1}{\sigma_u}\phi\left(\frac{u}{\sigma_u}\right) \mathrm{d}u\right] \tag{10.13}$$

This function contains an integral over the variable $u$, which has a $N(0, \sigma_u^2)$ distribution. As explained in a more general context in Section 10.2, the basic principle of MSL is that this integral may be approximated using the following average:

$$\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T}\Phi\left[yy_{it} \times \left(x_{it}'\beta + u_{r,i}\right)\right]\right) \tag{10.14}$$

where $u_{1,i} \ldots u_{R,i}$ are the transformed Halton draws for subject $i$ (remember that these are fixed over $t$; see Section 10.3.3). The draws are transformed so as to represent draws from a $N(0, \sigma_u^2)$ distribution.

We need to maximise the log-likelihood function (10.13) with respect to the three parameters, $\beta_1$, $\beta_2$, and $\sigma_u$. The MSL approach amounts to replacing the integral appearing in (10.13) with the average defined in (10.14), and then maximising the resulting function.

The simulated data set was described in Section 10.4.2. Each row represents a particular decision for a particular subject. The row contains the subject identifier ($i$), the task identifier ($t$), the value of the explanatory variable pertaining to the decision ($x_{it}$), and the binary decision itself ($y_{it}$, taking values of 0 or 1).

As in previous chapters, the `ml` routine in STATA will be used for maximisation of the log-likelihood (10.13). There are a number of different likelihood evaluators in STATA. The one that was used in the various examples in Chapter 6 was *lf* (linear form). As mentioned in Section 10.1, a feature of the log-likelihood function (10.13) is that it does not satisfy the linear form restrictions, and therefore *lf* cannot be used. This is because the likelihood contributions that need to be summed in order to obtain the sample log-likelihood are not each derived from the information in a single row of the data, but are instead derived from the entire block of rows corresponding to a given subject. There is only one likelihood contribution for each such block of rows. Because of this, the *d-family* evaluators are required in place of the *lf* evaluator. The simplest of these is the *d0* evaluator, which simply requires the log-likelihood contributions to be evaluated. This is the one that we will use. The *d1* and *d2* evaluators require analytical derivatives of the log-likelihood to be programmed as well as the function evaluation, while *d0* uses numeric derivatives. Convergence is expected to be somewhat faster if *d1* or *d2* are used instead of *d0*, but this would be at the cost of a prohibitive amount of analytic differentiation.

### 10.4.4 Preparation of data for estimation

Having read the data into STATA, a number of tasks need to be performed in preparation for estimation.

First of all, it is desirable, if not essential, for all variables to be stored in the correct storage type. Integer variables, such as $i$, $t$, and $y$, should be stored as integers, since to store them as anything else would be a waste of memory. To ensure that they are stored as integers, we use the command:

```
recast int i t y
```

Any (numeric) variable that is not an integer must be stored as a *double*. Variables stored as "double" are accurate to 16 digits, a level of accuracy known as "double precision". If they are not stored as doubles, they are likely to be stored as floats, which only have around seven digits of accuracy. The reason why this is important is that when variables appear as floats, the log-likelihood function might not be computed with sufficient accuracy, and may not have the property of smoothness that is essential in order for the maximum to be found. The command for ensuring that all variables other than integer variables (i.e. the variable $x$ in this example) are stored in double precision is:

```
recast double x
```

Next, we require two binary variables indicating whether the current observation is the first, and the last, observation for each subject. The commands are:

```
by i: gen int  first=1 if _n==1
by i: gen int last=1 if _n==_N
```

The variable "first" will take the value 1 in the first row of data for each subject (i.e. when $t = 1$), and missing in every other row. Likewise, the variable "last" will take the value 1 in the last row for each subject (i.e. when $t = T$).

The Halton draws are generated, transformed, and arranged into the panel structure using the procedure described in detail in Section 10.3.3, with the difference that here, the dimension of the panel is $n = 10$; $T = 20$. Also, $R = 125$ draws will be used in estimation.

Finally, although the choice made by the subject in each task is represented by the 0/1 variable $y$, for practical reasons explained in Chapter 6, this information also needs to be supplied in the form of a variable (yy) taking the values +1 and $-1$. This variable is created with the following command:

```
gen int yy=2*y-1
```

### 10.4.5 The likelihood evaluator

The likelihood evaluation program will now be outlined. In the STATA code, we adopt the following names for the parameters and the other components of the log-likelihood function (10.13).



| Component of LogL | STATA name |
|---|---|
| $\beta_1 + \beta_2 x$ | xb |
| $\ln(\sigma_u)$ | ln_s_u |
| $\sigma_u$ | s_u |
| $\Phi\left[yy_{it} \times (\beta_1 + \beta_2 x_{it} + u_{r,i})\right]$ | p |
| $L_i \vert u = \prod_{t=1}^{T} \Phi\left[yy_{it} \times (\beta_1 + \beta_2 x_{it} + u_{r,i})\right]$ | pp |
| $u_{r,i} \times \prod_{t=1}^{T} \Phi\left[yy_{it} \times (\beta_1 + \beta_2 x_{it} + u_{r,i})\right]$ | upp |
| $L_i = \frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T} \Phi\left[yy_{it} \times (\beta_1 + \beta_2 x_{it} + u_{r,i})\right]\right)$ | ppp |
| $\frac{1}{R}\sum_{r=1}^{R}\left(u_{r,i} \times \prod_{t=1}^{T} \Phi\left[yy_{it} \times (\beta_1 + \beta_2 x_{it} + u_{r,i})\right]\right)$ | uppp |
| $R$ | draws |
| $\ln L_i$ | lnppp |

We will now consider each section of the program in turn (a complete uninterrupted listing of the code is provided in the next sub-section). The first line of the program assigns it the name "`my_rep`". The next line (`args`) introduces the arguments of the function: `todo` is simply a scalar whose value is determined by which of the evaluators (d0, d1 or d2) is being called; `b` is the current guess for the vector of parameters, in the form of a row vector; `lnppp` is the name of the variable which will contain the contributions (per subject) to the sample log-likelihood evaluated at the parameter vector b. The next line (`tempvar`) introduces the various temporary variables which are required in the course of arriving at `lnppp`. The next line (`tempname`) introduces the temporary scalars that are required in the calculation; these are the (scalar) parameters of the model, the first of which is extracted from the vector b. The next line (local) defines a variable list containing the Halton variables $h1\_1, h1\_2, h1\_3\ldots$ that were defined outside the program.

```
program define my_rep
args todo b lnppp
tempvar xb p pp ppp upp uppp
tempname ln_s_u s_u
local hlist h1*
```

The next two commands are `mleval` commands, whose purpose is to extract the parameters of the model from the row vector b. The vector b actually may be seen as consisting of two "equations"; in the first, the "parameter" `xb` depends linearly on the variable $x$, with a constant term. Hence `xb` embodies two of the model's parameters, $\beta_1$ and $\beta_2$. Since $\beta_1$ and $\beta_2$ only enter the likelihood through `xb`, only `xb` is required in the construction of the likelihood.

The second "equation" in the vector b contains only one parameter, which is represented by the constant of the equation. This is why we include the option `scalar` at the end of the second `mleval` command; efficiency gains result from recognising that $\ln \sigma_u$ is a scalar and not a variable. The reason why $\ln\sigma_u$ is estimated instead of $\sigma_u$ is to ensure that $\sigma_u$ never becomes negative. The command immediately following this one transforms $\ln \sigma_u$ to $\sigma_u$ by applying the exponential function.

```
mleval `xb' = `b', eq(1)
mleval  `ln_s_u' = `b', eq(2) scalar
scalar `s_u'=exp(`ln_s_u')
```

Next we need to initialise the various local (i.e. temporary) variables. Usually these are set to missing valnes, and when they are used later in the program, the `replace` command is used. The exceptions here are the variables ppp and uppp, which are initialised to zero. This is because these two variables are built up as the cumulative sums of the quantities $\prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(\beta_1 + \beta_2 x_{it} + u_{r,i}\right)\right]$ and $u_{r,i} \times \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(\beta_1 + \beta_2 x_{it} + u_{r,i}\right)\right]$ respectively, over the $R$ Halton draws, $u_{1,i} \ldots u_{R,i}$. Cumulative sums must obviously start at zero.

```
quietly{
quietly gen double `p'=.
quietly gen double `pp'=.
quietly gen double `ppp'=0
quietly gen double `upp'=.
quietly gen double `uppp'=0
}
```

It is in the next stage of the program that the integral appearing in the log-likelihood function is evaluated. Recall that the integral is:

$$\int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left[yy_{it} \times (\beta_1 + \beta_2 x_{it} + u)\right] \frac{1}{\sigma_u} \phi\left(\frac{u}{\sigma_u}\right) \mathrm{d}u \tag{10.15}$$

and the method of evaluation is to find the following average over the $R$ Halton draws:

$$\frac{1}{R} \sum_{r=1}^{R} \left( \prod_{t=1}^{T} \Phi\left[yy_{it} \times \left(\beta_1 + \beta_2 x_{it} + u_{r,i}\right)\right] \right) \tag{10.16}$$

Evaluation of this mean requires a loop over the $R$ draws (i.e. over the $R$ variables in the variable list "`hlist`"), evaluating the product over $t$ (`pp` in the program) at each stage of the loop, and adding it to a cumulative sum (`ppp` in the program). The loop is programmed as follows:

```
foreach v of varlist `hlist' {
replace  `p'= normal(yy*(`xb' + `s_u'*`v'))
by i: replace `pp' = exp(sum(ln(`p')))
replace `pp'=. if last~=1
replace `upp'=`s_u'*`v'*`pp'
replace `ppp'=`ppp'+`pp'
replace `uppp'=`uppp'+`upp'
}
```

Regarding the above code, the `foreach` command is the command that requests the loop over the variables contained in `hlist`. The curly brackets contain the commands that are to be repeated. The first line inside the loop evaluates the standard normal cdf of the term in square brackets in (10.16). This term contains the subject specific effect $u_{r,i}$, which is assumed to be distributed $N(0, \sigma_u^2)$. The Halton draws (hl_1-hl_R) that have been generated using the method described in Section 10.3.3 correspond to a standard normal distribution. To obtain values of $u_{r,i}$, we therefore need to multiply the Halton draws by $\sigma_u$ (`s_u` in the code).



The second line inside the above loop is the one that obtains the product of probabilities over $t$ for each subject. The reasons for using this combination of STATA functions ($exp(sum(ln(.)))$) was explained fully in Section 8.5.5. This procedure will result in the required product appearing in the final row (of the variable `pp`) for each subject. For good measure, the third line replaces all values of `pp` other than those in the final row for each subject by missing values. The fourth line creates the variable corresponding to the integrand in the numerator of (10.12) (the posterior random effect formula). The final two lines of this loop add the new products to the cumulative sums of such products (`ppp` and `uppp`). When the loop is exited, these two sums will be sums over the $R$ draws.

Once outside the loop, we divide the two sums over draws by the number of draws, to give evaluations of the respective integrals (these are also named `ppp` and `uppp` in the program).

```
replace `ppp'=`ppp'/draws
replace `uppp'=`uppp'/draws
replace u_post=`uppp'/`ppp'
putmata u_post, replace
```

The variable named `ppp` now contains a value in the last row of each block, representing the likelihood contribution associated with a given subject. We then take the log of the likelihood, and use the `mlsum` command to indicate that this is the variable containing contributions to the log-likelihood. The subject-specific log-likelihood contribution appearing in this way, in the last row for each subject, is exactly what is required when the d0 evaluator is being used.

```
mlsum `lnppp'=ln(`ppp') if last==1
end
```

Notice that the above code includes generation of the (global) variable `u_post` which, again in the last row for each subject, contains the subject's posterior random effect, defined in (10.12). Even though this variable is global, in order to be readable outside of the program, it needs to be put into mata, and then extracted from mata (using `getmata`) outside the program.

The `end` command signifies the end of the evaluation program entitled "`my_rep`". Once outside the program, we need to perform a number of tasks before running the program. In the following piece of code, the first command defines a variable list containing the explanatory variables (of which there is only one in the present case). We next need to find starting values. A sensible approach is to obtain starting values from estimation of a simpler model. The second line of the following runs the simple probit model, while the third line stores the estimates. The fourth line then defines a vector of starting values for the random effects probit model, as the results from simple probit, appended by a guess (zero) of the value of the additional parameter $\ln \sigma_u$. We also initialise the global variable `u_post` which is generated inside the program.

```
local list_explan "x"
probit y `list_explan'
mat b_probit=e(b)
mat start = b_probit,0
gen double u_post=.
```

Most importantly of all, we need to call the program. This requires the following sequence of commands:

```
ml model d0 my_rep (  = 'list_explan')  /ln_s_u
ml init start, copy
ml max
nlcom s_u: exp(_b[ln_s_u:_cons])
```

The first line of the above specifies the evaluator, the evaluation program, and the parameter list. The second line brings in the vector of starting values. The third line executes the `ml` routine. The final line calls on the delta method to deduce an estimate of $\sigma_u$ from the estimate of $\ln\sigma_u$.[3]

Finally, we need to extract the posterior random effect variable (u_post) from mata. Having deleted the existing variable from memory, we use the getmata command to do this.

```
drop u_post
getmata u_post
```

## 10.4.6 Complete annotated code

Here, we show the complete uninterrupted code, including both the simulation and the estimation. The code is annotated: comment lines (commencing with *) provide brief descriptions of what follows in the code.

```
clear
set more off

* SET SAMPLE SIZE AND RANDOM NUMBER SEED

set obs 1200
set seed 7654321

*  GENERATE SUBJECT IDENTIFIER (i) AND TASK IDENTIFIER (t);
* ENSURE THAT THESE ARE STORED AS INTEGERS
egen i=seq(), f(1) b(20)
egen t=seq(), f(1) t(20)
recast int i t

*  DECLARE DATA TO BE PANEL DATA

xtset i t

* GENERATE x (FROM UNIFORM) AND e (FROM NORMAL)

gen double x=runiform()
gen double e=rnormal()

* GENERATE u (SUBJECT-SPECIFIC EFFECT)

by i: generate double u=0.5*(rnormal()) if _n==1
by i: replace u=u[1] if u==.
```

[3] In the `nlcom` command the parameter estimate is referred to with the name `_b[ln_s_u:_cons]`. This is known as the "legend" associated with the estimate. In order to find the legend of each parameter estimate, the command `ml coefleg` may be used immediately after the `ml` command.



```
* GENERATE LATENT VARIABLE y*, AND BINARY VARIABLES y and yy

gen double ystar=-1.0+2.0*x+u+e
gen int y=ystar>0
gen int yy=2*y-1

* ESTIMATE RANDOM EFFECTS PROBIT MODEL USING xtprobit COMMAND

xtprobit y x

* GENERATE INDICATOR VARIABLES FOR FIRST AND LAST OBSERVATION FOR EACH SUBJECT

by i: gen int first=1 if _n==1
by i: gen int last=1 if _n==_N

* APPEND (HORIZONTALLY) EACH SUBJECT'S FIRST ROW WITH 125 HALTON DRAWS
* (DIFFERENT BETWEEN SUBJECTS).  STORE NUMBER OF DRAWS AS "draws".

mat p=[3]
mdraws if first==1, neq(1) dr(125) prefix(h) primes(p)
scalar draws=r(n_draws)

*CREATE A VARIABLE LIST CONTAINING THE HALTON DRAWS
* ENSURE THEY ARE IN DOUBLE PRECISION
* COPY THE ROW OF HALTONS IN EACH BLOCK INTO ROWS 2-T OF SAME BLOCK

local hlist h1*

quietly{
foreach v of varlist 'hlist' {
recast double 'v'
by i: replace 'v'='v'[1] if 'v'==.
replace 'v'=invnorm('v')
}
}

* LIKELIHOOD EVALUATION PROGRAM "my_rep" STARTS HERE:

capt prog drop my_rep
program define my_rep

* SPECIFY ARGUMENTS

args todo b lnppp
tempvar xb p pp ppp upp uppp
tempname ln_s_u s_u
local hlist h1*

* EXTRACT ELEMENTS OF PARAMETER VECTOR b

mleval 'xb' = 'b', eq(1)
mleval  'ln_s_u' = 'b', eq(2) scalar
scalar 's_u'=exp('ln_s_u')

* INITIALISE TEMPORARY VARIABLES

quietly gen double 'p'=.
quietly gen double 'pp'=.
quietly gen double 'ppp'=0
quietly gen double 'upp'=.
quietly gen double 'uppp'=0

* LOOP FOR EVALUATION OF SUM (OVER r) OF PRODUCT (OVER t)
* pp AND ppp ARE FOR LIKELIHOOD FUNCTION;
* upp AND uppp ARE FOR NUMERATOR OF POSTERIOR RANDOM EFFECT FORMULA

quietly{
foreach v of varlist 'hlist' {
replace  'p'= normal(yy*('xb' + 's_u'*'v'))
by i: replace 'pp' = exp(sum(ln('p')))
replace 'pp'=. if last~=1
replace 'upp'='s_u'*'v'*'pp'
replace 'ppp'='ppp'+'pp'
```

```
replace `uppp'=`uppp'+`upp'
}

* DIVISION BY R TO GENERATE REQUIRED AVERAGES (OVER r)
* COMPUTE POSTERIOR RANDOM EFFECT VARIABLE (u_post) AND SEND THIS TO MATA

quietly {
replace `ppp'=`ppp'/draws
replace `uppp'=`uppp'/draws
replace u_post=`uppp'/`ppp'
}
putmata u_post, replace

* MLSUM COMMAND TO SPECIFY PER-SUBJECT LOG-LIKELIHOOD CONTRIBUTION

mlsum `lnppp'=ln(`ppp') if last==1
}
end

* "end" SIGNIFIES END OF LIKELIHOOD EVALUATION PROGRAM "my_rep"

* CREATE VARIABLE LIST (list_explan) FOR EXPLANATORY VARIABLES;
* ESTIMATE SIMPLE PROBIT MODEL
* STORE ESTIMATES FROM SIMPLE PROBIT MODEL
* CREATE VECTOR OF STARTING VALUES (start) FOR PANEL PROBIT MODEL
* INITIALISE VARIABLE CONTAINING POSTERIOR RANDOM EFFECT (u_post)

local list_explan "x"
probit y `list_explan'
mat b_probit=e(b)
mat start = b_probit,0
gen double u_post=.

* SPECIFY EVALUATOR (d0), EVALUATION PROGRAM (my_rep), AND PARAMETER LIST
* SPECIFY STARTING VALUE VECTOR
* RUN MAXIMUM LIKELIHOOD PROCEDURE; DEDUCE ESTIMATE OF s_u USING DELTA METHOD

ml model d0 my_rep ( = `list_explan') /ln_s_u
ml init start, copy
ml max
nlcom s_u: exp(_b[ln_s_u:_cons])

* EXTRACT POSTERIOR RANDOM EFFECT (u_post) GENERATED INSIDE EVALUATION PROGRAM
* PLOT IT AGAINST TRUE RANDOM EFFECT (u).

drop u_post
getmata u_post
lowess u_post u, xline(0) yline(0)
```

### 10.4.7 Results

The results from the MSL program described in detail in previous sections are as follows.

```
. ml max

initial:       log likelihood = -724.34505
rescale:       log likelihood = -724.34505
rescale eq:    log likelihood = -724.34505
Iteration 0:   log likelihood = -724.34505
Iteration 1:   log likelihood = -711.86049
Iteration 2:   log likelihood = -710.30549
Iteration 3:   log likelihood = -710.28495
Iteration 4:   log likelihood = -710.28494

                                                Number of obs   =       1200
                                                Wald chi2(1)    =     157.36
Log likelihood = -710.28494                     Prob > chi2     =     0.0000
```



```
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
           x |    1.84808   .1473262    12.54   0.000     1.559326    2.136834
       _cons |  -.9092081   .1087404    -8.36   0.000    -1.122335   -.6960808
-------------+----------------------------------------------------------------
ln_s_u       |
       _cons |  -.6117916   .1250433    -4.89   0.000    -.8568721   -.3667112
------------------------------------------------------------------------------

. nlcom s_u: exp(_b[ln_s_u:_cons])

         s_u:  exp(_b[ln_s_u:_cons])

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         s_u |   .5423782   .0678208     8.00   0.000      .409452    .6753045
------------------------------------------------------------------------------
```

For ease of comparison, we present again the results from using the `xtprobit` command (previously reported in Section 10.4.2).

```
. xtprobit y x, re

Random-effects probit regression                Number of obs      =      1200
Group variable: i                               Number of groups   =        60

Random effects u_i ~ Gaussian                   Obs per group: min =        20
                                                               avg =      20.0
                                                               max =        20

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(1)       =    157.25
Log likelihood  = -710.41547                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.847581    .147337    12.54   0.000     1.558806    2.136356
       _cons |  -.9181971   .1090388    -8.42   0.000    -1.131909   -.704485
-------------+----------------------------------------------------------------
    /lnsig2u |  -1.178868    .253014                     -1.674766   -.6829695
-------------+----------------------------------------------------------------
     sigma_u |   .5546412    .070166                      .4328418    .7107143
         rho |   .2352558   .0455199                      .1577898    .3355989
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =    95.05 Prob >= chibar2 = 0.000
```

We see that the three parameter estimates and accompanying standard errors obtained using the MSL routine are admirably close to the corresponding estimates obtained using the `xtprobit` command, with differences appearing only in the second or third decimal place. The maximised log-likelihoods are also in close agreement. This is all reassuring since it indicates that the MSL routine has been programmed correctly and is doing what we intended it to do.

Posterior random effects ($\hat{u}_i$) were defined in (10.12). The best way to think of these is as estimates of the random effect term $u_i$ for each subject. We have generated a variable `u_post` containing these posterior random effects, and a useful way of examining them is to plot them against the true random effect (u). This is done in Figure 10.6. The close correspondence between these two variables is further confirmation that the computations have been performed correctly.

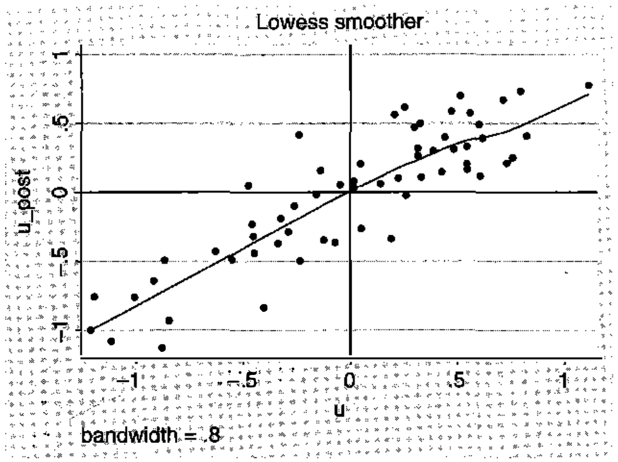



Figure 10.6: Scatter (with smoother) of posterior random effect against true random effect for 60 subjects in simulation

## 10.5 The Random Effects Two-limit Tobit Model

In this section, we apply the MSL approach to a different model, the random effects two-limit Tobit model. A further difference is that in this section we will use real data rather than simulated data.

In Section 6.6.3, we applied the two-limit Tobit model to the data set of Bardsley (2000) in order to test the theory of reciprocity. The data was on 98 subjects who each performed 20 tasks. The dependent variable was the contribution by a particular subject in a particular task, and the key explanatory variable was the median of contributions of subjects placed earlier in the sequence. The strongly positive effect of previous contributions on the subject's contribution was interpreted as strong evidence of reciprocity in this data. It was noted that the dependent variable was censored from below at zero, and censored from above at 10 (the maximum contribution), and this was the reason for estimating the two-limit Tobit model.

A point made in Section 6.6.3 is that, because this data set includes repeated observations for each subject, and because there are obvious between-subject differences in contribution behaviour, what is actually required is a panel data model, namely, the random effects two-limit Tobit model. Like the random effects probit model considered earlier in this chapter, the random effects two-limit Tobit model is a model that can be estimated using a STATA command (`xttobit`). Here, once again, we will set out to estimate the model using MSL, again on Bardsley's (2000) data, and then compare the results to those obtained using `xttobit`. We will also go further than the `xttobit` command allows by extending our program to compute posterior random effects for each subject.

Recall that the file **bardsley** contains data from 98 subjects who each performed 20 tasks. The variables to be used here are:



| | |
|---|---|
| i: | subject id; |
| t: | task id; |
| y: | subject's contribution to the public fund; |
| med: | median of contributions by others in the group who have already contributed; |
| tsk: | task number. |

### 10.5.1 Construction of log-likelihood function

The two-limit Tobit model, appropriate when the data consists of a single cross section, was developed and estimated in Section 6.6.3. Here, we extend the model to the panel data setting. The resulting model is the random effects two-limit Tobit model, defined as follows:

$$\begin{aligned} y_{it}^* &= \beta_0 + \beta_1 med_{it} + \beta_2 tsk_{it} + u_i + \epsilon_{it} \\ &= x_{it}'\beta + u_i + \epsilon_{it} \\ u_i &\sim N\left(0, \sigma_u^2\right) \\ \epsilon_{it} &\sim N\left(0, \sigma_\epsilon^2\right) \end{aligned} \quad (10.17)$$

$y_{it}^*$ is *desired* contribution by subject $i$ in task $t$. $y_{it}$ is *observed* contribution. The *censoring rules* are:

$$\begin{aligned} y_{it} &= 0 \text{ if } y_{it}^* \leq 0 \\ y_{it} &= y_{it}^* \text{ if } 0 < y_{it}^* < 10 \\ y_{it} &= 10 \text{ if } y_{it}^* \geq 10 \end{aligned} \quad (10.18)$$

As in Section 6.6.3, we say that there are three *regimes*. To obtain the likelihood contributions (conditional on $u_i$), we consider each regime in turn:

$$y_{it} = 0: \quad P(y_{it} = 0|u_i) = P\left(y_{it}^* \leq 0|u_i\right) = \Phi\left(-\frac{x_{it}'\beta + u_i}{\sigma_\epsilon}\right) \quad (10.19)$$

$$0 < y_{it} < 10: \quad f(y_{it}|u_i) = \frac{1}{\sigma_\epsilon}\phi\left(\frac{y_{it} - x_{it}'\beta - u_i}{\sigma_\epsilon}\right) \quad (10.20)$$

$$y_{it} = 10: \quad P(y_{it} = 10|u_i) = P\left(y_{it}^* \geq 10|u_i\right) = \Phi\left(\frac{x_{it}'\beta + u_i - 10}{\sigma_\epsilon}\right) \quad (10.21)$$

By applying the indicator function $I_{(.)}$ to each of (10.19) – (10.21), we may obtain the joint density (conditional on $u_i$) of the $T$ decisions of subject $i$.

$$L_i|u_i = \prod_{t=1}^{T}\left[\Phi\left(-\frac{x_{it}'\beta + u_i}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=0}}\left[\frac{1}{\sigma_\epsilon}\phi\left(\frac{y_{it} - x_{it}'\beta - u_i}{\sigma_\epsilon}\right)\right]^{I_{0<y_{it}<10}} \left[\Phi\left(\frac{x_{it}'\beta + u_i - 10}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=10}} \quad (10.22)$$

To obtain the marginal joint density from the conditional joint density (10.22), we integrate over the normal variate $u$:

$$L_i = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \left[\Phi\left(-\frac{x_{it}'\beta + u}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=0}} \left[\frac{1}{\sigma_\epsilon}\phi\left(\frac{y_{it} - x_{it}'\beta - u}{\sigma_\epsilon}\right)\right]^{I_{0<y_{it}<10}} \left[\Phi\left(\frac{x_{it}'\beta + u - 10}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=10}} \frac{1}{\sigma_u}\phi\left(\frac{u}{\sigma_u}\right) du \tag{10.23}$$

The $L_i$s defined in (10.23) are the subject likelihood contributions. These are logged and summed to give the sample log-likelihood function:

$$LogL = \sum_{i=1}^{n} \ln L_i \tag{10.24}$$

The posterior random effect may be obtained as:

$$\hat{u}_i = E(u_i|y_{i1}\dots y_{iT}) = \frac{1}{\hat{L}_i}\int_{-\infty}^{\infty}\left(u\prod_{t=1}^{T}\left[\Phi\left(-\frac{x_{it}'\hat{\beta}+u}{\hat{\sigma}_\epsilon}\right)\right]^{I_{y_{it}=0}} \times \left[\frac{1}{\hat{\sigma}_\epsilon}\phi\left(\frac{y_{it}-x_{it}'\hat{\beta}-u}{\hat{\sigma}_\epsilon}\right)\right]^{I_{0<y_{it}<10}} \left[\Phi\left(\frac{x_{it}'\hat{\beta}+u-10}{\hat{\sigma}_\epsilon}\right)\right]^{I_{y_{it}=10}} \frac{1}{\hat{\sigma}_u}\phi\left(\frac{u}{\hat{\sigma}_u}\right)\right)du \tag{10.25}$$

where hats over parameters indicate estimates, and where $\hat{L}_i$ is (10.23) with parameters replaced by estimates.

## 10.5.2 Estimation by MSL

Applying the MSL principle, we may write the log-likelihood function (10.24) as:

$$LogL = \sum_{i=1}^{n}\ln\left(\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T}\left[\Phi\left(-\frac{x_{it}'\beta+u_{r,i}}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=0}} \times \left[\frac{1}{\sigma_\epsilon}\phi\left(\frac{y_{it}-x_{it}'\beta-u_{r,i}}{\sigma_\epsilon}\right)\right]^{I_{0<y_{it}<10}} \left[\Phi\left(\frac{x_{it}'\beta+u_{r,i}-10}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=10}}\right)\right) \tag{10.26}$$

where $u_{1,i}\dots u_{R,i}$ are the Halton draws for subject $i$, transformed in such a way as to represent draws from a $N(0,\sigma_u^2)$ distribution.

In the STATA code that follows, we adopt the names shown in Table 10.1 for the parameters and the other components of the log-likelihood function (10.26).



| Component of LogL | STATA name |
|---|---|
| $x_{it}'\beta = \beta_0 + \beta_1 med_{it} + \beta_2 tsk_{it}$ | xb |
| $\ln(\sigma_u)$ | ln_s_u |
| $\sigma_u$ | s_u |
| $\ln(\sigma_e)$ | ln_s_e |
| $\sigma_e$ | s_e |
| $p_{r,i} \equiv \left[\Phi\left(-\frac{x_{it}'\beta+u_{r,i}}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=0}}\left[\frac{1}{\sigma_\epsilon}\phi\left(\frac{y_{it}-x_{it}'\beta-u_{r,i}}{\sigma_\epsilon}\right)\right]^{I_{0<y_{it}<10}}\left[\Phi\left(\frac{x_{it}'\beta+u_{r,i}-10}{\sigma_\epsilon}\right)\right]^{I_{y_{it}=10}}$ | p |
| $\prod_{t=1}^{T} p_{r,i}$ | pp |
| $u_{r,i} \times \prod_{t=1}^{T} p_{r,i}$ | upp |
| $\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T} p_{r,i}\right)$ | ppp |
| $\frac{1}{R}\sum_{r=1}^{R}\left(u_{r,i} \times \prod_{t=1}^{T} p_{r,i}\right)$ | uppp |
| $R$ | draws |
| $\ln L_i$ | logl |

Table 10.1: Components of LogL and corresponding STATA names

## 10.5.3 STATA code

The random effects two-limit Tobit model defined in (10.17) and (10.18) may be estimated by MSL using the following code. The code is annotated: comment lines (commencing with *) provide brief descriptions of what follows in the code.

```
* LIKELIHOOD EVALUATION PROGRAM "my_ret" STARTS HERE:

capt prog drop my_ret
program define my_ret

* SPECIFY ARGUMENTS

args todo b logl
tempvar xb p pp ppp upp uppp
tempname ln_s_u s_u ln_s_e s_e
local hlist h1*

* EXTRACT ELEMENTS OF PARAMETER VECTOR b

mleval 'xb' = 'b', eq(1)
mleval  'ln_s_u' = 'b', eq(2) scalar
mleval 'ln_s_e' = 'b', eq(3) scalar
scalar 's_u'=exp('ln_s_u')
scalar 's_e'=exp('ln_s_e')

* INITIALISE TEMPORARY VARIABLES

quietly gen double 'p'=.
quietly gen double 'pp'=.
```

```
quietly gen double `ppp'=0
quietly gen double `upp'=.
quietly gen double `uppp'=0

* LOOP FOR EVALUATION OF SUM (OVER r) OF PRODUCT (OVER t)
* pp AND ppp ARE FOR LIKELIHOOD FUNCTION;
* upp AND uppp ARE FOR NUMERATOR OF POSTERIOR RANDOM EFFECT FORMULA

quietly{
foreach v of varlist `hlist' {
replace  `p'= normal(-(`xb' + `s_u'*`v')/`s_e') if y==0
replace  `p'=(1/`s_e')*normalden((y-`xb'-`s_u'*`v')/`s_e') if (y>0)&(y<10)
replace  `p'= normal((-10+(`xb' + `s_u'*`v'))/`s_e') if y==10
by i: replace `pp' = exp(sum(ln(`p')))
replace `pp'=. if last~=1
replace `upp'=`s_u'*`v'*`pp'
replace `ppp'=`ppp'+`pp'
replace `uppp'=`uppp'+`upp'
}

* DIVISION BY R TO GENERATE REQUIRED AVERAGES (OVER r)
* COMPUTE POSTERIOR RANDOM EFFECT VARIABLE (u_post) AND SEND THIS TO MATA

quietly {
replace `ppp'=`ppp'/draws
replace `uppp'=`uppp'/draws
replace u_post=`uppp'/`ppp'
}
putmata u_post, replace

* MLSUM COMMAND TO SPECIFY PER-SUBJECT LOG-LIKELIHOOD CONTRIBUTION

mlsum `logl'=ln(`ppp') if last==1
}
end

* "END" SIGNIFIES END OF LIKELIHOOD EVALUATION PROGRAM "my_rep"

* READ DATA AND DECLARE TO BE PANEL DATA

use bardsley, clear
xtset i t

* GENERATE INDICATOR VARIABLES FOR FIRST AND LAST OBSERVATION FOR EACH SUBJECT

by i: gen int first=1 if _n==1
by i: gen int last=1 if _n==_N

* APPEND (HORIZONTALLY) EACH SUBJECT'S FIRST ROW WITH 125 HALTON DRAWS
* (DIFFERENT BETWEEN SUBJECTS).  STORE NUMBER OF DRAWS AS "draws".

mat p=[3]
mdraws if first==1, neq(1) dr(125) prefix(h) primes(p)
scalar draws=r(n_draws)

*CREATE A VARIABLE LIST CONTAINING THE HALTON DRAWS
* ENSURE THEY ARE IN DOUBLE PRECISION
* COPY THE ROW OF HALTONS IN EACH BLOCK INTO ROWS 2-T OF SAME BLOCK

local hlist h1*

quietly{
foreach v of varlist `hlist' {
recast double `v'
by i: replace `v'=`v'[1] if `v'==.
replace `v'=invnorm(`v')
}
}

* CREATE VARIABLE LIST (list_explan) FOR EXPLANATORY VARIABLES;
* ESTIMATE 2-LIMIT TOBIT MODEL
* STORE ESTIMATES FROM 2-LIMIT TOBIT MODEL
* CREATE VECTOR OF STARTING VALUES (start) FOR RANDOM EFFECTS 2-LIMIT TOBIT MODEL
```



```
* INITIALISE VARIABLE CONTAINING POSTERIOR RANDOM EFFECT (u_post)

local list_explan "med tsk"
tobit y `list_explan', ll(0) ul(10)
mat b_tobit=e(b)
mat ln_s_e=ln(b_tobit[1,4])
mat start=b_tobit[1,1..3],0,ln_s_e
gen double u_post=.

* SPECIFY EVALUATOR (d0), EVALUATION PROGRAM (my_rep), AND PARAMETER LIST
* SPECIFY STARTING VALUE VECTOR
* RUN MAXIMUM LIKELIHOOD PROCEDURE; DEDUCE ESTIMATES OF s_u and s_e USING DELTA
  METHOD

ml model d0 my_ret ( = `list_explan')  /ln_s_u  /ln_s_e
ml init start, copy
ml max
nlcom (s_u: exp(_b[ln_s_u:_cons])) (s_e: exp(_b[ln_s_e:_cons]))

* EXTRACT POSTERIOR RANDOM EFFECT (u_post) GENERATED INSIDE EVALUATION PROGRAM
* PLOT POSTERIOR RANDOM EFFECT AGAINST SUBEJECT'S MEAN CONTRIBUTION

drop u_post
getmata u_post
by i: egen mean_y=mean(y)
scatter u_post mean_y, yline(0)
}
}

* CREATE VARIABLE LIST (list_explan) FOR EXPLANATORY VARIABLES;
* ESTIMATE 2-LIMIT TOBIT MODEL
* STORE ESTIMATES FROM 2-LIMIT TOBIT MODEL
* CREATE VECTOR OF STARTING VALUES (start) FOR RANDOM EFFECTS 2-LIMIT TOBIT MODEL
* INITIALISE VARIABLE CONTAINING POSTERIOR RANDOM EFFECT (u_post)

local list_explan "med tsk"
tobit y `list_explan', ll(0) ul(10)
mat b_tobit=e(b)
mat ln_s_e=ln(b_tobit[1,4])
mat start=b_tobit[1,1..3],0,ln_s_e
gen double u_post=.

* SPECIFY EVALUATOR (d0), EVALUATION PROGRAM (my_rep), AND PARAMETER LIST
* SPECIFY STARTING VALUE VECTOR
* RUN MAXIMUM LIKELIHOOD PROCEDURE; DEDUCE ESTIMATES OF s_u and s_e USING DELTA
  METHOD

ml model d0 my_ret ( = `list_explan')  /ln_s_u  /ln_s_e
ml init start, copy
ml max
nlcom (s_u: exp(_b[ln_s_u:_cons])) (s_e: exp(_b[ln_s_e:_cons]))

* EXTRACT POSTERIOR RANDOM EFFECT (u_post) GENERATED INSIDE EVALUATION PROGRAM
* PLOT POSTERIOR RANDOM EFFECT AGAINST SUBEJECT'S MEAN CONTRIBUTION

drop u_post
getmata u_post
by i: egen mean_y=mean(y)
scatter u_post mean_y, yline(0)
```

## 10.5.4 Results from random effects two-limit Tobit model

The output from running the code listed in Section 10.5.3 is as follows:

```
. ml max

initial:      log likelihood = -3688.5167
rescale:      log likelihood = -3688.5167
```

```
rescale eq:    log likelihood = -3688.5167
Iteration 0:   log likelihood = -3688.5167  (not concave)
Iteration 1:   log likelihood = -3405.0938
Iteration 2:   log likelihood = -3347.8512
Iteration 3:   log likelihood = -3341.1692
Iteration 4:   log likelihood =   -3341.16
Iteration 5:   log likelihood =   -3341.16

                                                  Number of obs   =      1960
                                                  Wald chi2(2)    =    188.36
Log likelihood =   -3341.16                       Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
         med |   .4176157   .0330514    12.64   0.000     .3528363    .4823952
         tsk |  -.0701045   .0117759    -5.95   0.000    -.0931848   -.0470242
       _cons |  -.5672377   .5899609    -0.96   0.336     -1.72354    .5890645
-------------+----------------------------------------------------------------
ln_s_u       |
       _cons |   1.629583   .0771286    21.13   0.000     1.478414    1.780753
-------------+----------------------------------------------------------------
ln_s_e       |
       _cons |   1.356453   .0257532    52.67   0.000     1.305978    1.406929
------------------------------------------------------------------------------

. nlcom (s_u: exp(_b[ln_s_u:_cons])) (s_e: exp(_b[ln_s_e:_cons]))

         s_u:  exp(_b[ln_s_u:_cons])
         s_e:  exp(_b[ln_s_e:_cons])

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         s_u |   5.101748   .3934908    12.97   0.000     4.330521    5.872976
         s_e |   3.882399   .0999842    38.83   0.000     3.686433    4.078364
------------------------------------------------------------------------------
```

Again we can compare results with those obtained using the built-in STATA command:

```
Random-effects tobit regression                 Number of obs      =      1960
Group variable: i                               Number of groups   =        98

Random effects u_i ~ Gaussian                   Obs per group: min =        20
                                                               avg =      20.0
                                                               max =        20

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(2)       =    188.38
Log likelihood  = -3341.4572                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         med |   .4177111   .0330561    12.64   0.000     .3529224    .4824999
         tsk |  -.0701086   .0117791    -5.95   0.000    -.0931951   -.0470221
       _cons |   -.543441   .5858525    -0.93   0.354    -1.691691    .6048088
-------------+----------------------------------------------------------------
    /sigma_u |   5.032513   .4475302    11.25   0.000      4.15537    5.909656
    /sigma_e |   3.883217   .1000142    38.83   0.000     3.687193    4.079242
-------------+----------------------------------------------------------------
         rho |   .6267995   .0416354                      .5429457    .7050288
------------------------------------------------------------------------------
  Observation summary:        872  left-censored observations
                              952     uncensored observations
                              136 right-censored observations
```



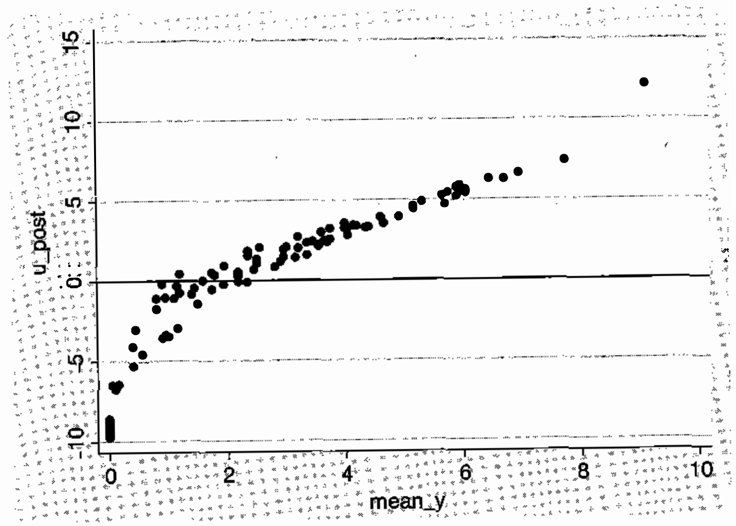



Figure 10.7: Posterior random effect against mean contribution for the 98 subjects in Bardsley's experiment

Notice once again the close agreement between the two sets of estimates. There is particularly close agreement on the two slope estimates (i.e. the effects of reciprocity and learning) and their standard errors, for which differences appear only in the fourth decimal place.

The estimate of the reciprocity parameter is +0.418, which not very different from the estimate of +0.429 obtained with "pooled" two-limit Tobit in Section 6.6.3. Likewise, the estimate of the learning parameter is −0.070, which is not very different from the pooled estimate of −0.066. It is perhaps surprising that these differences are not greater, on the basis of the evidence arising from the Monte Carlo simulation performed in Section 9.3.4, in which we found that neglecting between-subject heterogeneity in the context of the binary probit model led to a 30% bias in the estimation of the slope parameter. However, we also notice that the estimated standard errors of these two parameters are both around 25% lower in the random effects model than in the pooled model. This difference reflects the superior efficiency of the random effects estimator.

In Figure 10.7 we plot posterior random effect $\hat{u}_i$, defined in (10.25) and represented by "u_post" in the above code, against the subject's mean contribution. Unsurprisingly, subjects contributing a larger amount on average are estimated to have a larger random effect, and this positive relationship is very clear.

## 10.6 Summary and Further Reading

This chapter has been dedicated to explaining and demonstrating in a step-by-step fashion the method of MSL, a technique for dealing with heterogeneity which is used many times later in the book. Readers seeking more information about MSL and related estimation techniques are referred to Train (2003).

An important ingredient in the approach to MSL adopted here is the use of Halton draws. Readers interested in Halton sequences are referred to Halton (1960). Our chosen method for obtaining Halton sequences is the user-written STATA command mdraws, which is documented in Cappellari & Jenkins (2006).

For an application of MSL to the proposer's decision in the ultimatum game, the reader is referred to Bellemare et al. (2008). Applications to risky choice modelling include Von Gaudecker et al. (2011), Conte et al. (2011) and Moffatt et al. (2015).

## Exercise

Write down the first few numbers of the Halton sequence with $p = 5$. Check that your answers are correct using the STATA commands:

```
mat p=5
mdraws, neq(1) dr(1) prefix(h) primes(p)
```

# Chapter 11
# Dealing with Zeros: Hurdle Models

## 11.1 Introduction

The theme of this chapter is zero observations in experimental data, and how to allow for them in estimation. A class of model tailored for data sets containing zero observations that has recently gained prominence in mainstream Econometrics is the hurdle modelling framework. Hurdle models are potentially very useful in experimental economics. Two settings in which they are likely to be particularly useful are dictator games and public goods games. In these settings, it is common for subjects to make contributions of zero, and indeed this is the Nash equilibrium prediction in both cases. The incorporation of zero censoring is one way of dealing with such observations. However, experimental economists have been known to express unease at the principle of censoring – the idea that the contribution would be negative were it not for the exogenously set lower limit of zero. Many experimental economists prefer to think in terms of a "zero-type". In a dictator game, we might refer to a subset of subjects as of "selfish" type, and these subjects choose to donate zero to the other player; in a public goods context, we might refer to the "free-rider" type, who always contributes zero to the public fund. The hurdle class of model offers a natural means of incorporating the "zero-type", while also allowing zero-censoring. Moreover, the hurdle model allows estimation of the proportion of the population who are of the zero-type. Better still, it allows the probability of a subject being of zero-type to depend on subject characteristics and/or treatments.

At the outset, it is necessary to settle an important issue of terminology. Some authors working with data containing zeros have estimated two models separately – perhaps a binary probit model capturing the 0/+ dichotomy, and a truncated regression model caputuring the amount – and presented the combined set of results as "double hurdle" estimates. We would instead label such an estimation strategy a "two-part model". The two-part model is a useful strategy in some situations; indeed, as we shall demonstrate in the examples used in this chapter, it provides a useful means of obtaining adequate starting values for the estimation of the full model. However, we reserve the "hurdle" label for a model that combines the two equations and estimates both sets of parameters in a "full-information" sense. A demonstration of how to estimate all parameters simultaneously using full-information methods is the central objective of this chapter.

The hurdle model can be thought of as an example of a mixture model of the type introduced in Chapter 8. There are only two types in the mixture: selfish and non-selfish; or free-rider and non-free-rider. As in the examples of Chapter 8, one of the end-products of hurdle model estimation will be to assign each subject a posterior probability of being each of these types.

In addition to introducing the hurdle class of model to the experimental economics literature, this chapter makes the further advance of extending the technique to deal with panel data. This is clearly important in experimental economics. As we have seen many times in this text, it is common for subjects to engage in a sequence of tasks over the course of an experiment, and hence panel data models are called for. The extension of the hurdle model to panel data is made tricky by the fact that the outcome of the "first hurdle", that is, the determination of whether a subject is of the zero-type, must apply to that subject for the whole experiment. Switching in and out of the zero-type is ruled out. In contrast, the outcome of the second hurdle, that is, the amount actually contributed in any particular task, is determined at the level of individual observations.

## 11.2 Review of Tobit and Random Effects Tobit

It is essential to start by reviewing the Tobit model, since the hurdle model is an extension of it. However, the review will be brief, because the Tobit model has already been explained in some detail in Chapter 6, and the random effects version in Section 10.5.

The Tobit model is required when the dependent variable is censored, that is, when there is an accumulation of observations at the limits of the range of the variable. In Chapter 6, and again in Section 10.5, we were concerned with the version of the model with three regimes: lower censored, uncensored and upper censored. That version was the two-limit Tobit model. Here, since the focus is on zero observations, we will be mainly interested in the version with only two regimes: lower censored (at zero) and uncensored.

Firstly assuming only one observation per subject, the model is based on the following equation:

$$\begin{aligned} y_i^* &= x_i'\beta + \epsilon_i \\ \epsilon &\sim N(0, \sigma^2) \end{aligned} \tag{11.1}$$

where $y_i^*$, is a latent (unobserved) variable representing subject i's *desired* contribution. The desired contribution is assumed to be a linear function of the observed subject characteristics and/or treatment variables contained in the vector $x_i$, plus a normally distributed random error. The important feature of $y_i^*$ is that it can be negative: subjects are permitted to desire to contribute a negative amount. Of course, if a subject does desire to contribute a negative amount, the amount they will actually



contribute will be zero, as a consequence of the lower bound at zero; if they desire to contribute any positive amount, this positive amount will be their actual observed contribution. This gives rise to what is known as the *censoring rule* – the rule showing the relationship between desired ($y_i^*$) and observed ($y_i$) contributions:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* < 0 \end{cases} \tag{11.2}$$

In this situation in which we have only lower censoring at zero, there are two "regimes" of behaviour: zero observations and positive observations. To obtain the likelihood contributions, we consider the two regimes separately:

$$\begin{aligned} y_i = 0 : P(y_i = 0) &= p(y_i^* \leq 0) = P(x_i'\beta + \epsilon_i \leq 0) = P\left(\frac{\epsilon_i}{\sigma} \leq -\frac{x_i'\beta}{\sigma}\right) \\ &= \Phi\left(-\frac{x_i'\beta}{\sigma}\right) \\ y_i > 0 : f(y_i) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}\right) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \end{aligned} \tag{11.3}$$

Recall that $\phi(.)$ is the standard normal density function, $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$, and $\Phi(.)$ is the standard normal distribution function, $\Phi(a) = P(Z < a)$. The relationship between the two functions is of course $\Phi(a) = \int_{-\infty}^{a} \phi(z)dz$.

Putting the two likelihood contributions in (11.3) together, the sample log-likelihood becomes:

$$LogL = \sum_{i=1}^{n} \left[ I_{y_i=0} \ln\left(\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right) + I_{y_i>0} \ln\left(\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right)\right] \tag{11.4}$$

where $I(.)$ is the *indicator function*, taking the value one if the subscripted expression is true, and zero otherwise. $LogL$ is maximised with respect to the parameters contained in the vector $\beta$, and the standard deviation parameter $\sigma$.

In Section 10.5, we considered the random effects two-limit Tobit model that is appropriate for panel data. Let $i$ index subjects and $t$ indicate tasks. In the case of two regimes, the random effects tobit model is defined by the latent equation:

$$\begin{aligned} y_{it}^* &= x_{it}'\beta + u_i + \epsilon_{it} \\ u_i &\sim N\left(0, \sigma_u^2\right) \\ \epsilon_{it} &\sim N\left(0, \sigma_\epsilon^2\right) \end{aligned} \tag{11.5}$$

and the censoring rules:

$$\begin{aligned} y_{it} &= 0 \text{ if } y_{it}^* \leq 0 \\ y_{it} &= y_{it}^* \text{ if } y_{it}^* > 0 \end{aligned} \tag{11.6}$$

The sample log-likelihood for the random effects Tobit model defined in (11.5) and (11.6) is:

$$LogL = \sum_{i=1}^{n} \ln \int_{-\infty}^{\infty} \prod_{t=1}^{T} \left[ \Phi \left( -\frac{x_{it}'\beta + u}{\sigma_\epsilon} \right) \right]^{I_{y_{it}=0}} \left[ \frac{1}{\sigma_\epsilon} \phi \left( \frac{y_{it} - x_{it}'\beta - u}{\sigma_\epsilon} \right) \right]^{I_{y_{it}>0}} \frac{1}{\sigma_u} \phi \left( \frac{u}{\sigma_u} \right) \mathrm{d}u \tag{11.7}$$

To estimate the standard Tobit model (defined in (11.1) and (11.2)) in STATA requires the command:

```
tobit y x1 x2 x3, ll(0)
```

while the random effects Tobit model (defined in (11.5) and (11.6)) requires the commands:

```
xtset i t
xttobit y x1 x2 x3, ll(0)
```

## 11.3 The Need for Hurdle Models

A feature the Tobit models reviewed in Section 11.2 is that the process which results in a positive contribution is assumed to be the same as that which determines the extent of the contribution. Thus, for example, if a particular subject-characteristic, or treatment, is known to have a strong positive effect on the *extent* of contribution, then the presence of this characteristic or treatment would inevitably lead to the prediction of a *positive* contribution for the subject in question. While such assumptions may turn out to hold, there is no reason to expect this *a priori*.[1] One reason why such an assumption might fail is that there may exist a proportion of the population of subjects who would never contribute under any circumstances.

Such considerations lead us to a class of model in which the event of a subject being a potential contributor, and the extent of the contribution made by that subject, are treated separately. This type of model is known as the "double hurdle" model and is originally due to Cragg (1971). As the name suggests, the model assumes that a subject must cross two hurdles in order to be a contributor. Those who fall at the first hurdle are the subjects to whom we refer as "selfish" in a dictator game context, and as "free riders" in the public goods context. Passing the first hurdle places a subject in the class of "potential contributor". Whether a potential contributor actually contributes then depends on their current circumstances; if they do contribute, we say that they have crossed the second hurdle. Both hurdles have equations associated with them, incorporating the effects of subject characteristics and treatments. Such explanatory variables may appear in both equations or only in one. Most importantly, a variable appearing in both equations may have opposite effects in the two equations.

[1] A particular example serves to illustrate this point nicely. Consider a model of criminal behaviour, in which the dependent variable is the proceeds from criminal activity by the individual. One hopes that the majority of observations on this dependent variable will be zero; hence the relevance of the hurdle framework. A useful indicator of criminal activity may be that of whether the individual is classified as "white-collar" or "blue-collar". It is reasonable to suppose that, *ceteris paribus*, white-collar individuals are *less likely* to engage in criminal behaviour than blue-collar individuals. However, it is also reasonable to suppose that white-collar criminals earn *higher profits* from their criminal activities than do blue-collar criminals. If a hurdle model were applied to this situation, we would therefore expect the white-collar dummy to have an opposite effect in the two equations.



In later sections of this chapter, it is demonstrated that estimation of the double hurdle model and its variants is possible using the ML routine available in STATA.

## 11.4 The Double Hurdle Model and Variants

### *11.4.1 p-Tobit*

The over-restrictive feature of the Tobit model highlighted in Section 11.3 is that it only allows one type of zero observation, and the implicit assumption is that zeros arise as a result of subject characteristics and/or treatments. The obvious way to relax this is to assume the existence of an additional class of subject who would never contribute under any circumstances.

In the first instance, let us simply assume that the proportion of the population who are potential contributors is $p$, so that the proportion of the population who would never contribute is $1 - p$. For the former group, the Tobit model applies, while for the latter group, the contribution is automatically zero.

This assumption leads to the $p$-tobit model, originally proposed by Deaton & Irish (1984) in the context of household consumption decisions, where they were essentially allowing for a class of "abstinent" consumers for each good modelled. The log-likelihood function for the $p$-tobit model is:

$$LogL = \sum_{i=1}^{n} \left( I_{y_i=0} \ln \left[ 1 - p\Phi \left( \frac{x_i'\beta}{\sigma} \right) \right] + I_{y_i>0} \ln \left[ p \frac{1}{\sigma} \phi \left( \frac{y_i - x_i'\beta}{\sigma} \right) \right] \right) \tag{11.8}$$

Maximising (11.8) returns an estimate of the parameter $p$, in addition to those of $\beta$ and $\sigma$ obtained under Tobit.

### *11.4.2 Double hurdle*

Since the class of subjects who would never contribute may well be the focus of the analysis, it is desirable to investigate which types of subject are most likely to appear in this class. With this in mind, we assume that the probability of a subject being in the said class depends on a set of subject characteristics. In other words, we shall generalise the $p$-Tobit model of Section 11.4.1 by allowing the parameter $p$ to vary according to subject characteristics. This generalisation leads us to the "double hurdle" model.

Subjects must cross two hurdles in order to contribute positively. The "first hurdle" needs to be crossed in order to be a potential contributor. Given that the subject is a potential contributor, their current circumstances and/or treatment in the experiment then dictate whether or not they do in fact contribute: this is the "second hurdle".

The double hurdle model contains two equations. We write:

$$
\begin{aligned}
d_i^* &= z_i'\alpha + \epsilon_i \\
y_i^{**} &= x_i'\beta + u_i \\
\begin{pmatrix}\epsilon_i \\ u_i\end{pmatrix} &= N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & \sigma^2\end{pmatrix}\right]
\end{aligned}
\tag{11.9}
$$

The diagonality of the covariance matrix in (11.9) implies that the two error terms, $\epsilon_i$ and $u_i$, are assumed to be independently distributed.

The first hurdle is then represented by:

$$
d_i = \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{if } d_i^* \leq 0 \end{cases} \tag{11.10}
$$

The second hurdle closely resembles the censoring rule for the Tobit model (11.2) :

$$
y_i^* = \max(y_i^{**}, 0) \tag{11.11}
$$

Finally, the observed variable, $y_i$, is determined as:

$$
y_i = d_i y_i^* \tag{11.12}
$$

The log-likelihood function for the double hurdle model, defined in (11.9)–(11.12), is:

$$
LogL = \sum_{i=1}^{n}\left(I_{y_i=0}\ln\left[1 - \Phi(z_i'\alpha)\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right] + I_{y_i>0}\ln \left[\Phi(z_i'\alpha)\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right]\right) \tag{11.13}
$$

Figure 11.1 is useful for understanding the model defined in (11.9)–(11.12). The concentric ellipses are contours of the joint density of the latent variables $d^*$ and $y^{**}$. These ellipses are centred on the point $(z_i'\alpha, x_i'\beta)$, so that the whole distribution moves around with changes in the values taken by the explanatory variables. The likelihood contribution associated with non-contribution (i.e. the first term in square brackets in (11.13)) is represented by the probability mass under the L-shaped region comprising the north-west, south-west, and south-east quadrants of the graph; the



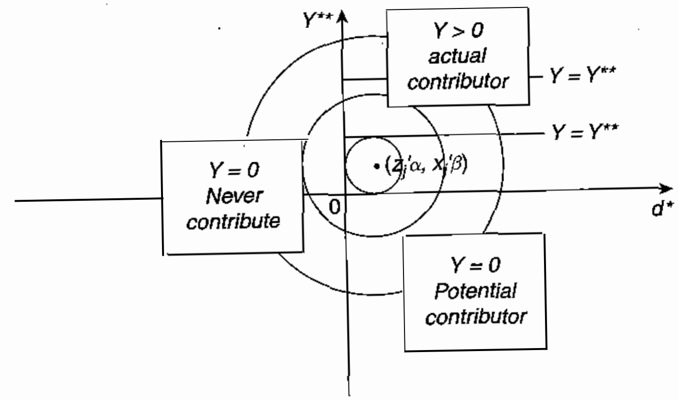



Figure 11.1: The relationship between latent ($d^*$ and $y^{**}$) and observed ($y$) variables in the double hurdle model

contribution associated with a contribution (the second bracketed term in (11.13)) is represented by a thin strip of the probability mass within the north-east quadrant at the value of the observed contribution (two such values are depicted in the diagram).

Finally, consider a double hurdle model in which there are no explanatory variables in the first hurdle equation. There is only an intercept, $\alpha_0$. The likelihood function becomes:

$$
LogL = \sum_{i=1}^{n}\left(I_{y_i=0}\ln\left[1 - \Phi(\alpha_0)\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right] + I_{y_i>0}\ln \left[\Phi(\alpha_0)\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right]\right) \tag{11.14}
$$

Note that $\Phi(\alpha_0)$ is now a scalar. If we rename this scalar $p$, we have the $p$-Tobit model defined in (11.8).

This gives us a way of estimating the $p$-Tobit model. We estimate the double hurdle model with no explanatory variables in the first hurdle. We then transform the estimate of the intercept parameter in the first hurdle, $\alpha_0$, using:

$$
p = \Phi(\alpha_0)
$$

This gives the estimate of the parameter $p$ in the $p$-Tobit model. The delta method is required in order to obtain a standard error for this estimate.

### 11.4.3 The single hurdle model

The single hurdle model is a model that has the property of "first hurdle dominance" (Jones, 1989). This essentially requires that any individual who passes the first hurdle necessarily has a positive outcome. Hence there is only one source of zeros: the zero-type; censored zeros are ruled out.

The formal definition of the single hurdle model is similar to that of the double hurdle model given in Section 11.4.2, with the only difference that the second hurdle equation (11.11) changes from a rule embodying zero censoring to one embodying zero truncation:

$$y_i^* = \begin{cases} y_i^{**} & \text{if } y_i^{**} > 0 \\ \text{unobserved} & \text{if } y_i^{**} \leq 0 \end{cases} \tag{11.15}$$

As explained in the next section, logical problems arise when we attempt to extend the single hurdle model to the panel data setting. For this reason, we do not pay close attention to this model.

## 11.5 The Panel Hurdle Model

### 11.5.1 The basic model

The models considered in Section 11.4 were all estimable with a single cross-section of data. We now progress to panel data. Here, we assume that we have $n$ subjects, each of whom participates in $T$ tasks. We denote as $y_{it}$ the decision (i.e. the contribution) of subject $i$ in task $t$. The two hurdles are defined as follows:

**First hurdle:**

$$\begin{aligned} d_i^* &= z_i'\alpha + \epsilon_{1,i} \\ d_i &= \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{if } d_i^* \leq 0 \end{cases} \\ \epsilon_{1,i} &\sim N(0, 1) \end{aligned} \tag{11.16}$$

**Second hurdle:**

$$\begin{aligned} y_{it}^{**} &= x_{it}'\beta + u_i + \epsilon_{2,it} \\ y_{it}^* &= \begin{cases} y_{it}^{**} & \text{if } y_{it}^{**} > 0 \\ 0 & \text{if } y_{it}^{**} \leq 0 \end{cases} \\ \epsilon_{2,it} &\sim N(0, \sigma^2); \quad u_i \sim N(0, \sigma_u^2) \end{aligned} \tag{11.17}$$

**Observed:**

$$y_{it} = d_i y_{it}^* \tag{11.18}$$

The central feature of this model is that, in accordance with (11.16), the first hurdle has only one outcome per subject, and, in accordance with (11.18), that outcome applies to all observations for that subject. For example, if subject $i$ falls at the first hurdle ($d_i = 0$), then all observations on $y$ for subject $i$ must be zero ($y_{it} = 0, t = 1, \dots, T$). This feature of the model is essential in order to capture the concept of the "zero-type". If a subject is genuinely a "zero-type", then, necessarily, they will contribute zero on every occasion on which they are observed.

Note also that the second hurdle (11.17) contains a subject-specific random effect term ($u_i$) that allows between-subject heterogeneity.



### 11.5.2 The panel single hurdle model

The single hurdle model was introduced in Section 11.4.3. This is a model satisfying first hurdle dominance: passing the first hurdle necessarily implies a positive outcome. In the panel setting, first hurdle dominance gives rise to a logical problem. If an individual passes the first hurdle, their outcomes would need to be positive in every time period. We already know that if the individual falls at the first hurdle, the outcome is zero at every time period. Hence first hurdle dominance rules out a mixture of zero and positive outcomes for a given individual. This is clearly a serious problem because most panel data sets would be expected to contain such mixtures. For this reason, we shall restrict attention to the framework of the panel double hurdle model introduced in Section 11.5.1, in which the zero-censoring assumed in the second hurdle allows mixtures of zeros and positive observations for a given individual.

### 11.5.3 Construction of likelihood function

Conditional on $d_i = 1$ (and also on the heterogeneity term $u_i$), we obtain something very similar to the random-effects Tobit likelihood:

$$(L_i|d_i = 1, u_i) = \prod_{t=1}^{T} \left[1 - \Phi\left(\frac{x_{it}'\beta + u_i}{\sigma}\right)\right]^{I_{y_{it}=0}} \left[\frac{1}{\sigma}\phi\left(\frac{y_{it} - x_{it}'\beta - u_i}{\sigma}\right)\right]^{I_{y_{it}>0}} \tag{11.19}$$

Conditional on $d_i = 0$, the likelihood is trivial, and depends simply on whether or not all observations are zero for subject $i$:

$$(L_i|d_i = 0) = \begin{cases} 0 & \text{if } \sum_{t=1}^{T} y_{it} > 0 \\ 1 & \text{if } \sum_{t=1}^{T} y_{it} = 0 \end{cases} \tag{11.20}$$

The likelihood (conditional on $u_i$) for subject $i$ is then obtained as a weighted average of (11.19) and (11.20), with weights given by the probabilities $P(d_i = 1)$ and $P(d_i = 0)$ which are obtained from the first hurdle equation (11.16):

$$(L_i|u_i) = \Phi(z_i'\alpha)(L_i|d_i = 1, u_i) + \left[1 - \Phi(z_i'\alpha)\right](L_i|d_i = 0) \tag{11.21}$$

Finally, the marginal likelihood for subject $i$ is obtained by integrating (11.21) over $u$:

$$L_i = \int_{-\infty}^{\infty} (L_i|u) f(u) du \tag{11.22}$$

where $f(u)$ is the $N(0, \sigma_u^2)$ density function for $u$.

The sample log-likelihood function is then given by:

$$LogL = \sum_{i=1}^{n} \ln L_i \tag{11.23}$$

### 11.5.4 Panel hurdle with upper censoring

In some applications, there is an accumulation of observations at the upper limit as well as at zero. For example, in public goods experiments, it is common for subjects to contribute their entire endowment to the public fund. The natural way to deal with this is to assume that contributions are upper censored at the amount of the endowment.

When this is assumed, (11.19) becomes:

$$(L_i|d_i = 1, u_i) = \prod_{t=1}^{T} \left[1 - \Phi\left(\frac{x'_{it}\beta + u_i}{\sigma}\right)\right]^{I_{y_{it}=0}}$$
$$\left[\frac{1}{\sigma}\phi\left(\frac{y_{it} - x'_{it}\beta - u_i}{\sigma}\right)\right]^{I_{0<y_{it}<y_{\max}}} \times \left[\Phi\left(\frac{y_{\max} - x'_{it}\beta - u_i}{\sigma}\right)\right]^{I_{y_{it}=y_{\max}}} \tag{11.24}$$

where $y_{\max}$ is the upper limit to contribution.

### 11.5.5 Panel hurdle model with tremble

A potential problem with the model defined in (11.16)–(11.18) is that it is too rigid. In particular, an individual can only be classified as falling at the first hurdle (i.e. a free-rider, or selfish type) if their observations on $y$ are zero in *every* task. We might wish for a subject who reports zero in *nearly* every task to be classifiable as a free-rider/selfish, with the rationalisation that the individual suffered a lapse of concentration on the small number of tasks for which the observation was positive.

To allow this, we introduce a tremble parameter, $\omega$. This is the probability of a lapse of concentration on any individual task. When this occurs, the distribution of $y$ is assumed to be uniform on $[0, y_{\max}]$ where $y_{\max}$ is the maximum possible value of $y$ (i.e. the endowment).



With the introduction of $\omega$, (11.19) and (11.20) become:

$$(L_i|d_i = 1, u_i) = \prod_{t=1}^{T} \left[(1-\omega)\left(1 - \Phi\left(\frac{x'_{it}\beta + u_i}{\sigma}\right)\right)\right]^{I_{y_{it}=0}} \times$$
$$\left[(1-\omega)\frac{1}{\sigma}\phi\left(\frac{y_{it} - x'_{it}\beta - u_i}{\sigma}\right) + \frac{\omega}{y_{\max}}\right]^{I_{y_{it}>0}} \tag{11.25}$$

$$(L_i|d_i = 0) = (1-\omega)^{N_i(y_{it}=0)} \left(\frac{\omega}{y_{\max}}\right)^{N_i(y_{it}>0)} \tag{11.26}$$

where $N_i(.)$ is the number of occurrences of the event contained in parentheses, for subject $i$.

The marginal likelihood is then constructed as in (11.21), but using (11.25) and (11.26) in place of (11.19) and (11.20).

### 11.5.6 Panel hurdle model with dependence

The hurdle models developed so far assume that there is no correlation between the error terms in the two hurdles. In this section, this assumption is relaxed.

Subject $i$'s idiosyncratic propensity to pass the first hurdle is represented by the error term $\epsilon_{1,i}$; their idiosyncratic propensity to contribute, conditional on passing the first hurdle, is represented by $u_i$. It is between these two terms that we introduce a correlation:

$$\text{corr}(\epsilon_1, u) = \rho \tag{11.27}$$

How is the correlation parameter $\rho$ incorporated in estimation? Let us return to the first hurdle:

$$\begin{aligned} d_i^* &= z_i'\alpha + \epsilon_{1,i} \\ d_i &= \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \\ \epsilon_{1,i} &\sim N(0,1) \end{aligned} \tag{11.28}$$

Since $\text{corr}(\epsilon_1, u) = \rho$, we may represent $\epsilon_1$ as:

$$\epsilon_1 = \rho\frac{u}{\sigma_u} + \sqrt{1-\rho^2}\,\xi \tag{11.29}$$

where $\xi \sim N(0, 1)$ and $\xi \perp u$. The requirement for passing the first hurdle becomes:

$$d_i = 1 \quad \text{if} \quad \xi > -\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-p^2}} \tag{11.30}$$

whence the probability of passing the first hurdle (conditional on $u$) becomes:

$$P(d_i = 1|u) = \Phi\left(\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-p^2}}\right) \tag{11.31}$$

In estimation, the Halton draws used to represent realisations of $u$ in the second hurdle, also appear in the probability of passing the first hurdle, in accordance with (11.31).

In the standard double hurdle model with dependence (one observation per subject), there are known to be problems identifying the correlation coefficient $\rho$ (see Smith, 2003). However, with panel data, and using the estimation approach outlined in this section, the parameter may be estimated quite precisely.

### 11.5.7 Obtaining posterior probabilities

In the discussion of finite mixture models in Chapter 8, a lot of attention was paid to the problem of computing posterior type probabilities for each subject. In the context of the hurdle model, we consider the same problem. Of course, in the hurdle model, there are only two types, the "zero-type" and the "potential contributor" type, so the problem simplifies to that of finding the posterior probability of each subject being "zero-type".

As in Chapter 8, we apply Bayes' rule in order to obtain the formula for the required posterior probability. Assuming the model with dependence (Section 11.5.6) we have:

$$P(d_1 = 0|y_{i1}...y_{iT}) = \frac{P(y_{i1}...y_{iT}|d_i = 0)P(d_i = 0)}{P(y_{i1}...y_{iT}|d_i = 0)P(d_i = 0) + P(y_{i1}...y_{iT}|d_i = 1)P(d_i = 1)} \tag{11.32}$$

The various components of the formula are obtained as integrals over $u$:

$$P(y_{i1}...y_{iT}|d_i = 0) = \int_{-\infty}^{\infty} P(y_{i1}...y_{iT}|d_i = 0, u)f(u)\mathrm{d}u \tag{11.33}$$

$$P(y_{i1}...y_{iT}|d_i = 1) = \int_{-\infty}^{\infty} P(y_{i1}...y_{iT}|d_i = 1, u)f(u)\mathrm{d}u \tag{11.34}$$

$$P(d_i = 0) = \int_{-\infty}^{\infty} P(d_i = 0|u)f(u)\mathrm{d}u \tag{11.35}$$

The form of the posterior probabilities depend crucially on whether a tremble parameter is assumed. If the tremble parameter is zero (i.e. not assumed), then $P(y_{i1}...y_{iT}|d_i = 0) = 0$ whenever at least one of the observations $y_{i1}...y_{iT}$ is positive. Hence, when this is the case, $P(d_i = 0|y_{i1}...y_{iT}) = 0$; that is, if a subject contributes on at least one occasion, they are definitely not a zero-type. If the observations $y_{i1}...y_{iT}$ are all zero, both $P(y_{i1}...y_{iT}|d_i = 0)$ and $P(y_{i1}...y_{iT}|d_i = 1)$ are positive, and hence the posterior probability of subject $i$ being zero-type is between zero and one.



When the tremble parameter is present, $P(y_{i1}...y_{iT}|d_i = 0) > 0$ even if some of the observations $y_{i1}...y_{iT}$ are positive. This is because, according to (11.26) above,

$$P(y_{i1}...y_{iT}|d_i = 0) = (1-\omega)^{N_i(y_{it}=0)}\left(\frac{\omega}{y_{\max}}\right)^{N_i(y_{it}>0)}$$

where $\omega$ is the tremble probability.

### 11.5.8 Estimation

Estimation of the panel hurdle model is performed using the method of MSL (Train, 2003). As explained fully in Chapter 10, this requires the use of Halton draws, which, when converted to normality, represent simulated realisations of the random effect term $u$. In the model with dependence, in accordance with (11.31), the simulated values also appear in the probability of passing the first hurdle. Maximisation of the simulated likelihood function is performed using the `ml` routine in STATA.

### 11.5.9 STATA code and simulation

The following code first simulates panel data (with $n = 100$ and $T = 20$) from the panel hurdle model with dependence, and then estimates the model using an MSL routine.

The data generating process is as follows:

**First hurdle:**

$$d_i^* = -1.5 + 2.0z_i + \epsilon_{1,i}$$
$$z_i \sim U(0,1); \quad \epsilon_{1,i} \sim N(0,1)$$

**Second hurdle:**

$$y_{it}^{**} = -5.0 + 10.0x_{it} + u_i + \epsilon_{2,it}$$
$$x_{it} \sim U(0,1); \quad \epsilon_{2,it} \sim N(0,5^2); \quad u_i \sim N(0,2.5^2)$$

**Censoring:**

$$y_{it}^* = \begin{cases} 0 & \text{if} \quad y_{it}^{**} < 0 \\ y_{it}^{**} & \text{if} \quad 0 \le y_{it}^{**} \le 10 \\ 10 & \text{if} \quad y_{it}^{**} > 10 \end{cases}$$

**Observed:**

$$y_{it} = d_i \times y_{it}^* \qquad i = 1, \ldots, 100; \quad t = 1, \ldots, 20$$

**Tremble:**

With probability 0.04, $y_{it} \sim \text{Uniform}(0,10)$

In the code that follows, we adopt the names for the parameters and for the other components of the log-likelihood function as shown in Table 11.1.

Two models are estimated in order to obtain starting values for the hurdle model. Starting values for the first hurdle come from a probit model estimated for the cross-section of subjects, in which the dependent variable takes the value one if the subject has at least one positive outcome, and zero otherwise. Starting values for the second hurdle come from a random effects Tobit model, with the sample restricted to subjects who have at least one positive outcome.

The code is annotated. For a more complete explanation of the way the MSL routine is programmed, and other aspects of this type of program, the reader should refer back to Chapter 10.

| Component of LogL | STATA name |
|---|---|
| $z_i'\alpha$ | za |
| $x_i'\beta$ | xb |
| $\sigma$ | s_e |
| $\sigma_u$ | s_u |
| $\rho$ | rr |
| $\omega$ | w |
| $\left[(1-\omega)\left(1-\Phi\left(\frac{x_{it}'\beta+u}{\sigma}\right)\right)\right]^{I_{y_{it}=0}} \times \left[(1-\omega)\frac{1}{\sigma}\phi\left(\frac{y_{it}-x_{it}'\beta-u}{\sigma}\right)\right]^{I_{0<y_{it}<10}} \left[\Phi\left(\frac{10-x_{it}'\beta-u}{\sigma}\right)\right]^{I_{y_{it}=10}}$ | pl |
| $(L_i \mid d_i = 1, u) = \prod_{t=1}^{T}\left[(1-\omega)\left(1-\Phi\left(\frac{x_{it}'\beta+u}{\sigma}\right)\right)\right]^{I_{y_{it}=0}} \times \left[(1-\omega)\frac{1}{\sigma}\phi\left(\frac{y_{it}-x_{it}'\beta-u}{\sigma}\right)\right]^{I_{0<y_{it}<10}} \left[\Phi\left(\frac{10-x_{it}'\beta-u}{\sigma}\right)\right]^{I_{y_{it}=10}}$ | ppl |
| $(L_i \mid d_i = 0) = \left(\frac{\omega}{10}\right)^{N_i(y_{it}>0)} (1-\omega)^{N_i(y_{it}=0)}$ | pp0 |

Table 11.1: (Continued on next page)



| | |
|---|---|
| $P(d_i = 1 \mid u) = \Phi\left(\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-\rho^2}}\right)$ | pd |
| $(L_i \mid u) = P(d_i = 1 \mid u)(L_i \mid d_i = 1, u) + [1 - P(d_i = 1 \mid u)](L_i \mid d_i = 0)$ | pp |
| $L_i = \int_{-\infty}^{\infty} (L_i \mid u) f(u) du$ | ppp |
| $LogL = \sum_{i=1}^{n} L_i$ | logl |
| $\int_{-\infty}^{\infty} P(y_{i1} \ldots y_{iT} \mid d_i = 1, u) f(u) du$ | pppl |
| $P(d_i = 1) = \int_{-\infty}^{\infty} \Phi\left(\frac{z_i'\alpha + \rho\frac{u}{\sigma_u}}{\sqrt{1-\rho^2}}\right) f(u) du$ | ppd |
| $P(d_i = 0 \mid y_{i1} \ldots y_{iT})$ | pd0 |

Table 11.1: Components of LogL and corresponding STATA names

```
* SET RANDOM NUMBER SEED

set seed 971156

* SET SAMPLE SIZE; GENERATE SUBJECT NUMBER (i) AND TASK NUMBER (t)

set obs 2000

egen int i=seq(), f(1) b(20)
egen int t=seq(), f(1) t(20)

summ i
scalar N=r(max)
summ t
scalar T=r(max)

xtset i t

* SET PARAMETER VALUES

scalar a1=2.0
scalar a0=-1.5
scalar b1=10.0
scalar b0=-5.0
scalar sig_u=2.5
scalar sig_e=5.0
```

```
scalar rho=0.40
scalar w=0.04
scalar y_max=10

* GENERATE VARIABLES

gen double x=uniform()
gen double e2=invnorm(uniform())

by i: generate double u=(invnorm(uniform())) if _n==1
by i: replace u=u[1] if u==.

by i: generate double z=uniform() if _n==1
by i: generate double e1=rho*u+sqrt(1-rho^2)*invnorm(uniform()) if _n==1

generate double ds=a0+a1*z+e1
generate int d=ds>0

by i: replace z=z[1]
by i: replace d=d[1]

gen double yss=b0+b1*x+sig_u*u+sig_e*e2
gen double ys=yss*(yss>0)
gen double y=ys*d

replace y=y_max if y>y_max

gen int tremble=uniform()<w
replace y=y_max*uniform() if tremble==1

* DROP ALL LATENT VARIABLES

drop e1 e2 u ds d yss ys tremble

* GENERATE INDICATORS FOR FIRST AND LAST OBSERVATION FOR EACH SUBJECT

by i: generate int first=1 if _n==1
by i: generate int last=1 if _n==_N

* GENERATE HALTON DRAWS IN FIRST ROW OF EACH SUBJECT
* STORE NUMBER OF DRAWS AS "DRAWS"

mat p=[3]
mdraws if first==1 , neq(1) dr(31) prefix(h) primes(p) burn(3)
scalar draws=r(n_draws)

* COLLECT HALTON DRAWS IN VARIABLE LIST "hlist"

local hlist h1*
recast double h1*

* FOR EACH SUBJECT, COPY ROW OF HALTON DRAWS INTO EVERY ROW;
*  CONVERT TO STANDARD NORMAL

quietly{

foreach v of varlist ‘hlist’ {
by i: replace ‘v’=‘v’[1] if ‘v’==.
replace ‘v’=invnorm(‘v’)
}
}

* LIKELIHOOD EVALUATION PROGRAM ("PANEL_HURDLE") STARTS HERE:

program drop _all
program define panel_hurdle

* SPECIFY ARGUMENTS

args todo b logl
tempvar za xb pp0 p1 pp1 ppp1 pp ppp pd ppd
tempname s_u s_e rr w
local hlist h1*
```



```
* EXTRACT ELEMENTS OF PARAMETER VECTOR b

mleval ‘za’ = ‘b’, eq(1)
mleval ‘xb’ = ‘b’, eq(2)
mleval ‘s_u’ = ‘b’, eq(3) scalar
mleval ‘s_e’ = ‘b’, eq(4) scalar
mleval ‘rr’ = ‘b’, eq(5) scalar
mleval ‘w’=‘b’, eq(6) scalar

* INITIALISE TEMPORARY VARIABLES

quietly gen double ‘p1’=.
quietly gen double ‘pp1’=.
quietly gen double ‘ppp1’=0
quietly gen double ‘pp0’=.
quietly gen double ‘pp’=.
quietly gen double ‘ppp’=0
quietly gen double ‘pd’=.
quietly gen double ‘ppd’=0

* LOOP FOR EVALUATION OF SUM (OVER r) OF PRODUCT (OVER t)
* ppp IS FOR LIKELIHOOD FUNCTION; ppd IS FOR p(d=1) (USED TO COMPUTE
* POSTERIOR PROBABILITY OF d=0)

quietly{

foreach v of varlist ‘hlist’ {
replace  ‘p1’= (1-‘w’)*(1/‘s_e’)*normalden((y-(‘xb’ + ‘s_u’*‘v’)) ///
/‘s_e’)+‘w’/y_max if (y>0)&(y<y_max)
replace  ‘p1’=(1-‘w’)*(1-normal((‘xb’ + ‘s_u’*‘v’)/‘s_e’)) if y==0
replace ‘p1’=(1-‘w’)*(1-normal((y_max-(‘xb’ + ‘s_u’*‘v’))/‘s_e’)) ///
if y==y_max
by i: replace ‘pp1’ = exp(sum(ln(‘p1’)))
replace ‘pp1’=. if last~=1
replace ‘ppp1’=‘ppp1’+‘pp1’

replace ‘pd’=normal((‘za’+‘rr’*‘v’)/sqrt(1-‘rr’^2))
replace ‘pd’=. if last~=1

replace ‘pp0’= ((‘w’/y_max)^n_pos)*((1-‘w’)^n_zero)
replace ‘pp0’=. if last~=1

replace ‘pp’=‘pd’*‘pp1’+(1-‘pd’)*‘pp0’
replace ‘ppp’=‘ppp’+‘pp’

replace ‘ppd’=‘ppd’+‘pd’
}
* END OF LOOP

* DIVISION BY R TO GENERATE REQUIRED AVERAGES (OVER r)
* COMPUTE POSTERIOR PROBABILITY OF d=0 (pd0) AND SEND THIS TO MATA

replace ‘ppp’=‘ppp’/draws

replace ‘ppp1’=‘ppp1’/draws
replace ‘ppd’=‘ppd’/draws

replace pd0=‘pp0’*(1-‘ppd’)/(‘pp0’*(1-‘ppd’)+‘ppp1’*‘ppd’)
putmata pd0, replace

* MLSUM COMMAND TO SPECIFY PER-SUBJECT LOG-LIKELIHOOD CONTRIBUTION

mlsum ‘logl’=ln(‘ppp’) if last==1

}
end

* "end" SIGNIFIES END OF LIKELIHOOD EVALUATION PROGRAM "panel_hurdle"

* GENERATE BINARY d INDICATING AT LEAST ONE POSITIVE CONTRIBUTION BY SUBJECT;
* GENERATE NUMBER OF POSITIVE CONTRIBUTIONS BY SUBJECT (n_pos) AND NUMBER OF
* ZERO CONTRIBUTIONS (n_zero)
```

```
quietly{

by i: gen int sum_y=sum(y)
by i: replace sum_y=sum_y[_N]

gen int d=sum_y>0

gen int y_pos=y>0

by i: gen int n_pos = sum(y_pos)
by i: replace n_pos = n_pos[_N]

gen int y_zero=y==0
by i: gen int n_zero = sum(y_zero)
by i: replace n_zero = n_zero[_N]

}

* INITIALISE VARIABLE TO REPRESENT POSTERIOR PROBABILITY OF
* FALLING AT FIRST HURDLE (pd0)

quietly gen double pd0=.

* USE PROBIT (1 OBS PER SUBJECT) TO OBTAIN FIRST HURDLE STARTING VALUES

probit d x   if last==1
mat bprobit=e(b)

* USE RANDOM EFFECTS TOBIT TO OBTAIN SECOND HURDLE STARTING VALUES

xttobit y  x if d==1 , ll(0) ul(10)
mat bxttobit=e(b)

* DEFINE VECTOR OF STARTING VALUES (INCLUDING GUESSES FOR rho AND w)

mat start=bprobit, bxttobit, 0.00, 0.02

* SPECIFY EVALUATOR (d0), EVALUATION PROGRAM (panel_hurdle), AND PARAMETER LIST
* SPECIFY STARTING VALUE VECTOR
* RUN MAXIMUM LIKELIHOOD PROCEDURE

ml model d0 panel_hurdle ( = z ) (  = x ) /s_u /s_e /rr /w
ml init start, copy
ml max, trace search(norescale)

* EXTRACT POSTERIOR PROBABILITY OF d=0 (pd0) GENERATED INSIDE EVALUATION PROGRAM
* PLOT IT AGAINST NUMBER OF POSITIVE CONTRIBUTIONS.

drop pd0
getmata pd0

label variable pd0 "posterior prob zero type"
label variable n_pos "number of positive contributions"

scatter pd0 n_pos, jitter(1)  ylabel(0(0.1)1)
```

The final output from running the above simulation is shown below. These are the results from the panel hurdle model.

```
                                                Number of obs   =       2000
                                                Wald chi2(1)    =      17.79
Log likelihood = -1518.8587                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
           z |   2.369987   .5619582     4.22   0.000     1.268569    3.471405
       _cons |  -1.747444   .3340628    -5.23   0.000    -2.402195   -1.092693
-------------+----------------------------------------------------------------
eq2          |
           x |   11.06413   .8861659    12.49   0.000     9.327278    12.80098
```



```
       _cons |  -4.672788   1.331671    -3.51   0.000    -7.282815   -2.062762
-------------+----------------------------------------------------------------
s_u          |
       _cons |   2.467505   .5636856     4.38   0.000     1.362701    3.572308
-------------+----------------------------------------------------------------
s_e          |
       _cons |   4.770671   .2305484    20.69   0.000     4.318805    5.222538
-------------+----------------------------------------------------------------
rr           |
       _cons |   .1441294   .5070172     0.28   0.776    -.8496061    1.137865
-------------+----------------------------------------------------------------
w            |
       _cons |   .0353052   .0048938     7.21   0.000     .0257135    .0448968
------------------------------------------------------------------------------
```

The estimates are collected in Table 11.2. We include 95% confidence intervals for each parameter. The second column of the table contains the true parameter values being used in the simulation. That all true parameter values are contained within their respective confidence intervals is confirmation of the correctness of the estimation program presented above.

In Figure 11.2, we plot the posterior probability of falling at the first hurdle (i.e. of being "zero-type") against the number of positive contributions, for each of the 100 subjects in the simulation. We see that subjects with less than five positive contributions (out of 20) are all classified as very likely to be "zero types". Recall that it is possible to classify occasionally contributing subjects as zero-types by virtue of the presence of the tremble parameter, $\omega$, whose true value is 0.04. These occasional positive contributions are attributable to lapses of concentration. For subjects contributing on five or more occasions, the posterior probability of being a zero-type is very close to zero.

## 11.6 A Panel Hurdle Model of Dictator Game Giving

### *11.6.1 The experiment*

In this section, the panel hurdle model developed in Section 11.5 is applied to a real data set. The data is from a dictator game experiment conducted by Erkal

| Parameter | True value | 95% CI lower | MLE | 95% CI upper |
|---|---|---|---|---|
| $\alpha_1$ | 2.0 | 1.27 | 2.37 | 3.47 |
| $\alpha_0$ | −1.5 | −2.40 | −1.75 | −1.09 |
| $\beta_1$ | 10.0 | 9.33 | 11.06 | 12.80 |
| $\beta_0$ | −5.0 | −7.28 | −4.67 | −2.06 |
| $\sigma_u$ | 2.5 | 1.36 | 2.47 | 3.57 |
| $\sigma_\epsilon$ | 5 | 4.32 | 4.77 | 5.22 |
| $\rho$ | +0.4 | −0.85 | +0.14 | 1.14 |
| $\omega$ | 0.04 | 0.026 | 0.035 | 0.045 |

Table 11.2: Point and interval estimates from simulation of panel hurdle model, alongside true parameter values

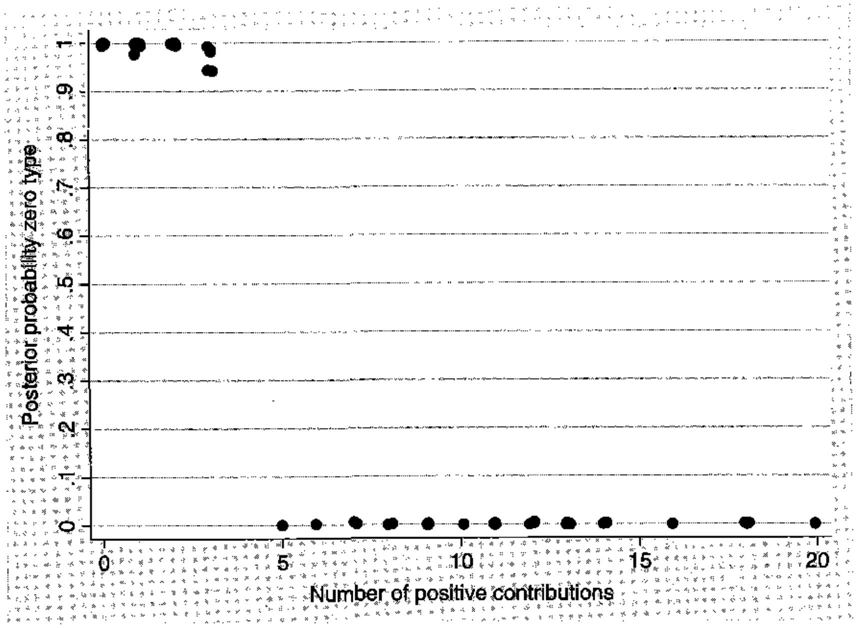



Figure 11.2: Posterior zero-type probability against number of positive contributions

et al. (2011), whose principal objective was to investigate the relationship between income and giving. The data is contained in the file **Erkal**.

Subjects are arranged into groups of four. The experiment consists of two stages. In the first stage, subjects carry out a real effort task, and are ranked according to their performance in this task. In treatments 1 and 2, subjects receive payment in accordance with their ranking: the subject ranked first receives \$60; second receives \$45; third receives \$30; and fourth receives \$15. Then stage 2 follows, in which, with knowledge of their own and others' income in stage 1, subjects decide how much to transfer to each of the other three. Subjects are made aware that only one of their chosen transfers will be made, and this will be determined randomly.

There are 27 groups of four subjects for whom the experiment is played out in the manner described above. Firstly note that it is natural to treat the resulting data set as a panel because there is a sample of 108 dictators, each of whom is observed making three different transfer decisions (only one of which will ultimately be realised).

Secondly, note that of the 324 transfer decisions, 264 (81%) were zeros. The distribution of transfers is shown in Figure 11.3. The predominance of zeros in the data clearly calls for the hurdle approach.

The issue to be addressed in estimation is what determines the transfer. Clearly, the key determinants are expected to be own income and recipient's income. Own income is a characteristic of the subject, so it may enter the first hurdle. In contrast, recipient's income is a characteristic of the task, and therefore can only enter the second hurdle.



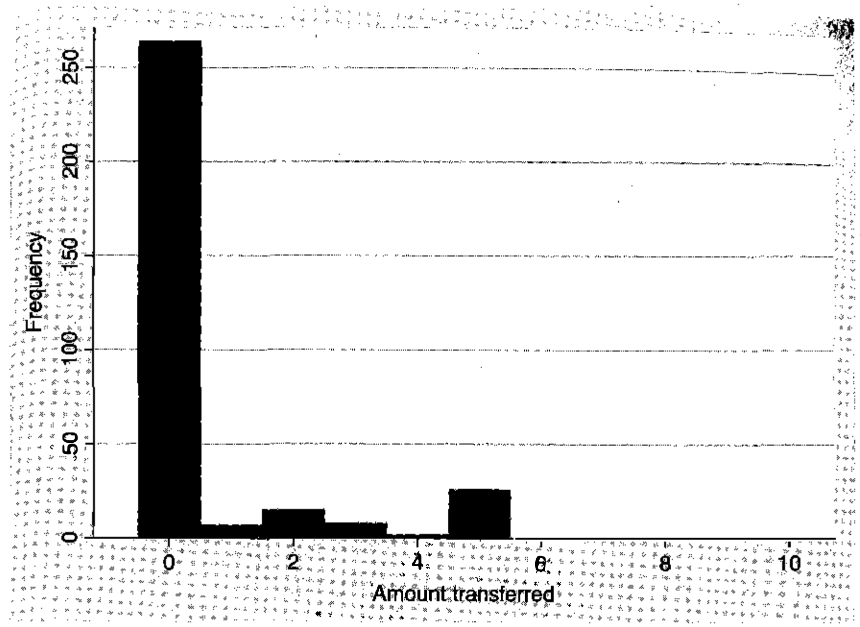



Figure 11.3: A frequency histogram of amount transferred in Erkal et al. (2011) treatments 1 and 2 (324 observations)

| Dictator's rank | Earnings | Number of dictators | % giving |
|---|---|---|---|
| 1 | \$60 | 27 | 26% |
| 2 | \$45 | 27 | 67% |
| 3 | \$30 | 27 | 37% |
| 4 | \$15 | 27 | 22% |
| ALL | | 108 | 38% |

Table 11.3: Proportion of dictators giving positive amounts (at least once) by own-earnings

To give a feel for the data, in Table 11.3 we tabulate the proportion of dictators giving positive amounts at least once. The key feature of Table 11.3 is that the dictators who are ranked first and earn \$60 appear to be less likely to give than those who are ranked second and earn \$45. In fact, the proportion of \$60 earners who give is comparable to that of \$15 earners. Explaining why high earners are less likely to give is in fact the principal focus of the paper by Erkal et al. (2011).

It is also interesting to consider to whom dictators give the most. In Table 11.4, we show the mean transfer being given to receivers of each rank. We see that, although the amounts given are always small, the amount given appears to rise steeply as receivers' income falls. Subjects earning the lowest amount are given ten times as much, on average, as those earning the highest amount. Moreover, this relationship appears to be monotonic, unlike the one uncovered in Table 11.3.

| Receiver's rank | Earnings | Number of receivers | Mean of amount received ($) |
|---|---|---|---|
| 1 | $60 | 81 | 0.16 |
| 2 | $45 | 81 | 0.20 |
| 3 | $30 | 81 | 0.70 |
| 4 | $15 | 81 | 1.60 |
| ALL | | 324 | 0.67 |

Table 11.4: Mean transfer to receivers by earnings of receiver

### *11.6.2 Estimation*

The panel hurdle model (with dependence) developed in Section 11.5 has been applied to the data set of Erkal et al. (2011). The code used is very similar to that used for the simulation in Section 11.5.

Some words of advice are in order for anyone setting out to apply this sort of estimation routine to a real data set such as this one. The quality of starting values is important. If the starting values are too far from the MLE, maximisation is unlikely to be achieved. A preliminary set of starting values is obtained by estimating a two-part model; that is, a probit model followed by random effects probit on a restricted sample. However, these starting values are not always sufficiently accurate to allow estimation of the hurdle model in a single run. It is often necessary to maximise the log-likelihood in stages, starting with a heavily constrained model and gradually removing the constraints. Below is the code that has been applied successfully to the present data set. The likelihood maximisation program "panel_hurdle" is the same as the one used in Section 11.5.

A tremble term was found to be insignificantly different from zero, so was excluded from the final model. The final model is one which constrains the tremble parameter to zero, but for which all other parameters are free.

```
* CREATE VARIABLE LISTS

local listd "male  myrank22 myrank23 myrank24  "
local listy "male   yourrank22 yourrank23 yourrank24    "

* USE ONLY TREATMENTS 1 AND 2

keep if treat<3

* OBTAIN STARTING VALUES USING PROBIT AND XTTOBIT.
* STORE VARIANCE ESTIMATES FROM XTTOBIT

probit d   `listd'      if last==1
mat bprobit=e(b)

xttobit y  `listy'  if  d==1, ll(0)
mat bxttobit=e(b)

scalar sig_u=_b[sigma_u:_cons]
scalar sig_e=_b[sigma_e:_cons]

* CREATE STARTING VALUE VECTOR

mat start=bprobit, bxttobit,0,0

* SPECIFY CONSTRAINTS ON PARAMETERS
```



```
constraint 3 [sig_u]_b[_cons]=sig_u
constraint 4 [sig_e]_b[_cons]=sig_e
constraint 5 [r]_b[_cons]=0.0
constraint 6 [w]_b[_cons]=0

* ESTIMATE PANEL HURDLE MODEL WITH sig_u, sig_e, r, w ALL CONSTRAINED

ml model d0 panel_hurdle (= `listd') ( = `listy' ) /sig_u /sig_e  /r /w, ///
 constraints( 3 4 5 6)
ml init start, copy
ml max, trace  search(norescale)

* STORE ESTIMATES AS STARTING VECTOR FOR NEXT ESTIMATION

mat start=e(b)

* ESTIMATE PANEL HURDLE MODEL WITH r AND w CONSTRAINED

ml model d0 panel_hurdle (= `listd' ) ( = `listy') /sig_u /sig_e  /r /w, ///
 constraints( 5 6)
ml init start, copy
ml max, trace  search(norescale)

* STORE ESTIMATES AS STARTING VECTOR FOR NEXT ESTIMATION

mat start=e(b)

* ESTIMATE PANEL HURDLE MODEL WITH ONLY w CONSTRAINED (SET STARTING VALUE FOR r)

mat start[1,`=colnumb(start, "r:_cons")']=-0.3

ml model d0 panel_hurdle (= `listd' ) ( = `listy') /sig_u /sig_e  /r /w, ///
  constraints(  6)
ml init start, copy
ml max, trace  search(norescale)
```

### *11.6.3 Results*

The results are presented in Table 11.5. The first and second columns show respectively the results from the probit model and the panel Tobit model that are used to obtain starting values for the first and second hurdle parameters of the panel hurdle model. The contents of the first two columns may be treated as estimates from a "two-part model". The final two columns contain estimates from two panel hurdle models, one with independence and one with dependence.

Results from the two panel hurdle models are very similar. This is because the correlation coefficient ($\rho$) is not estimated as significantly different from zero, and hence its free estimation does not significantly improve the model's explanatory power. Both models confirm the main result from Erkal et al.'s (2011) paper: that those ranked second are significantly more likely to give than those ranked first. This is seen from the significantly positive coefficient of "myrank2" in the first hurdle equation and from noting that "myrank1" is the excluded dummy. The second hurdle estimates are not surprising: the estimates of the included "yourrank" dummies are all positive (with "yourrank1" being the base case) and rise progressively as rank falls, implying that, conditional on a dictator being a "giving type", they are prepared to give more to receivers who earn less.

A gender dummy has also been included in both hurdles. Here we see that males are significantly less likely to be "giving types", but a male who is a giving type is

| | Probit | Panel Tobit (excluding all-zero subjects) | Panel hurdle ($\rho$=0) | Panel hurdle with dependence |
|---|---|---|---|---|
| **First hurdle** | | | | |
| Male | −0.36(0.30) | | −0.58*(0.35) | −0.58*(0.36) |
| Myrank2 | 1.03**(0.36) | | 1.03**(0.42) | 0.99**(0.41) |
| Myrank3 | 0.20(0.37) | | 0.24(0.42) | 0.27(0.40) |
| Myrank4 | −0.25(0.39) | | −0.07(0.49) | 0.23(0.51) |
| Constant | −0.38(0.33) | | −0.09(0.41) | −0.09(0.41) |
| **Second hurdle** | | | | |
| Male | | 1.22*(0.71) | 1.56(0.98) | 1.89*(1.15) |
| Yourrank2 | | 0.59(0.93) | 0.59(0.99) | 0.61(0.99) |
| Yourrank3 | | 3.90**(0.80) | 3.97**(0.86) | 3.97**(0.86) |
| Yourrank4 | | 6.68**(0.82) | 6.88**(0.89) | 6.94**(0.90) |
| Constant | | −3.33(0.84) | −4.35(1.12) | −4.14(1.15) |
| $\sigma_u$ | | 1.71 | 2.39 | 2.42(0.53) |
| $\sigma_\epsilon$ | | 2.11 | 2.26 | 2.25(0.29) |
| $\rho$ | | | | −0.30(0.40) |
| $n$ | 108 | 41 | 108 | 108 |
| $T$ | | 3 | 3 | 3 |
| LogL | −63.83 | −170.40 | −228.82 | −228.61 |

Table 11.5: Maximum likelihood estimates from hurdle and related models applied to Erkal et al. (2011) data, Treatments 1 and 2
$* \ p < 0.05$; $** \ p < 0.01$.

predicted to give significantly (in the model with dependence) more than a female who is a giving type. This contradiction in the effect of gender has already been noted in the context of dictator game giving by Andreoni & Vesterlund (2001). It is an important result because it can easily be seen that these effects are hidden when simple tests are used. For example, consider a Mann-Whitney test of the effect of gender on amount transferred.

```
. ranksum y, by(male)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

         male |      obs    rank sum    expected
-------------+---------------------------------
            0 |      150     24656.5       24375
            1 |      174     27993.5       28275
-------------+---------------------------------
     combined |      324       52650       52650

unadjusted variance    706875.00
adjustment for ties   -382851.63
                      ----------
adjusted variance      324023.37

Ho: y(male==0) = y(male==1)
             z =   0.495
    Prob > |z| =   0.6209
```



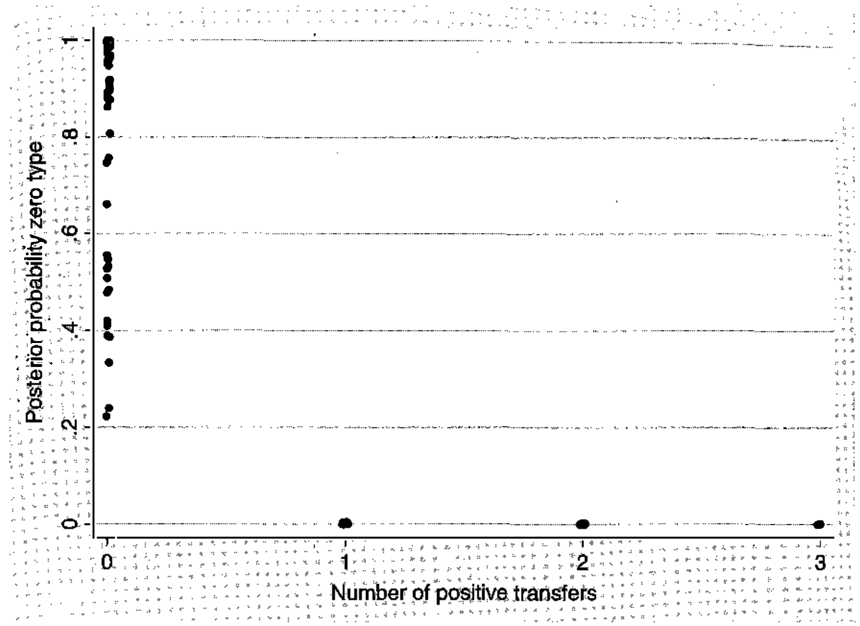



Figure 11.4: Posterior probability of "zero type" against number of positive transfers

The p-value of 0.62 indicates that, according to this simple test, there is no gender effect. However, this is simply because the lower likelihood of giving by males is being compensated by the higher amount given by males who do give. This sort of effect is a clear example of the usefulness of the hurdle approach, which successfully separates the contradictory effects, resulting in the significant estimates we see in the final column of Table 11.5.

Having estimated the panel hurdle model with dependence, we may find the posterior probabilities of each subject being of "zero type", using the method described in Section 11.5.7. In Figure 11.4, we present a jittered plot of these posterior probabilities against the subject's number of positive transfers. Of the 108 subjects, 41 transfer a positive amount at least once, and, since the model does not include a tremble parameter, these 41 subjects all have zero posterior probability of being zero type. For the 67 subjects who always transfer zero, the posterior probability of being zero type is positive, but not always close to 1; for some it is less than 0.5.

As mentioned above, the estimation reported in Table 11.5 confirms the main result of Erkal et al. (2011), that the highest earning dictators are less likely to give than moderately earning dictators. The question to which this result gives rise is whether this is an income effect *per se* (i.e. giving is an "inferior good") or whether it is because self-interested individuals tend to work harder, and hence to self-select to rank1. In order to address this important question, Erkal et al. (2011) conduct a Treatment 3 (the "luck" treatment), in which earnings of the group members are determined randomly, rather than being determined by performance in any task.

| | Probit | Panel Tobit (excluding all-zero subjects) | Panel hurdle ($\rho$=0) |
|---|---|---|---|
| **First hurdle** | | | |
| Male | −0.31(0.35) | | −0.47(0.44) |
| Myrank2 | −0.07(0.50) | | −0.04(0.56) |
| Myrank3 | −0.60(0.49) | | −0.54(0.54) |
| Myrank4 | −0.93*(0.50) | | −0.65(0.60) |
| Constant | 0.55(0.41) | | 0.73(0.48) |
| **Second hurdle** | | | |
| Male | | 1.45(1.17) | 1.69(1.50) |
| Yourrank2 | | 1.84(2.10) | 1.74(2.17) |
| Yourrank3 | | 6.83**(1.98) | 6.74**(2.04) |
| Yourrank4 | | 11.29**(1.95) | 11.35**(2.03) |
| Constant | | −5.88(1.86) | −6.71(2.06) |
| $\sigma_u$ | | 1.25(1.21) | 1.91(1.25) |
| $\sigma_\epsilon$ | | 4.11(0.58) | 4.26(0.61) |
| $\rho$ | | | |
| $n$ | 56 | 28 | 56 |
| $T$ | | 3 | 3 |
| LogL | −35.80 | −145.44 | −178.82 |

Table 11.6: Maximum likelihood estimates from hurdle and related models applied to Erkal et al. (2011) data, Treatment 3
* $p < 0.05$; ** $p < 0.01$.

In Treatment 3, there are 14 groups of four subjects. Hence there are 56 dictators, each, as before, making three decisions. Once again we estimate the panel hurdle model. For this data set, the dependence model is not successfully estimated, perhaps due to the smaller sample. However, results from the model with independence are obtained, and are presented in Table 11.6.

In the final column of Table 11.6, we see one very important difference from the corresponding set of estimates in Table 11.5. It is no longer the case that those ranked first are less likely to give than those ranked second – none of the earnings dummies in the first hurdle show statistical significance. This leads us to the conclusion that when a dictator's income is determined randomly, the level of this income has no effect on the propensity to give. It is also confirmed that the reason for higher (earned) income individuals being less likely to give (as seen in Table 11.5) is a consequence of self-interested individuals self-selecting into the high income group.

## 11.7 A Panel Hurdle Model of Contribution in a Public Goods Game

In this section, we apply the hurdle model to a data set from a public goods experiment first analysed by Clark (2002) and subsequently by Harrison (2007).



The focus of this study is to test the house money effect. The data is contained in the file **clark**.

The concept of the public goods experiment was explained fully in Section 2.5.4.

In Clark's (2002) experiment, there are ten rounds, and subjects are in groups of five, with the make-up of groups changing between rounds. In each round, the endowment is 80 tokens, and subjects are required to allocate between an individual account and a group account. There are 150 subjects, of whom 75 are in the "house money" treatment (given 80 tokens at the start of each round), and 75 are in the "own money" treatment (nothing given to them).[2]

A number of hypotheses relating to the effect of house money in this context are proposed by Clark (2002). Firstly, house money is used by subjects to purchase more "public-spiritedness" than they would show in real life; that is, they are more likely to behave altruistically. Secondly, house money is used by subjects to purchase reciprocal fairness; that is, they become more likely to behave as reciprocators. Thirdly, house money makes subjects more risk-seeking, and consequently more likely to experiment with "strategic giving"; that is, they are more likely to behave as strategists.

Note that some of the types alluded to in the last paragraph are the same types as assumed in the mixture model of public goods analysed in Section 8.5. Note also that all three of Clark's hypotheses embody the idea that the effect of house money is to move subjects from the "free-rider" type (i.e. the "zero-type") to one of the contributing types. Hence in the context of the hurdle model, in order to capture these hypotheses, house money should be an explanatory variable appearing in the first hurdle.

Previous findings from the analysis of Clark's data set are as follows. Clark (2002) is concerned about subjects' influence on others' behaviour within sessions, and conducts "ultra-conservative" tests[3] using session averages. Unfortunately there are only five sessions for each treatment, so the tests are based on only ten observations. Unsurprisingly with such a small sample, these tests detected no difference between the two treatments, and therefore no evidence of a house money effect. Harrison (2007) examines the same data set at the individual level, and finds a house money effect. In particular, he finds that house money impacts on the proportion of free-riders.

Figure 11.5 shows that, both with and without house money, the data is left- and (less intensely) right-censored. The key difference between the two histograms is that participants appear to be more likely to contribute zero if they are using their own money.

We can conduct a simple test of the impact of treatment on free-riding. Of the 150 subjects, 20 *never* contribute. At a simple level, we may classify these 20 subjects as free-riders. Of the 20, 13 (out of 75) are in the own money treatment, and 7 (out of 75) are in the house money treatment. We may use the

[2] Clark finds a subtle means of administering the "own money" treatment by which the possibility of a subject experiencing a net loss is avoided. See Clark (2002) for further details.

[3] See Section 4.4.3 for a discussion of ultra-conservative tests.

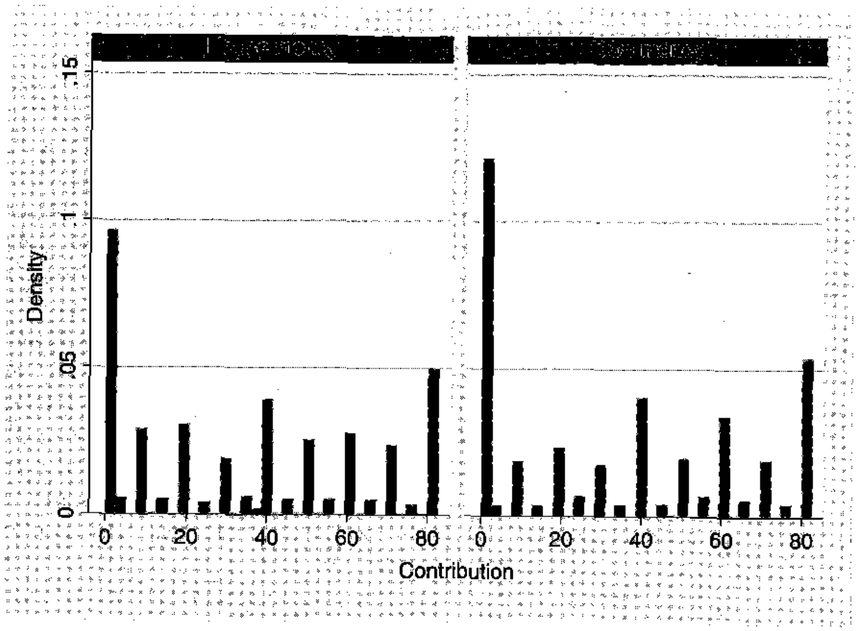



Figure 11.5: Histogram of contributions by treatment

tabi command in order to obtain the required test statistics directly from this information:

```
. tabi 62 68 \ 13 7, exact chi2 col
+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |          col
       row |         1          2 |     Total
-----------+----------------------+----------
         1 |        62         68 |       130
           |     82.67      90.67 |     86.67
-----------+----------------------+----------
         2 |        13          7 |        20
           |     17.33       9.33 |     13.33
-----------+----------------------+----------
     Total |        75         75 |       150
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =   2.0769   Pr = 0.150
           Fisher's exact =                 0.229
   1-sided Fisher's exact =                 0.115
```

Although the proportion of free-riders is certainly lower in the house money treatment (9.33%) than in the own money treatment (17.33%), neither the chi-squared test nor the Fisher's exact test are returning significant test statistics. However, it may be that these tests are invalid, if, as Clark supposed, there is



dependence in behaviour between subjects in the same session. Of course, this is a compelling reason for adopting the panel hurdle framework, in which between-subject heterogeneity is allowed for, and behaviour of others within the same session is controlled for. To the panel hurdle model we now turn.

Two more plots are useful. These are contributions against two of the explanatory variables that will appear in the second hurdle. Figure 11.6 shows plots of contribution against round number, and contribution against mean of others' contribution in previous rounds, each with a Lowess smoother. The first smoother reveals that contributions slightly decline as the experiment progresses. Hence the inclusion of round number in the second hurdle is justified. Regarding the second smoother, an important feature of the experiment is that in each round subjects are made aware of the group mean in the previous round. It is clear from the smoother that participants are sensitive to this measure of recent experience, justifying its inclusion as an explanatory variable in the second hurdle. Moreover, it appears that this effect levels off as previous contributions increase. This leads us to use the square of previous contributions, in addition to the variable itself, as an explanatory variable in the second hurdle.

When using the lagged mean contribution of other group members as an explanatory variable, a slight problem that arises is that this variable contains missing values in the first round. Contributions from the first round could be

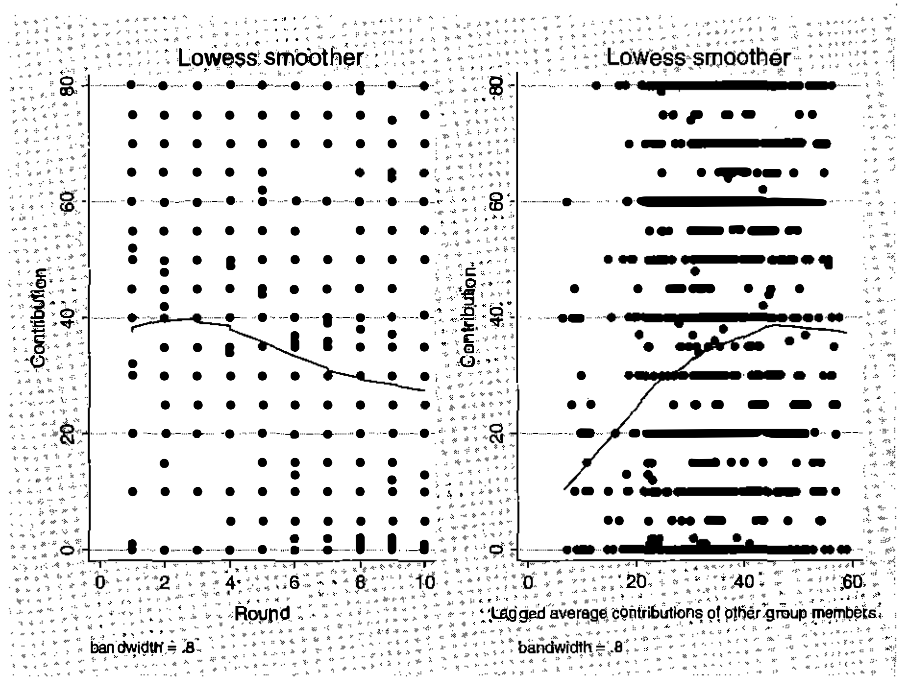



Figure 11.6: Plot of contribution against round number; plot of contribution against mean of others' contribution in previous round

excluded from estimation. However, rather than doing this, we assume that, in the first round, subjects make a judgement (perhaps based on lifetime experience prior to the experiment) of what a typical contribution of others is likely to be.[4] This value is then used as the value of the lagged mean in the first round. The value we use for this purpose is, in fact, 66 units, and this is obtained, via a grid search, as the value which maximises the sample log-likelihood. This relatively high value implies that subjects tend to start the experiment optimistically. While this assumption may be viewed as somewhat arbitrary, the main result of the paper (relating to the house money effect) is in fact robust to the choice of value made here.

Results from various models are shown in Table 11.7. The first is a binary probit model estimated using the cross-section of 150 subjects, with the dependent variable taking the value 1 if at least one of the subject's contributions is positive, and zero otherwise. This model gives estimates of the first hurdle, which are used as starting values in the estimation of the corresponding part of the hurdle model. The second model is a panel Tobit model estimated using only those subjects with at least one positive contribution, which gives estimates of the second hurdle; these are also used as starting values. As previously, the contents of these first two columns may be treated as estimates from a "two-part model". The remaining three columns

| | Probit | Panel Tobit | Hurdle ($\rho = 0$; $\omega = 0$) | Hurdle dep. no tremble | Hurdle dep. with tremble |
|---|---|---|---|---|---|
| **First hurdle** | | | | | |
| Const | 0.94(0.17) | | 0.97(0.18) | 0.98(0.18) | 0.69(0.15) |
| House money | 0.38(0.26) | | 0.39(0.28) | 0.34(0.28) | 0.50*(0.25) |
| **Second hurdle** | | | | | |
| Const | | 42.46(10.04) | 42.61(10.39) | 44.75(10.63) | 30.05(9.67) |
| Round | | −3.12**(0.40) | −3.09**(0.40) | −3.08**(0.40) | −2.46**(0.39) |
| Previous | | 0.95*(0.39) | 0.95*(0.40) | 0.95*(0.40) | 0.95**(0.37) |
| Previous$^2$ | | −0.01**(0.00) | −0.01**(0.00) | −0.01**(0.00) | −0.01**(0.00) |
| $\sigma$ | | 26.07(0.69) | 26.09(0.69) | 26.08(0.69) | 21.06(0.85) |
| $\sigma_u$ | | 28.65(2.10) | 29.99(2.38) | 30.04(2.30) | 35.60(3.40) |
| $\rho$ | | | 0 | −0.29(0.35) | 0.93(0.08) |
| $\omega$ | | | 0 | 0 | 0.09(0.02) |
| $n$ | 150 | 130 | 150 | 150 | 150 |
| $T$ | — | 10 | 10 | 10 | 10 |
| LogL | −57.85 | −4649.02 | −4707.01 | −4706.83 | −4679.63 |
| AIC | | | 9430.02 | 9431.66 | 9379.26 |

Table 11.7: Results of hurdle and related models applied to Clark's (2002) public goods data
* significant ($p < 0.05$); ** strongly significant ($p < 0.01$).
Note: AIC is defined as $(2k - 2\text{LogL})$.

[4] This method is similar to that of Bardsley & Moffatt (2007, see their Table II, note 4).



contain estimates from three panel hurdle models: the basic model; the model with dependence; the model with dependence and tremble. The last of these is the best model on the basis of Akaike's Information Criterion (AIC).

The effects seen in the second hurdle are as expected. Firstly, in agreement with other public good studies, and with Figure 11.6 (left pane), the negative coefficient of the round number provides evidence of a decline in contributions over the course of the experiment. Secondly, the positive coefficient of the mean previous contribution by others, and the negative coefficient of its square, confirm the conclusion drawn from Figure 11.6 (right pane), that contributions rise but level off with others' contributions. This amounts to evidence of reciprocity.

The key result is the evidence of the house money effect in the final column: house money increases the probability of passing the first hurdle, that is, it reduces the probability of being a "free-rider". Note that this result is only seen as significant in the final column.

The role of the tremble parameter is crucial in reaching this conclusion. In order to see why, in Figure 11.7 we plot the posterior probability of each subject being a free-rider (obtained using Bayes' rule, as explained in Section 11.5.7) against the number of positive contributions by the subject. The left-hand plot is from the model with no tremble (whose results are presented in the penultimate column of

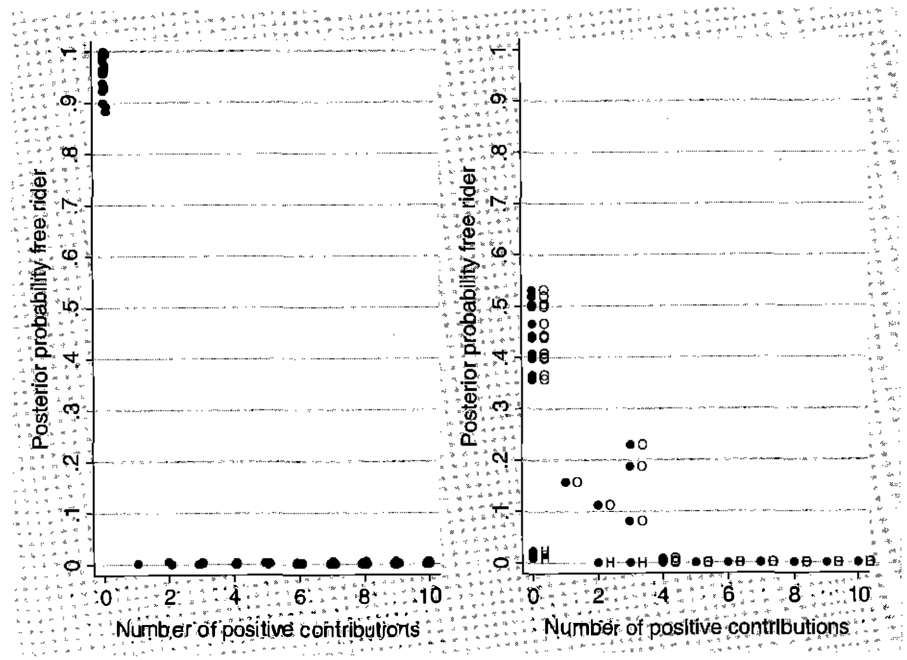



Figure 11.7: Posterior probability of being a free rider against number of positive contributions
Left pane: tremble-free model; right pane: model with tremble
Note: in right pane, "H" and "O" represent subjects in house-money and own-money treatments respectively.

Table 11.7). The right-hand plot is from the model with a tremble (whose results are reported in the final column of Table 11.7).

The left-hand plot of Figure 11.7 shows that, in the absence of a tremble, subjects who never contribute at all have a high posterior probability of being a free-rider, while all subjects with positive contributions have a zero probability of being a free-rider. The right-hand plot of Figure 11.7 shows that, in the presence of a tremble, some subjects with a small number of positive contributions now have a positive posterior probability of being a free-rider, having risen off the horizontal axis. Note further that these subjects tend to be in the "own-money" treatment, and it is this that drives the difference in results between the penultimate and final columns of Table 11.2. The small number of subjects in the own-money treatment who are being (partly) classified as "free-riders", despite their occasional contributions, are tilting the balance between no house money effect and a house money effect.

## 11.8 Summary and Further Reading

The concept of the double hurdle model is well-established in Econometrics. The central idea is that there are two stages to the decision making process: "whether" and "how much". The practice is particularly well-developed in the context of count data outcomes (see Cameron & Trivedi, 2010), although perhaps less so for continuous outcomes. The model has been applied by Jones (1989) to cigarette consumption by individuals, by Burton et al. (1994) to meat consumption by households, and by Dionne et al. (1996) and Moffatt (2005*a*) to data on loan default by borrowers.

There are many natural applications of the hurdle approach in experimental economics. The hurdle concept is clear and intuitive in these applications: there is a mixture of a priori clearly defined types: "zero types" and "contributors". These applications typically involve panel data, and it has therefore been necessary to develop an estimation procedure for the panel hurdle model. The panel version allows estimation of the "dependence" parameter, which is poorly identified in the cross-section case (Smith, 2003). The panel hurdle model was originally developed by Dong & Kaiser (2008) who applied the model to household milk consumption.

The chapter has used two different real data sets to demonstrate estimation of the panel hurdle model. The first application was to Erkal et al.'s (2011) dictator game data, and their main result was reproduced through estimation of the panel hurdle model with dependence. The second application was to Clark's (2002) public goods experiment, where, again using the panel hurdle model, we were able to detect a "house money effect". The role of the tremble parameter was important here, and it was useful to see how the posterior "zero-type" probabilities shifted on the introduction of this parameter.

What is most important for users of the panel hurdle model to understand is that the first hurdle equation can contain characteristics of the subject, which are fixed between time periods. The second hurdle can contain characteristics of the individual *and* the features of the task, and, if desired, interactions between the two.



The most interesting results from the estimation of hurdle models are those where an effect in the first hurdle is contradicted by an effect with the opposite sign in the second hurdle. One example of this is the Andreoni & Vesterlund (2001) effect of gender on giving.

McDowell (2003) has provided advice "from the STATA help desk" on the programming that is required in the estimation of hurdle models. More recently, Garcia (2013) and Engel & Moffatt (2014) have developed STATA software for the estimation of hurdle models, in the latter case, including the panel hurdle model.

Applications of the panel hurdle model in experimental economics include Engel & Moffatt (2012) and Cheung (2014).

## Exercises

1. Derive the likelihood function for the single hurdle model introduced in Section 11.4.3. Explain the logical problem that arises when one attempts to generalise the single hurdle model to a panel data setting.
2. Consider the study of Eckel & Grossman (1998) on dictator game giving and gender. The data is presented within the article. Use their data to estimate a hurdle model with gender in both hurdles. Is the Andreoni & Vesterlund (2001) gender effect evident in this data?
3. Apply the panel hurdle model to the public goods data of Bardsley (2000), contained in the file **bardsley**, that was analysed using different methods in Chapters 6, 8 and 10.

# Chapter 12
# Choice under Risk: Theoretical Issues

## 12.1 Introduction

The purpose of this chapter is to introduce a number of theoretical concepts, an understanding of which is essential in the econometric modelling of choice under risk.

We start by defining two commonly used measures of risk aversion, and then introduce the most popular utility functions, focusing on their advantages and disadvantages with regard to risky choice modelling. We then introduce the notation used to represent lotteries in risky choice problems, we present formal definitions of expected value and expected utility, we demonstrate the method for assessing which of two lotteries is the riskier, and we define stochastic dominance. Then we consider non-expected utility models, with particular attention paid to the probability weighting functions that are an important component of cumulative prospect theory. The chapter concludes by describing various approaches to the construction of stochastic models of risky choice. This final section leads naturally to the following chapter whose purpose is to demonstrate the econometric estimation of these stochastic models.

## 12.2 Utility Functions and Risk Aversion

One of the central features of any model of choice under risk is the utility function that is assumed. We first assume that the outcome of the task is an amount of income, or wealth, which we shall label $x$. The utility function, $U(x)$, is defined over this variable. A number of different specifications of $U(x)$ are possible.

An important point is that the type of utility function required here differs in a particular way from the more usual utility function defined over quantities consumed. The latter type of utility function can be transformed by any order-preserving transformation, and implied behaviour is unaffected. With the utility function $U(x)$ however, it is only permitted to apply transformations that leave the shape unchanged. It is sometimes said that $U(x)$ is "unique up to positive affine transformations". This basically means that the shape of $U(x)$, and in particular its curvature, *is* important in the determination of behaviour.

Two concepts, attributable to Pratt (1964), are central to the modelling of behaviour under risk: absolute risk aversion and relative risk aversion. Both of these measures of risk aversion relate closely to the curvature of the utility function. The coefficient of absolute risk aversion is defined by:

$$A(x) = -\frac{U''(x)}{U'(x)} \tag{12.1}$$

The coefficient of relative risk aversion is defined by:

$$R(x) = -\frac{xU''(x)}{U'(x)} \tag{12.2}$$

Let us remark first that if the utility function is linear, e.g. $U(x) = x$, then both of the above measures of risk aversion are zero, and hence we would say that an individual with this utility function is "risk neutral". We know from everyday experience that most individuals are risk averse, and hence we expect the measures normally to be positive. A risk-seeking individual would have negative values for the two measures.

In certain applications, it is also useful to consider higher-order derivatives. The third derivative is said to represent "prudence" while the fourth represents "temperance" (see Eeckhoudt & Schlesinger, 2006). Here, however, attention is restricted to risk aversion, which is represented by the second derivative.

A straightforward utility function is the "power" utility function, defined by:

$$\begin{aligned} U(x) &= x^{\alpha} \quad x \geq 0; \alpha > 0 \\ &= ln(x) \quad \alpha = 0 \\ &= -x^{\alpha} \quad \alpha < 0 \end{aligned} \tag{12.3}$$

Equation (12.3) is a simple version of the constant relative risk aversion (CRRA) utility function, and the coefficient of relative risk aversion is $1-\alpha$. Note that $\alpha < 0$ for a very risk averse individual, $0 < \alpha < 1$ for a risk averse individual, $\alpha = 1$ for a risk-neutral individual, and $\alpha > 1$ for a risk-loving individual.

A different way of parameterising the CRRA utility function is as follows:

$$\begin{aligned} U(x) &= \frac{x^{1-r}}{1-r} \quad x \geq 0; -\infty < r < \infty; r \neq 1 \\ &= ln(x) \quad r = 1 \end{aligned} \tag{12.4}$$

In (12.4), the coefficient of relative risk aversion is $r$. For a risk-neutral individual, $r = 0$. Note that $r$ can be very large and positive, indicating extreme risk aversion, and it can also be very large and negative, indicating extreme risk-lovingness. Note finally that the function appearing in the first line of (12.4) is not defined when $r = 1$. The limit of the function as $r$ approaches 1 is $ln(x)$ (see Exercise 1), and hence this is what is used when $r = 1$.



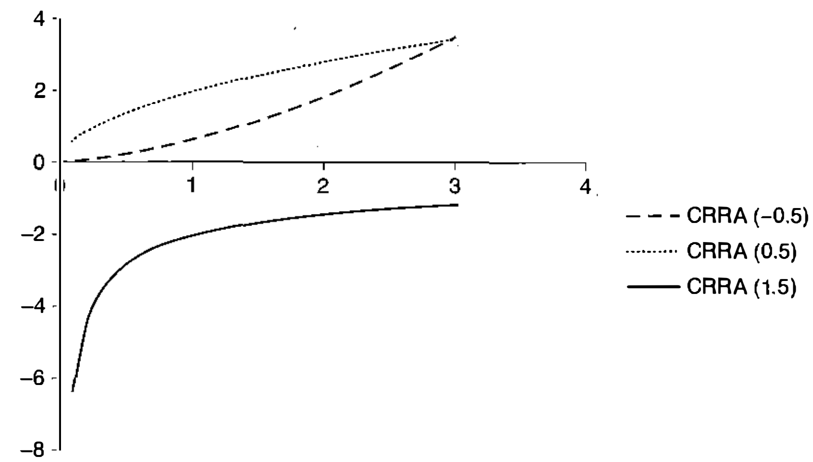



Figure 12.1: CRRA utility functions (12.4) with $r = -0.5, 0.5, 1.5$

Figure 12.1 shows CRRA utility functions, as defined in (12.4), for three different values of $r$. The dashed curve is the function with $r = -0.5$; this function is convex, indicating risk lovingness. The dotted curve is for $r = 0.5$; this function is concave, indicating risk aversion. The solid curve is for $r = 1.5$; this function is even more concave, indicating greater risk aversion. Note also that this curve lies below the horizontal axis, as it must whenever $r > 1$.

One problem with the CRRA functions (12.3) and (12.4) is that they do not fully accommodate zero outcomes. It is very common for the lowest outcome in a lottery to be zero, and hence we need to be able to evaluate $U(0)$. However, when the "power" term is non-positive, i.e. if $\alpha < 0$ in (12.3) or $r \geq 1$ in (12.4), $U(0)$ is not defined. This does not mean that the CRRA utility function cannot be used when the lowest outcome is zero. It simply means that, when the lowest outcome is zero, the function is incapable of explaining high degrees of risk aversion.

A different sort of utility function is the constant absolute risk aversion (CARA) function, defined by:

$$U(x) = 1 - \exp(-rx) \quad x \geq 0; r > 0 \tag{12.5}$$

Here, $r$ is the coefficient of absolute risk aversion. Note that in (12.5), $r$ must be positive; this is because a negative value of $r$ leads to a utility function that is decreasing in $x$. One way of solving this problem is to use the normalised CARA function (see Conte et al., 2011):

$$U(x) = \frac{1 - \exp(-rx)}{1 - \exp(-rx_{max})} \quad 0 \leq x \leq x_{max}; -\infty < r < \infty \tag{12.6}$$

Here, $x_{max}$ is the upper limit of the income variable $x$. Equation (12.6) is increasing in $x$ even for negative values of $r$, and negative values of $r$ represent risk-lovingness.

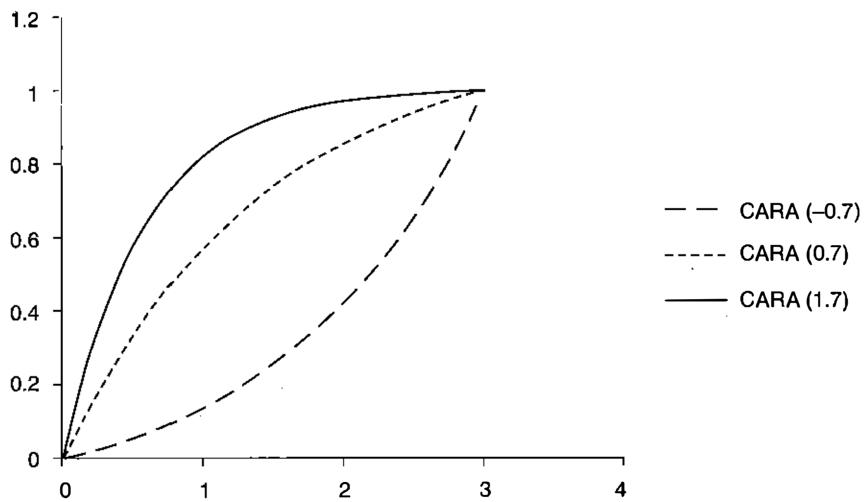



Figure 12.2: CARA utility functions (12.6) with $r = -0.7, 0.7, 1.7$

Note that (12.5) and (12.6) are defined for all values of $x$ including zero. Hence the CARA function avoids the problem identified above relating to the CRRA: the CARA is able to explain high degrees of risk aversion even when the lowest outcome is zero.

Figure 12.2 shows normalised CARA utility functions (12.6), with $x_{max} = 3$, for three different values of $r$. The dashed curve is the function with $r = -0.7$. This function is convex, indicating risk lovingness. The dotted curve is for $r = 0.7$. This function is concave, indicating risk aversion. The solid curve is for $r = 1.7$. This function is even more concave, indicating greater risk aversion.

If we wish to assume constant relative risk aversion, we must necessarily assume decreasing absolute risk aversion (DARA), that is, $A(x)$ falling with $x$. If we wish to assume constant absolute risk aversion, we must necessarily assume increasing relative risk aversion (IRRA).

The differences between CRRA and CARA are clearly seen when we consider altering the money outcomes of a choice problem in particular ways. Specifically, if an individual has CRRA, multiplying all of the outcomes by the same positive number will leave the individual's choice unchanged. If an individual has CARA, adding a fixed money amount to all outcomes will leave the choice unchanged. CARA is therefore a useful assumption to make if we wish to assume away the impact of initial wealth, since under CARA, initial wealth has no effect on choice. In contrast, if CRRA is assumed, the effect of higher initial wealth is less risk aversion in choices.

A more general utility function that has become popular in applications in recent years is the expo-power utility function (Saha, 1993), defined as follows:

$$U(x) = 1 - \exp(-\beta x^{\alpha}) \quad x \geq 0; \alpha \neq 0; \beta \neq 0; \alpha\beta > 0 \tag{12.7}$$



Loosely speaking, (12.7) is a utility function that combines CARA and CRRA. The degree of absolute risk aversion is determined by the parameter $\alpha$: $\alpha < 1$ implies DARA; $\alpha = 1$ implies CARA; and $\alpha > 1$ implies IARA. The degree of relative risk aversion is determined by the parameter $\beta$: $\beta < 0$ implies DRRA, and $\beta > 0$ implies IRRA. Note however that $\beta \neq 0$, so exact CRRA is not included as a special case of (12.7).

Abdellaoui et al. (2007) have introduced the one-parameter expo-power utility function, defined as:

$$\begin{aligned} U(x) &= -\exp\left(\frac{-x^r}{r}\right) \quad x \geq 0; r \neq 0 \\ &= -\frac{1}{x} \quad r = 0 \end{aligned} \tag{12.8}$$

In (12.8), the single parameter $r$ has a clear interpretation as an index of concavity: the smaller is $r$, the more concave the function. Also, as can be easily verified, the function combines the assumptions of DARA and IRRA

## 12.3 Lottery Choice

What we refer to as lotteries are in some theoretical treatments referred to as "prospects" (see e.g. Wakker, 2010). We define a lottery by a vector of probabilities and a corresponding vector of outcomes. In a situation with $n$ outcomes, a lottery would be represented by a vector of outcomes $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_1 < x_2 < \cdots < x_n$, and a corresponding vector of probabilities $\mathbf{p} = (p_1, \ldots, p_n)$. Let us refer to the lottery as $(\mathbf{p}, \mathbf{x})$. In studies of choice under risk, the number of outcomes ($n$) is rarely greater than three.[1]

An important concept is the expected value of the lottery, defined as:

$$EV(\mathbf{p}, \mathbf{x}) = \sum_{j=1}^{n} p_j x_j \tag{12.9}$$

A still more important concept is the expected utility of the lottery, defined by:

$$EU(\mathbf{p}, \mathbf{x}) = \sum_{j=1}^{n} p_j U(x_j) \tag{12.10}$$

where $U(.)$ is one of the utility functions described in Section 12.2 above.

[1] It should be said that a recent strand of the literature is concerned with the extension of risky choice models to allow for *complexity aversion*, for example Sonsino et al. (2002) and Moffatt et al. (2015). In order to ensure the desired variation in complexity levels, some choice problems in these studies contain lotteries with much larger numbers of outcomes.


Figure 12.3: Diagrammatic representation of the two example lotteries defined in (12.11)

A lottery choice problem usually consists of two lotteries between which the subject is asked to choose. If the first lottery is $(\mathbf{p}, \mathbf{x})$, then the second might be $(\mathbf{q}, \mathbf{x})$, where $\mathbf{q} = (q_1, \ldots, q_n)$. It is usually the case that both lotteries have the same vector of outcomes, $\mathbf{x}$, and hence it is acceptable to refer to the two lotteries simply as $\mathbf{p}$ and $\mathbf{q}$.

For an example, let us consider the two three-outcome lotteries:

$$
\begin{aligned}
(\mathbf{p}, \mathbf{x}) &= \left((0.333, 0.167, 0.5)', (0, 10, 20)'\right) \\
(\mathbf{q}, \mathbf{x}) &= \left((0.167, 0.667, 0.167)', (0, 10, 20)'\right)
\end{aligned} \tag{12.11}
$$

A convenient and intuitive means of representing the two lotteries, used elsewhere in this book, is by means of circles divided into sectors, with each outcome being represented by a sector with area proportional to the probability of the outcome. The circles representing the example pair (12.11) are presented in Figure 12.3.

It is usual that one of the two lotteries can be classified as the "safer" and the other as the "riskier". For each lottery, we need to consider the function of a scalar variable $X$, which is the probability of receiving an outcome of *at least* $X$. This is clearly a non-increasing function of $X$. We may then plot the functions arising from the two lotteries against $X$ on the same graph. Typically, the two functions will cross, and the one that crosses from below will be the one associated with the riskier lottery. If the two functions do not cross, one lottery is said to *stochastically dominate* the other (see next sub-section).

Such a plot is shown for the example pair (12.11) in Figure 12.4. We see that the function associated with $\mathbf{p}$ is the one that crosses the other from below, and we may therefore classify $\mathbf{p}$ as the riskier of the two lotteries.

An individual may use a number of different rules in order to choose between the lotteries $\mathbf{p}$ and $\mathbf{q}$. If she is an expected-value maximiser, she would choose the one with the higher expected value (12.9). If she is an expected utility maximiser, she would choose the one with the higher expected utility (12.10). It is also possible for an individual to behave according to a non-EU rule and some of these are explained in later sections of this chapter.

The expected values of the two example lotteries are:

$$
EV(\mathbf{p}) = 0 + \left(\frac{1}{6} \times 10\right) + \left(\frac{1}{2} \times 20\right) = \underline{11.67}
$$



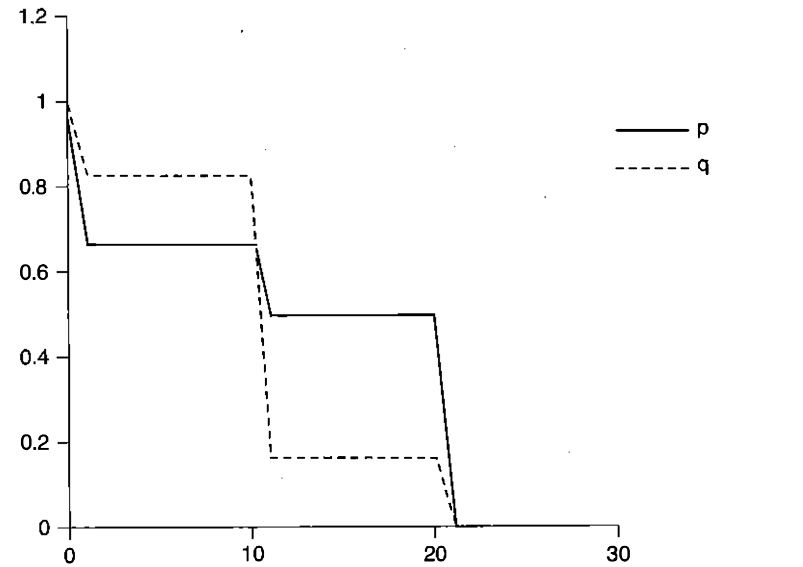



Figure 12.4: Probability of receiving an outcome of at least $X$ against $X$, for the two example lotteries defined in (12.11)

$$
EV(\mathbf{q}) = 0 + \left(\frac{2}{3} \times 10\right) + \left(\frac{1}{6} \times 20\right) = \underline{10.00} \tag{12.12}
$$

Hence, an individual who maximises expected value (i.e. a risk-neutral individual) would choose the riskier lottery, $\mathbf{p}$. However, $\mathbf{q}$ will be chosen by a subject who is sufficiently risk-averse. Assume, for example, that the individual has CRRA utility function (12.4) with $r = 0.75$, and is an expected utility maximiser. This individual's expected utilities of the two lotteries are:

$$
\begin{aligned}
EU(\mathbf{p}) &= 0 + \left(\frac{1}{6} \times \frac{10^{0.25}}{0.25}\right) + \left(\frac{1}{2} \times \frac{20^{0.25}}{0.25}\right) = \underline{5.415} \\
EU(\mathbf{q}) &= 0 + \left(\frac{2}{3} \times \frac{10^{0.25}}{0.25}\right) + \left(\frac{1}{6} \times \frac{20^{0.25}}{0.25}\right) = \underline{6.152}
\end{aligned} \tag{12.13}
$$

Hence this individual will choose the safer lottery, $\mathbf{q}$, since it has the higher expected utility. In fact, as can be verified, for any value of $r$ above 0.42, an EU maximising individual will choose $\mathbf{q}$. It follows that any subject choosing $\mathbf{q}$ is revealing that their current coefficient of relative risk aversion is at least 0.42; a subject choosing $\mathbf{p}$ is revealing that it is less than 0.42. This is, in fact, the basis of the random preference model covered in detail in the next chapter: repeated choice data may be used to infer the distribution of the risk aversion parameter both within and between individuals.

## 12.4 Stochastic Dominance

Lottery **p** stochastically dominates lottery **q** if the probability of earning a fixed amount $X$ or larger is always at least as great (and sometimes greater) if you choose **p** instead of **q**. An example is:

$$(\mathbf{p}, \mathbf{x}) = \left((0.25, 0, 0.75)', (0, 10, 20)'\right)$$
$$(\mathbf{q}, \mathbf{x}) = \left((0.25, 0.25, 0.5)', (0, 10, 20)'\right) \qquad (12.14)$$

This pair of lotteries is presented in Figure 12.5. That **p** stochastically dominates **q** is evident in Figure 12.6, in which the probability function corresponding to **p** is always above (or at least coincides with) that corresponding to **q**. We shall refer to a choice problem such as this as a "dominance problem".

The significance of "dominance problems" is that any rational subject (and rationality of course includes EU-maximisation, EV maximisation, and most other sensible theories) will choose the lottery that stochastically dominates. In choice experiments, however, it is considered perfectly normal for a small number of subjects to violate stochastic dominance. For example, Loomes & Sugden (1998) report that 1.5% of the choices made in dominance problems are violations of dominance.


Figure 12.5: Example of a lottery pair in which lottery **p** stochastically dominates lottery **q**

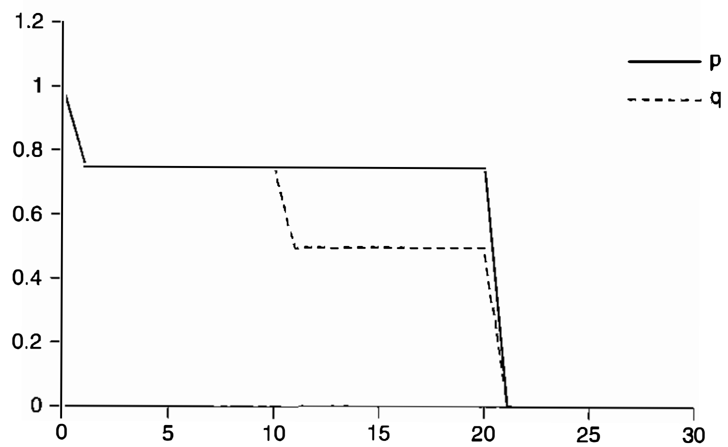



Figure 12.6: Probability of receiving an outcome of at least X against X, for the two example lotteries defined in (12.14)



The question then arising is how to explain these anomalous observations given that standard theories do not allow them to occur. The answer is in the stochastic specification – this must be made sufficiently flexible to allow a small probability of dominance violation. We return to this point in our discussion of stochastic specifications in Section 12.6.

## 12.5 Non-expected Utility Models

It is well known that the assumption of EU does not stand up well when confronted by experimental data. Many types of violation of EU are seen, including the Allais (1953) paradox which was discussed in Chapter 3. For this reason, we need to extend the EU model in ways that explain EU violations. There are many "non-expected utility models" (see Starmer, 2000). Probably the most well-known is cumulative prospect theory (CPT) of Tversky & Kahneman (1992).

The two key features of CPT are the weighting of probabilities and loss aversion. Weighting of probabilities amounts to the phenomenon of individuals not using the true probabilities when evaluating a lottery, but instead using "transformed probabilities" obtained using what is known as a weighting function. Typical features seen in this transformation process are the overweighting of low probabilities of favourable outcomes, and the overweighting of low probabilities of undesirable outcomes.

Loss aversion is an encapsulation of the principle that "losses loom larger than gains" (Kahneman & Tversky, 1979). This is captured by a loss aversion parameter, usually $\lambda$, representing the multiple by which losses are peceived as worse (in terms of utility) than gains. When an experiment contains choice problems in which some outcomes are negative, it is necessary to extend the choice model to include the loss aversion parameter, $\lambda$. However, in all of the examples in this text, negative outcomes are avoided.

Our own implementation of CPT will therefore involve probability weighting, but not loss aversion. When CPT is applied to a situation in which all outcomes are non-negative, the model actually becomes equivalent to rank-dependent (RD) utility theory (Quiggin, 1982).

### *12.5.1 Weighting functions*

In this section we focus on the parametric specification of the weighting function, one of the key components of CPT. A useful survey of weighting functions has been provided by Stott (2006).

Again we denote as $(\mathbf{p}, \mathbf{x})$ the lottery whose possible outcomes are contained in the vector $\mathbf{x} = (x_1, \ldots, x_n)$ with $x_1 < x_2 < \cdots < x_n$ , and whose corresponding probabilities are contained in the vector $\mathbf{p} = (p_1, \ldots, p_n)$. According to

CPT, an individual does not use the true probabilities contained in $\mathbf{p}$ when evaluating the lottery, but rather uses the transformed probabilities $\tilde{\mathbf{p}} = (\tilde{p}_1, \ldots, \tilde{p}_n)$. The transformed probabilities are obtained from the true probabilities as follows:

$$\tilde{p}_j = w\left(\sum_{k=j}^{n} p_k\right) - w\left(\sum_{k=j+1}^{n} p_k\right) \qquad j = 1 \ldots n-1$$
$$\tilde{p}_n = w(p_n) \tag{12.15}$$

where $w(.)$ is the weighting function, with the properties $w(0) = 0$, $w(1) = 1$. Note that (12.15) ensures that the transformed probabilities sum to one, i.e. that $\sum_{j=1}^{n} \tilde{p}_j = 1$.

Again, it is useful to consider the case with only three outcomes, for which (12.15) becomes:

$$\begin{aligned} \tilde{p}_3 &= w(p_3) \\ \tilde{p}_2 &= w(p_2 + p_3) - w(p_3) \\ \tilde{p}_1 &= 1 - w(p_2 + p_3) \end{aligned} \tag{12.16}$$

The three-outcome case is illustrated in Figure 12.7, in which the curve is the weighting function. The inverse-S shape of this curve causes the probabilities of the best and worst outcomes to be overweighted (i.e. $\tilde{p}_3 > p_3$ and $\tilde{p}_1 > p_1$ respectively), implying of course that the probability of the middle outcome is underweighted (i.e. $\tilde{p}_2 < p_2$). Note that if the weighting function coincided with the 45 degree line (i.e. if $w(p) = p$), the probabilities would be correctly weighted, and we would be back to EU.

The inverse-S shape seen in Figure 12.7 is in fact the standard assumption for weighting functions. It is standard to assume that small probabilities of both the best and worst outcomes are overweighted, at the expense of the middle outcome(s).

Three parametric functions that appear in the literature are specified below:

$$\begin{aligned} &\text{Power:} && w(p) = p^{\gamma}, \quad \text{with } \gamma > 0 \\ &\text{Tversky and Kahneman (1992):} && w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}, \\ & && \text{with } \gamma > 0.279 \\ &\text{Prelec (1998):} && w(p) = \exp\left(-\alpha(-\ln p)^{\gamma}\right), \\ & && \text{with } \alpha > 0; \gamma > 0 \end{aligned} \tag{12.17}$$

All three functions are shown with arbitrarily chosen parameter values in Figure 12.8. The first of these, the power weighting function, might be seen as undesirably restrictive since it does not allow an inverse S-shape; it is either completely above (if $\gamma < 1$) or completely below (if $\gamma > 1$) the 45-degree line. The second function is due to Tversky & Kahneman (1992). Despite having only one parameter,



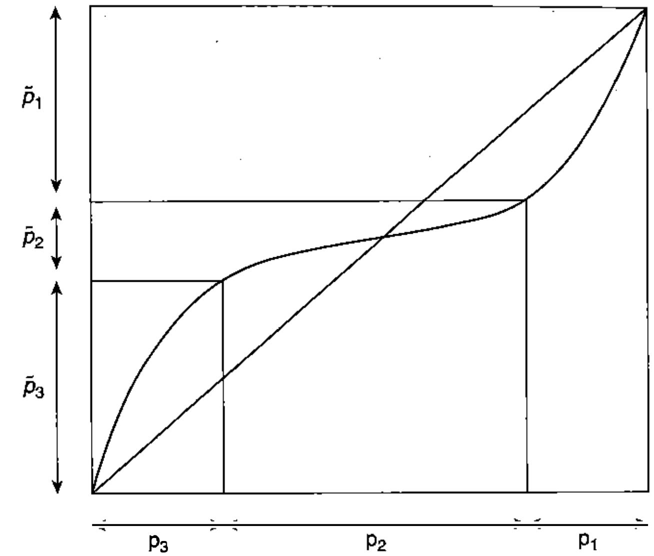



Figure 12.7: An inverted S-shaped weighting function; three outcomes assumed

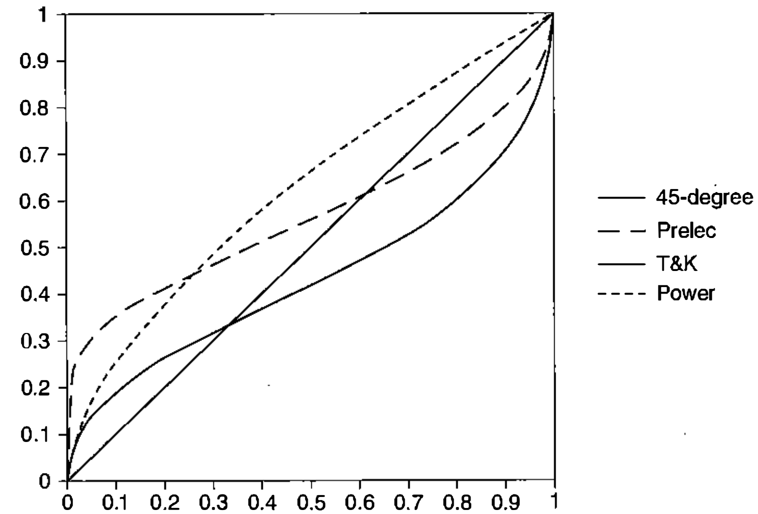



Figure 12.8: (i) Power function with $\gamma = 0.6$; (ii) T&K function with $\gamma = 0.6$; Prelec function with $\gamma = 0.5$ and $\alpha = 0.7$

this function has the required inverse-S shape (if $0.279 < \gamma < 1$), crossing the 45-degree line at a point that depends on the value of $\gamma$. The lower limit of $\gamma$ is required for monotonicity. If $\gamma = 1$, we have EU.

The third function is due to Prelec (1998) and has two parameters. When both of these parameters are equal to one, we have EU. The function crosses the 45-degree line at a point given by $p = \exp\left(-\alpha^{\frac{1}{1-\gamma}}\right)$ (see Exercise 3). The parameter $\alpha$ reflects pessimism (with values less than 1 indicating optimism) and the parameter $\gamma$ determines the pronouncedness of the inverse-S shape (the lower the value of $\gamma$, the more pronounced the inverse-S; $\gamma > 1$ would instead indicate an S-shape). It is typically found that both parameters are somewhat less than one.

Having found the transformed probabilities, we may compute an individual's valuation of the lottery as:

$$V(\mathbf{p}) = \sum_{j=1}^{n} \tilde{p}_j U(x_j) \tag{12.18}$$

## 12.6 Stochastic Models of Choice under Risk

In earlier sections of this chapter, we have discussed various theoretical concepts essential to the understanding of risky choice. Assuming EU, we asserted that the condition for choosing lottery $\mathbf{p}$ over lottery $\mathbf{q}$ is $EU(\mathbf{p}) > EU(\mathbf{q})$, where the $EU(.)$ function is evaluated by combining probabilities with the utilities of outcomes. Assuming CPT (or RD theory), the condition for choosing $\mathbf{p}$ is given by $V(\mathbf{p}) > V(\mathbf{q})$ where the $V(.)$ function is evaluated similarly to the $EU(.)$ function, but with transformed probabilities in place of true probabilities.

For convenience, we introduce the symbol $\nabla$ (or $\tilde{\nabla}$) to represent the valuation differential of $\mathbf{p}$ over $\mathbf{q}$. That is, under EU, $\nabla = EU(\mathbf{p}) - EU(\mathbf{q})$, and under CPT, $\tilde{\nabla} = V(\mathbf{p}) - V(\mathbf{q})$. Now, the condition for choosing $\mathbf{p}$ is simply $\nabla > 0$ or $\tilde{\nabla} > 0$.

Naturally, in order to make these models operational, we need to introduce a stochastic component. There are a number of possible approaches to this problem, and to introduce these is the purpose of this section.

Perhaps the most obvious approach is simply to apply an additive error term to the valuation differential, so that the condition for choosing $\mathbf{p}$ becomes (assuming EU):

$$\text{choose } \mathbf{p} \text{ if } \nabla + \varepsilon > 0;\ \varepsilon \sim N(0, \sigma^2) \tag{12.19}$$

It is natural to assume a mean-zero normal distribution for the error term $\varepsilon$, although another possibility would be the logistic distribution. As soon as the error term appears, behaviour is no longer deterministic; it is now described in terms of probabilities. Given the distributional assumption in (12.19), the probability of choosing $\mathbf{p}$ is:

$$P(\text{choose } \mathbf{p}) = P(\nabla + \varepsilon > 0) = P(\varepsilon > -\nabla) = P\left(\frac{\varepsilon}{\sigma} > -\frac{\nabla}{\sigma}\right) = \Phi\left(\frac{\nabla}{\sigma}\right) \tag{12.20}$$

where $\Phi(.)$ is the standard normal cdf. The parameter $\sigma$ is interpreted as the noisiness in choice: if $\sigma = 0$, there is zero noise, and choice is deterministic; if $\sigma = \infty$, choice is driven entirely by noise, and both choices have probability 0.5.



The practice of applying an additive error to the valuation differential, as in (12.19), has come to be known as the Fechner approach after Fechner (1860). The approach was used by Hey & Orme (1994).

A perhaps more subtle approach is the random preference (RP) approach, due to Loomes & Sugden (1998). Here, variation in choice is explained in terms of variation (either within or between individuals) in the risk attitude parameter. For the purposes of an example, let us focus on the EU model with power utility (12.3), and let us assume that there are three outcomes: 0, 1, 2. With these assumptions, the valuation differential is:

$$\nabla(\alpha) = EU(\mathbf{p}) - EU(\mathbf{q}) = \left(p_2 + p_3 2^\alpha\right) - \left(q_2 + q_3 2^\alpha\right) = d_2 + d_3 2^\alpha \tag{12.21}$$

where $d_2 = p_2 - q_2$ and $d_3 = p_3 - q_3$. The condition for choosing $\mathbf{p}$ is therefore:

$$d_2 + d_3 2^\alpha > 0 \Leftrightarrow \alpha > \frac{\ln\left(-\frac{d_2}{d_3}\right)}{\ln 2} \tag{12.22}$$

In order to allow for randomness in choices, we need to assume that the power parameter $\alpha$ is a random variable. For a particular individual, let us assume that:

$$\ln\alpha \sim N(\mu, \sigma^2) \tag{12.23}$$

The reason for choosing the lognormal distribution in (12.23) is that there is a requirement that the power parameter in (12.3) is positive. Combining (12.22) and (12.23) we obtain the choice probability:

$$\begin{aligned} P(\text{choose } \mathbf{p}) &= P\left(\alpha > \frac{\ln\left(-\frac{d_2}{d_3}\right)}{\ln 2}\right) = P\left(\ln\alpha > \ln\left(\frac{\ln\left(-\frac{d_2}{d_3}\right)}{\ln 2}\right)\right) \\ &= \Phi\left(\frac{\mu - \ln\left(\frac{\ln\left(-\frac{d_2}{d_3}\right)}{\ln 2}\right)}{\sigma}\right) \end{aligned} \tag{12.24}$$

A problem with the RP model is that it is unable to explain violations of stochastic dominance. This is seen when we consider that a choice problem for which $\mathbf{p}$ dominates $\mathbf{q}$ is such that $|\, d_2 \,| < d_3$, and therefore $\ln\left(-\frac{d_2}{d_3}\right) < 0$, and since $\alpha > 0$, the first equality in (12.24) tells us that $\mathbf{p}$ will always be chosen. That is, under the RP model, choice of a stochastically dominated alternative is impossible. The problem with this is that experimenters who include dominance problems usually find that at least some subjects occasionally violate dominance. Whatever the reason for this, these choices need to be explained.

The popular solution to this problem (see Loomes et al., 2002; Moffatt & Peters, 2001) is to include a *tremble term*. The tremble term is a parameter, $\omega$ say, that represents the probability of an individual losing concentration and choosing randomly between the two alternatives. If we incorporate a tremble term into (12.24), we obtain the "RP model with tremble":

$$P(\text{choose } \mathbf{p}) = (1-\omega)\Phi\left(\frac{\mu - \ln\left(\frac{\ln\left(-\frac{d_2}{d_3}\right)}{\ln 2}\right)}{\sigma}\right) + \frac{\omega}{2} \qquad (12.25)$$

Harless & Camerer (1994) explained variation in choices purely in terms of trembles. We prefer to think of the role of the tremble term as one of complementing other models, such as RP. The tremble can also be interpreted in the Fechner model, although it is less important there since the Fechner model is capable of explaining dominance violations.

## 12.7 Summary and Further Reading

Economic theory has given rise to many models for decision under risk, starting with Neumann & Morgenstern (1947). Pratt (1964) provided the foundations for measuring individual risk attitude using the curvature of the utility function. Saha (1993) is a useful reference for understanding the differences between CRRA and CARA, and also for producing a utility function that combines the two. Readers interested in prudence and temperance should consult Eeckhoudt & Schlesinger (2006).

Departures from EU theory often fall under the headings of prospect theory, cumulative prospect theory, or rank dependent theory, and key references for these non-EU theories include Kahneman & Tversky (1979), Tversky & Kahneman (1992) and Wakker (2010). Stott (2006) provided a useful survey of functional forms for weighting functions, including that of Prelec (1998) which is the functional form assumed in the econometric model developed in the next chapter. Starmer (2000) reviews experimental evidence of EU violations.

We considered various approaches to incorporating a stochastic element to models for decision under risk. Readers should consult Wilcox (2008) for a much more detailed treatment of these approaches.

## Exercises

1. Consider the utility functions:

$$U_1(x) = \frac{x^{1-r}}{1-r} \qquad r \neq 1$$
$$U_2(x) = 1 - \exp(-rx)$$



   a. Derive the coefficient of relative risk aversion for $U_1$, and the coefficient of absolute risk aversion for $U_2$. Hence explain why the two utility functions are respectively labelled "CRRA" and "CARA".
   b. Prove that: CRRA implies DARA; and CARA implies IRRA.
   c. For each of $U_1$ and $U_2$, consider the effects on choice (i.e. on the individuals propensity to choose the safer choice) of:
      i. a doubling of all outcomes in the choice problem;
      ii. adding a fixed amount $c$ to all outcomes in the choice problem.
   d. For each of the utility functions, what is the impact of an increase in initial wealth on the propensity to choose the safer choice?
   e. Show that, as $r \to 1$, $U_1(x) \to \ln(x)$.

2. Consider the expo-power utility function:

$$U(x) = 1 - \exp(-\beta x^{\alpha}) \qquad x \geq 0; \alpha \neq 0; \beta \neq 0; \alpha\beta > 0$$

   a. Show that $\alpha < 1$ implies DARA, $\alpha = 1$ implies CARA, and $\alpha > 1$ implies IARA.
   b. Show that $\beta < 0$ implies DRRA, and $\beta > 0$ implies IRRA.

3. Consider the Prelec (1998) weighting function introduced in Section 12.5.1:

$$w(p) = \exp\left(-\alpha(-\ln p)^{\gamma}\right) \qquad \alpha > 0; \gamma > 0$$

   Find an expression for the value of $p$ at which this function crosses the 45 degree line. To do this, you need to set $w(p) = p$ and then solve for $p$ in terms of $\alpha$ and $\beta$.

# Chapter 13
# Choice under Risk: Econometric Modelling

## 13.1 Introduction

In many ways this chapter is a focal point of the book. This is partly because it brings together a large number of concepts and techniques introduced in other chapters, and applies them all to a key area of experimental economics: the econometric modelling of risky choice.

The main links with other chapters are as follows. This chapter is concerned with the problem of estimating the parameters of the type of theoretical model (including utility functions and weighting functions) introduced in Chapter 12. Certain econometric issues such as the choice between stochastic specifications, have already been introduced in a simpler setting in Chapter 6. The experimental design that is used to obtain the data analysed in this chapter is an optimal design in the sense to be explained in Chapter 14. The data is, in fact obtained using simulation, and the methods of simulation were introduced and demonstrated in Chapter 9. The method of estimation is the method of simulated likelihood (MSL) which has been introduced in Chapter 10. Finally Chapter 5 contained a model of decision time in a risky choice setting, in which one of the key explanatory variables was "closeness-to-indifference" in a given choice problem for a given subject. The method used to obtain this variable, following estimation of the choice model, is explained in this chapter.

It is important to make clear at the outset why our chosen focus is an experiment consisting of a long sequence of choice tasks, each being encountered in isolation. An alternative setting is the Multiple Price List (MPL) of Holt & Laury (2002), which has been discussed in Section 6.5.1. The MPL is an ordered list of lottery pairs that are normally presented to subjects on a single screen. The lottery pairs are designed in such a way that all subjects are expected to choose the left-hand (say) lottery in the first pair, while all are expected to choose the right-hand lottery in the last pair, and the position in which they *switch* from left to right is then used to infer their risk attitude. While this is a very useful, and popular, means of eliciting risk attitude, we do not consider it to be suitable when behaviour under risk is the focus of the investigation. This is because, under the MPL, only one piece of information is extracted from each subject: the switching point. Some researchers treat the sequence of implied "choices" as a set of independent observations, and conduct estimation on this basis, but this is incorrect. To do this correctly, a highly restrictive correlation structure between observations would need to be incorporated, which, if done correctly, would essentially result in only one piece of information being conveyed: the switching point. It is our view that a single observation for each subject is insufficient if the focus of the investigation is behaviour under risk. This is because within-subject variation is not identified, and estimation of other parameters is imprecise. It is clear from the literature that the MPL is a popular device for computing an estimate of risk attitude for use as an extraneous estimate in a study of something other than behaviour under risk. For example, see Andersen et al.'s (2008) study of the joint elicitation of risk and time preferences. However, we are unrelenting in the view that more informative designs, such as the one used as an example in this chapter, are required when behaviour under risk is the focus.

The risky choice data used in this chapter is simulated. The chosen design contains only three outcomes, \$0, \$10 and \$20. There are 50 distinct choice problems, each consisting of two lotteries with these outcomes. The probabilities have been chosen in such a way as to maximise the information in the sample in the sense to be explained in detail in Chapter 14. Appendix C contains the probabilities defining the 50 choice problems.

Three of the choice problems are presented for illustrative purposes in Figure 13.1. In tasks 10 and 40, lottery **p** is the riskier, while in task 49, lottery **p** dominates.

In the simulation, there are 60 subjects. Each subject faces the 50 choice problems in random order, and then faces the same 50 choice problems the next day in a different order. Hence the total number of choice tasks for each subject is 100. For each task the subject must choose one of the two alternatives, and indifference is

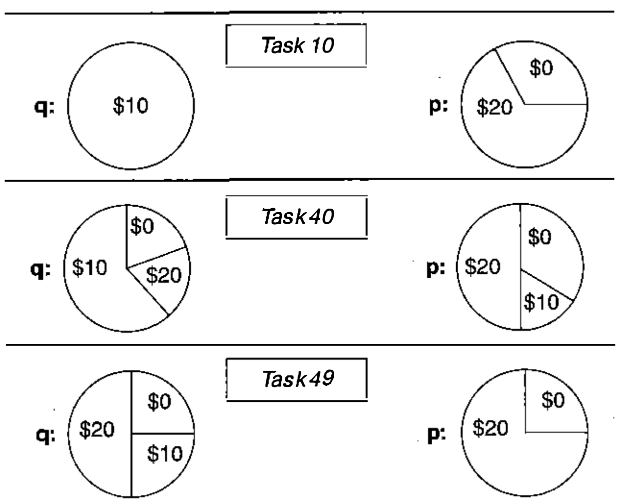



Figure 13.1: Typical choice tasks in the risky-choice experiment



not allowed. The random lottery incentive system is used, meaning that, at the end of the sequence of choice problems, one of the 100 problems is selected at random and played for real. This is a means of ensuring (notwithstanding the Holt critique, Holt, 1986) that subjects treat each choice task as if it is the only task.

The straightforward "power" utility function is assumed. As explained in Chapter 12, this is a version of the constant relative risk aversion (CRRA) utility function. The probability weighting function assumed is the one due to Prelec (1998). Two alternative stochastic specifications are assumed: the Fechner model and the RP model – these were also discussed in Chapter 12. In addition, a tremble parameter is assumed. Some models of risky choice assume that all variation is explained by a tremble. However, we do not consider this to be a sensible approach and it is not pursued.

The risky choice models that were considered in Chapter 6 were "cross-section" models, in the sense that it was assumed that there was only one observation for each subject. In this chapter we progress to the more usual setting in which each subject has engaged in a sequence of choice problems. Hence the available data is panel data, and, in estimation, we need to separate the within-subject variation from the between-subject variation. The random effects approach is used for this purpose. One of the reasons for choosing random effects over fixed effects is that it allows us to obtain estimates of risk attitude for even the most risk-averse or risk-loving subjects; this is sometimes impossible, for reasons of identification, when estimating separately by subject. The method of estimation is the method of simulated likelihood (MSL), explained in detail Chapter 10.

As in previous chapters, the estimation code, and the code that surrounds it, will be explained in some detail. We consider it important for readers to understand each step as fully as possible, since this will prepare them better for the task of tailoring the code to their own application.

The code consists of three sections: simulation; estimation; and post-estimation. In the simulation section, we simulate two data sets: one that assumes the Fechner model; the other that assumes the RP model. The estimation section contains programs for evaluating the log-likelihood for each of these two models, and calls to these programs. Hence there are four different estimations: each model is applied to each data set. This is useful since it gives a feel for the consequences of applying the "wrong" model, as well as giving the reassurance of correct estimation when applying the "correct" model.

The post-estimation stage contains the manipulation of various variables saved during the estimation stage. One task that is performed in this stage is the carrying out of the Vuong test and the Clarke test. These are non-nested tests that are appropriate for testing between the Fechner and RP models. Another task that is performed is the construction of a posterior estimate of risk attitude for each subject in the sample, and the derivation from this of a variable representing the estimated closeness-to-indifference for each subject in each choice problem. As we know from Chapter 5, closeness-to-indifference is one of the key determinants of decision time. In fact, the final task carried out in the STATA program is to simulate the decision times for the choice data, using closeness-to-indifference as well as other features of the task as explanatory variables. The resulting data is contained in the file **decision_times_sim** that was used in Chapter 5.

## 13.2 The Choice Models

### 13.2.1 The framework

All choice problems in the (simulated) experiment involve combinations of the three outcomes: \$0, \$10, \$20. Henceforth all money amounts will be measured in units of \$10, so the three outcomes become simply 0, 1, and 2. We index the problems in the experiment by $t$ ($t = 1, \dots, T$). For most choice problems, one of the two lotteries may be classified as the "riskier" lottery, and the other as the "safer" lottery. When this is not possible, the problem is one of "dominance", since one lottery first-order stochastically dominates the other (see Chapter 12). If task $t$ is a non-dominance problem, we will label the riskier lottery as $\mathbf{p}_t$, and the safer as $\mathbf{q}_t$. For a dominance problem, $\mathbf{p}_t$ will be the dominating lottery, and $\mathbf{q}_t$ the dominated. $\mathbf{p}_t = (p_{1t}, p_{2t}, p_{3t})'$ and $\mathbf{q}_t = (q_{1t}, q_{2t}, q_{3t})'$ are vectors containing three probabilities corresponding to the three possible outcomes.

We index subjects by $i$ ($i = 1, \dots, n$), and we assume the Power utility function:[1]

$$U_i(x) = x^{r_i} \qquad r_i > 0 \tag{13.1}$$

As explained in Chapter 12, (13.1) is a simple version of the Constant Relative Risk Aversion (CRRA) utility function, and the coefficient of relative risk aversion (for subject $i$) is $1 - r_i$. Note that $0 < r_i < 1$ for a risk averse-subject, $r_i = 1$ for a risk-neutral subject, and $r_i > 1$ for a risk-loving subject. Note also that $U_i(0) = 0$. The reason why we do not allow $r_i < 0$ is that choice problems will typically involve the outcome zero, and $U_i(0)$ is not defined when $r_i < 0$.

Assuming (13.1), and under the assumption of Expected Utility (EU) maximisation, subject $i$'s valuations of the two lotteries in choice problem $t$ are:

$$\begin{aligned} EU_i(\mathbf{p}_t) &= p_{2t} + p_{3t}2^{r_i} \\ EU_i(\mathbf{q}_t) &= q_{2t} + q_{3t}2^{r_i} \end{aligned} \tag{13.2}$$

Let:

$$\begin{aligned} d_{1t} &= p_{1t} - q_{1t} \\ d_{2t} &= p_{2t} - q_{2t} \\ d_{3t} &= p_{3t} - q_{3t} \end{aligned} \tag{13.3}$$

The three quantities in (13.3) are probability differences. Note that, for a non-dominance problem, $d_{1t} \geq 0$, $d_{2t} \leq 0$, $d_{3t} \geq 0$ and furthermore $|d_{2t}| \geq d_{3t}$. For a dominance problem, this final equality is not met, so that $|d_{2t}| > d_{3t}$. An example of a dominance problem is task 49 shown in Figure 13.1. The formal definition of stochastic dominance was given in Chapter 12.

[1] When discussing the power utility function in Chapter 12, $\alpha$ was used as the power parameter. This was because $r$ was being used for a different specification. Here, only the power utility function is used, and $r$ is used for the power parameter.



Continuing to assume EU, subject $i$ chooses $\mathbf{p}_t$ over $\mathbf{q}_t$ if their *valuation differential* is positive, that is, if:

$$\nabla_{it}(r_i) \equiv EU_i(\mathbf{p}_t) - EU_i(\mathbf{q}_t) = d_{2t} + d_{3t}2^{r_i} > 0 \tag{13.4}$$

To allow estimation, we need to incorporate a stochastic element to (13.4). Broadly, there are two approaches to this: the Fechner approach, and the random preference (RP) approach.

### 13.2.2 Incorporating violations of EU

As explained in Chapter 12, violations of EU are often considered within the framework of Cumulative Prospect Theory (CPT, Tversky & Kahneman, 1992). In the present setting in which outcomes are always non-negative, CPT is equivalent to Rank-Dependent (RD) Theory (Quiggin, 1982). The central feature of the model is the presence of a weighting function which transforms true probabilities into perceived probabilities, and allows, for example, over-weighting of small probabilities of favourable outcomes.

We use Prelec's (1998) weighting function, defined as:

$$w(p) = \exp[-\alpha(-\ln p)^{\gamma}] \tag{13.5}$$

As explained in Chapter 12, when there are three outcomes with probabilities $p_1, p_2, p_3$, (13.5) is applied as follows:

$$\begin{aligned} \tilde{p}_3 &= w(p_3) \\ \tilde{p}_2 &= w(p_2 + p_3) - w(p_3) \\ \tilde{p}_1 &= 1 - w(p_2 + p_3) \end{aligned} \tag{13.6}$$

Having obtained the transformed probabilities (13.6), we can define, analogously with (13.2), the valuations of the two lotteries:

$$\begin{aligned} V_i(\mathbf{p}_t) &= \tilde{p}_{2t} + \tilde{p}_{3t}2^{r_i} \\ V_i(\mathbf{q}_t) &= \tilde{q}_{2t} + \tilde{q}_{3t}2^{r_i} \end{aligned} \tag{13.7}$$

Further, we define analogously with (13.3), the transformed probability differences:

$$\begin{aligned} \tilde{d}_{1t} &= \tilde{p}_{1t} - \tilde{q}_{1t} \\ \tilde{d}_{2t} &= \tilde{p}_{2t} - \tilde{q}_{2t} \\ \tilde{d}_{3t} &= \tilde{p}_{3t} - \tilde{q}_{3t} \end{aligned} \tag{13.8}$$

Finally, we obtain, analogously to (13.4), the transformed valuation differential:

$$\tilde{\nabla}_{it}(r_i) \equiv V_i(\mathbf{p}_t) - V_i(\mathbf{q}_t) = \tilde{d}_{2t} + \tilde{d}_{3t}2^{r_i} > 0 \tag{13.9}$$

### 13.2.3 The Fechner model

The basis of the Fechner model (Fechner, 1860; Hey & Orme, 1994) is that each subject ($i$) has a risk-attitude parameter ($r_i$) that is constant between tasks. The model is obtained by appending a homoscedastic error term to (13.9), whereupon the condition for the choice of $\mathbf{p}_t$ becomes:

$$\begin{aligned} \tilde{\nabla}_{it}(r_i) + \epsilon_{it} &> 0 \\ \epsilon_{it} &\sim N(0, \sigma^2) \end{aligned} \tag{13.10}$$

Defining the binary indicator $yy_{it} = 1(-1)$ if subject $i$ chooses $\mathbf{p}_t$ ($\mathbf{q}_t$), the probabilities of the two choices (conditional on the risk-attitude parameter $r_i$) are given by:

$$P(yy_{it}|r_i) = \Phi\left(yy_{it} \times \frac{\tilde{\nabla}_{it}(r_i)}{\sigma}\right) \tag{13.11}$$

where $\Phi(.)$ is the standard normal c.d.f.

### 13.2.4 The Random Preference (RP) model

The random preference model is obtained by assuming that a subject's risk-attitude parameter varies randomly between tasks. So, we assume that the risk attitude of subject $i$ in task $t$ comes from the distribution:

$$\ln(r_{it}) \sim N\left(m_i, \sigma^2\right) \tag{13.12}$$

The reason for assuming a log-normal distribution in (13.12) is that there is a logical requirement that the power parameter $r$ is strictly positive.

From (13.9), the condition for choosing $\mathbf{p}_t$ over $\mathbf{q}_t$ is

$$\tilde{d}_{2t} + \tilde{d}_{3t} 2^{r_i} > 0 \tag{13.13}$$

which may be rearranged to give:

$$r_{it} > \frac{\ln\left(-\frac{\tilde{d}_{2t}}{\tilde{d}_{3t}}\right)}{\ln(2)} \equiv \tilde{r}_t^* \tag{13.14}$$

Note that in (13.14) we are defining $\tilde{r}_t^*$ to be the risk attitude parameter that would (under RD theory) make a subject exactly indifferent between the two lotteries in task $t$; we shall refer to this as the "threshold risk attitude" associated with task $t$. If the subject's current risk attitude is larger than this threshold, they will choose the riskier lottery $\mathbf{p}_t$. Note that, according to (13.14), $\tilde{r}_t^*$ is not defined if task $t$ is a dominance problem. However, the decision rule for a dominance problem in the RP model is simple: the dominating alternative, $\mathbf{p}_t$ will be chosen with certainty. Hence we will set $\tilde{r}_t^* = 0$ for dominance problems.



For a non-dominance problem, the probability that the riskier lottery is chosen, conditional on the subject's mean risk attitude, is (using 13.12):

$$\begin{aligned} P(y_{it} = 1|m_i) &= P(r_{it} > \tilde{r}_t^*|m_i) = P\left(\ln(r_{it}) > \ln(\tilde{r}_t^*)|m_i\right) \\ &= P\left(Z > \frac{\ln(\tilde{r}_t^*) - m_i}{\sigma}\right) = \Phi\left(\frac{m_i - \ln(\tilde{r}_t^*)}{\sigma}\right) \end{aligned} \tag{13.15}$$

where $\Phi(.)$ is the standard normal c.d.f. Note that $Z$ is being used to represent a standard normal random variable. $\tilde{r}_t^*$ is the threshold risk-attitude defined in (13.14).

As we did with the Fechner model, we can define the binary indicator $yy_{it} = 1(-1)$ if subject $i$ chooses $\mathbf{p}_t(\mathbf{q}_t)$. The probabilities of the two choices (conditional on $m_i$) are then given by:

$$P(yy_{it}|m_i) = \Phi\left(yy_{it} \times \frac{m_i - \ln(\tilde{r}_t^*)}{\sigma}\right) \tag{13.16}$$

### 13.2.5 Dominance problems in the RP model

If the lottery $\mathbf{p}_t$ first-order stochastically dominates the lottery $\mathbf{q}_t$, then, according to the RP model specified above, a subject will choose $\mathbf{p}_t$ whatever value is taken by their current risk attitude. This is because, for a dominance problem, $|d_{2t}| < d_{3t}$, so the valuation differential (13.4) is positive for all permissible values of $r_{it}$. That is:

$$EU_i(\mathbf{p}_t) - EU_i(\mathbf{q}_t) = d_{2t} + d_{3t} 2^{r_{it}} > 0 \qquad \forall r_{it} > 0 \tag{13.17}$$

This clearly presents a problem for the RP model. Any observed choice of a dominated lottery (i.e. a dominance violation) cannot be explained by the RP model alone. Hence, if the data set contains just one dominance violation, the RP model, as presented above, fails.

It is this consideration, among others, that leads us to introduce the concept of the tremble parameter.

### 13.2.6 The tremble parameter

We shall follow Loomes et al. (2002) by introducing a tremble parameter $\omega$ to both the Fechner and RP models. Issues surrounding the estimation of and testing for tremble effects have been discussed in some detail by Moffatt & Peters (2001).

$\omega$ is the probability that on any task a subject loses concentration and chooses randomly between the two alternatives, with equal probability. Note that a loss of concentration does not necessarily imply that the "incorrect" choice is made. Under such a loss of concentration, the "correct" and "incorrect" choices are equally likely.

Some care is required in combining the tremble probability with the choice probabilities derived above. Generally speaking, the probability of choosing the riskier lottery is derived as follows:

$$P(\mathbf{p}\text{ chosen}) = P(\mathbf{p}\text{ chosen}|\text{no tremble})P(\text{no tremble}) + P(\mathbf{p}\text{ chosen}|\text{tremble})P(\text{tremble}) \quad (13.18)$$

If there is no tremble, the probability of the riskier lottery being chosen is exactly as presented in (13.11) or (13.16) above. If there is a tremble, the probability of this is one half. So (13.18) becomes:

$$P(\mathbf{p}\text{ chosen}) = P(\mathbf{p}\text{ chosen}|\text{no tremble})(1-\omega) + \frac{\omega}{2} \quad (13.19)$$

Applying this extension to the Fechner model (13.11), we obtain:

$$P(yy_{it}|r_i) = (1-\omega)\Phi\left(yy_{it} \times \frac{\tilde{\nabla}_{it}(r_i)}{\sigma}\right) + \frac{\omega}{2} \quad (13.20)$$

where $\Phi(.)$ is the standard normal c.d.f.

Introducing the tremble to the RP model (13.16), we obtain:

$$P\left(yy_{it}|m_i\right) = (1-\omega)\Phi\left(yy_{it} \times \frac{m_i - \ln(\tilde{r}_t^*)}{\sigma}\right) + \frac{\omega}{2} \quad (13.21)$$

### 13.2.7 The role of experience

It is possible that subjects' behaviour may change systematically in the course of the experiment, revealing the effect of experience. In order to allow for this, certain parameters may be allowed to depend on the amount of experience accumulated, which may be measured using the position of the current task in the sequence of tasks. For this purpose we define a variable $\tau_{it}$ to be the position of task $t$ in the sequence of tasks undertaken by subject $i$. Note that a sensible feature of a design is that each subject encounters the sequence of tasks in a different order, that is, $\tau_{it} \neq \tau_{jt}, i \neq j$.

One parameter which it may be considered natural to allow to change with experience is the tremble parameter. This parameter has been defined as the probability that a subject loses concentration when undertaking a particular task. This interpretation may be broadened: another reason why a subject might choose randomly between the two alternatives is a lack of understanding of the task. We might hypothesise that such a lack of understanding is more likely to arise earlier in the sequence than later; hence we would expect the tremble probability to start off at a relatively high value, but to decay towards zero over the course of the experiment.

A suitable specification would therefore be:

$$\omega_{it} = \omega_0 \exp(\omega_1 \tau_{it}) \quad (13.22)$$



Thus $\omega_0$ represents the tremble probability at the start of the experiment, while $\omega_1$ (assuming it is negative) represents the speed of its decay. This specification for the tremble has in fact already been used in the context of a public goods experiment in Section 8.5.

### 13.2.8 Between-subject variation and the sample log-likelihood

The Fechner and RP models were constructed above in terms of the behaviour of an individual subject. The choice probabilities were derived, in the Fechner model conditional on the subjects risk attitude, $r_i$, and in the RP model conditional on the subject's mean risk attitude, $m_i$.

In each case, we need to allow for between-subject variation by assuming that these parameters vary across the population. In the Fechner model, we assume that:

$$\log(r) \sim N(\mu, \eta^2) \quad (13.23)$$

and the sample log-likelihood is then:

$$\log L = \sum_{i=1}^{n} \ln\left(\int_{-\infty}^{\infty} \prod_{t=1}^{T}\left[(1-\omega_{it})\Phi\left(yy_{it} \times \frac{\tilde{\nabla}_{it}(r)}{\sigma}\right) + \frac{\omega_{it}}{2}\right] f(r;\mu,\eta)dr\right) \quad (13.24)$$

where $f(r|\mu,\eta)$ is the lognormal density of the risk aversion parameter:

$$f(r|\mu,\eta) = \frac{1}{r\eta\sqrt{2\pi}}\exp\left[-\frac{(\ln r - \mu)^2}{2\eta^2}\right] \quad r > 0 \quad (13.25)$$

In the RP model, we assume that:

$$m \sim N(\mu, \eta^2) \quad (13.26)$$

and the sample log-likelihood is:

$$\log L = \sum_{i=1}^{n} \ln\left(\int_{-\infty}^{\infty} \prod_{t=1}^{T}\left[(1-\omega_{it})\Phi\left(yy_{it} \times \frac{m - \ln(\tilde{r}_t^*)}{\sigma}\right) + \frac{\omega_{it}}{2}\right] f(m;\mu,\eta)dm\right) \quad (13.27)$$

where $f(m|\mu,\eta)$ is this time the normal density function evaluated at $m$:

$$f(m|\mu,\eta) = \frac{1}{\eta\sqrt{2\pi}}\exp\left[-\frac{(m-\mu)^2}{2\eta^2}\right] \quad -\infty < m < \infty \quad (13.28)$$

Equations (13.24) and (13.27) represent versions of the random effects probit model, of the type estimated in a similar context by Loomes et al. (2002) and Conte et al. (2011).

### *13.2.9 Posterior estimation of risk attitude*

For certain purposes, it is very useful to have a measure of risk attitude for each individual subject. It is more appropriate to use the Fechner model for this purpose. This is because the Fechner model assumes that each individual has their own unique risk attitude parameter which remains the same over time. The RP model assumes that risk attitude varies randomly over time, giving rise to a logical problem when attempting to come up with a single number representing a subject's risk attitude.

Having estimated the Fechner model, we will apply Bayes' rule to obtain the posterior expectation of each subject's risk attitude, conditional on their $T$ choices, as follows (hats indicate evaluation at the MLE):

$$\hat{r}_i = E(r_i | y_{it}, \ldots, y_{iT}) = \frac{\int_{-\infty}^{\infty} r \prod_{t=1}^{T} \left[ (1 - \hat{\omega}_{it}) \Phi \left( yy_{it} \times \frac{\hat{\nabla}_{it}(r)}{\hat{\sigma}} \right) + \frac{\hat{\omega}_{it}}{2} \right] f(r; \hat{\mu}, \hat{\eta}) dr}{\int_{-\infty}^{\infty} \prod_{t=1}^{T} \left[ (1 - \hat{\omega}_{it}) \Phi \left( yy_{it} \times \frac{\hat{\nabla}_{it}(r)}{\hat{\sigma}} \right) + \frac{\hat{\omega}_{it}}{2} \right] f(r; \hat{\mu}, \hat{\eta}) dr} \quad (13.29)$$

## 13.3 Simulation and Estimation

In this section, we shall supply a complete, uninterrupted set of STATA code whose function is to simulate the risky choice data, estimate the models, and perform post-estimation tasks. The code is annotated, providing comment lines at each significant stage of the progam.

Two sets of data are simulated: one assuming the Fechner model is the true model; the other assuming the RP model. Then, both models are estimated on both data sets, leading to four different sets of estimates.

Some readers may find that the annotation is not detailed enough. If this is the case, they are referred back to Chapter 9 for information relating to simulation, and to Chapter 10 for information relating to MSL estimation, including information about the use of Halton draws and use of the d0 likelihood evaluator.

### *13.3.1 Data generating process*

We will assume the following "true" parameter values for the two models:

| Parameter | Fechner | RP |
|---|---|---|
| $\mu$ | −0.88 | −0.88 |
| $\eta$ | 0.20 | 0.20 |
| $\sigma$ | 0.05 | 0.15 |
| $\omega_0$ | 0.06 | 0.06 |
| $\omega_1$ | −0.01 | −0.01 |
| $\beta$ | 0.90 | 0.90 |
| $\gamma$ | 0.80 | 0.80 |



| | i | t | p1 | p2 | p3 | q1 | q2 | q3 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | .05 | 0 | .95 | 0 | 1 | 0 |
| 2 | 1 | 2 | .09 | 0 | .91 | 0 | 1 | 0 |
| 3 | 1 | 3 | .11 | 0 | .89 | 0 | 1 | 0 |
| 4 | 1 | 4 | .13 | 0 | .87 | 0 | 1 | 0 |
| 5 | 1 | 5 | .15 | 0 | .85 | 0 | 1 | 0 |
| 6 | 1 | 6 | .17 | 0 | .83 | 0 | 1 | 0 |
| 7 | 1 | 7 | .19 | 0 | .81 | 0 | 1 | 0 |
| 8 | 1 | 8 | .22 | 0 | .78 | 0 | 1 | 0 |
| 9 | 1 | 9 | .26 | 0 | .74 | 0 | 1 | 0 |
| 10 | 1 | 10 | .3 | 0 | .7 | 0 | 1 | 0 |
| 11 | 1 | 11 | .35 | 0 | .65 | 0 | 1 | 0 |
| 12 | 1 | 12 | .4 | 0 | .6 | 0 | 1 | 0 |
| 13 | 1 | 13 | .45 | 0 | .55 | 0 | 1 | 0 |
| 14 | 1 | 14 | .5 | 0 | .5 | 0 | 1 | 0 |
| 15 | 1 | 15 | .6 | 0 | .4 | 0 | 1 | 0 |
| 16 | 1 | 16 | .75 | 0 | .25 | 0 | 1 | 0 |
| 17 | 1 | 17 | .9 | 0 | .1 | 0 | 1 | 0 |
| 18 | 1 | 18 | .5 | 0 | .5 | .48 | .52 | 0 |
| 19 | 1 | 19 | .5 | 0 | .5 | .44 | .56 | 0 |
| 20 | 1 | 20 | .5 | 0 | .5 | .42 | .58 | 0 |
| 21 | 1 | 21 | .5 | 0 | .5 | .4 | .6 | 0 |
| 22 | 1 | 22 | .5 | 0 | .5 | .38 | .62 | 0 |
| 23 | 1 | 23 | .5 | 0 | .5 | .36 | .64 | 0 |
| 24 | 1 | 24 | .5 | 0 | .5 | .34 | .66 | 0 |
| 25 | 1 | 25 | .5 | 0 | .5 | .32 | .68 | 0 |
| 26 | 1 | 26 | .5 | 0 | .5 | .3 | .7 | 0 |
| 27 | 1 | 27 | .5 | 0 | .5 | .28 | .72 | 0 |
| 28 | 1 | 28 | .5 | 0 | .5 | .26 | .74 | 0 |
| 29 | 1 | 29 | .5 | 0 | .5 | .24 | .76 | 0 |
| 30 | 1 | 30 | .5 | 0 | .5 | .22 | .78 | 0 |
| 31 | 1 | 31 | .5 | 0 | .5 | .2 | .8 | 0 |
| 32 | 1 | 32 | .5 | 0 | .5 | .16 | .64 | 0 |
| 33 | 1 | 33 | .5 | 0 | .5 | .1 | .9 | 0 |

Figure 13.2: First 33 rows of design data set

The chosen values of $\mu$ and $\eta$ give rise to typical levels of risk aversion seen in experimental data elsewhere in the literature, and also a reasonable between-subject spread in risk aversion. The chosen values of $\sigma$ differ between the two models; this is because the models differ in terms of the scale of the variable to which the subject-specific error is being applied. The tremble probability is assumed to start at 0.06 and to decay in the course of the experiment, reaching around 0.02 after 100 choice tasks. The rank-dependent parameters $\beta$ and $\gamma$ are both a reasonable distance from 1, implying a clear deviation from EU.

Of course, when we come to estimate the models, we will pretend that we do not know the "true" parameter values given in the table above. Accordingly, we will "guess" a set of starting values that are somewhat different from the true values. In the code, the starting values will be placed in the row vector "start".

Another input to the simulation is the "design" data set. This data set contains eight columns: i, t, p1 − p3, q1 − q3. There are 6,000 rows: each of 60 subjects will carry out 50 tasks on two occasions. The first few rows of the design data are presented in Figure 13.2. The simulation algorithm will result in a number of columns being appended to this data set.

### *13.3.2 The STATA code*

```
* LIKELIHOOD EVALUATION PROGRAM FOR FECHNER MODEL STARTS HERE:

program define fechner
```

```
* DECLARE VARIABLES, PARAMETERS, AND VARIABLE LIST FOR HALTON DRAWS

args todo b logl
tempvar pp1 pp2 pp3 qq1 qq2 qq3 dd1 dd2 dd3 r w z p pp ppp rpp rppp
tempname mu ln_eta eta sig w0 w1 aa gg
local hlist h1*

* EXTRACT SCALAR PARAMETERS FROM VECTOR b.

mleval `mu' = `b', eq(1) scalar
mleval `eta' = `b', eq(2) scalar
mleval `sig' = `b', eq(3) scalar
mleval `w0'=`b', eq(4) scalar
mleval `w1'=`b', eq(5) scalar
mleval `aa'=`b', eq(6) scalar
mleval `gg'=`b', eq(7) scalar

* INITIALISE VARIABLES

quietly{
gen double `pp1'=.
gen double `pp2'=.
gen double `pp3'=.
gen double `qq1'=.
gen double `qq2'=.
gen double `qq3'=.
gen double `dd1'=.
gen double `dd2'=.
gen double `dd3'=.
gen double `r'=.
gen double `w'=`w0'*exp(`w1'*tau)
gen double `z'=.
gen double `p'=.
gen double `pp'=.
gen double `rpp'=.
gen double `ppp'=0
gen double `rppp'=0
}

quietly{

* TRANSFORM TRUE PROBABILITIES (p1,p2,p3,q1,q2,q3) USING PRELEC WEIGHTING FUNCTION.
* TRANSFORMED PROBABILITIES ARE pp1,pp2,pp3,qq1,qq2,qq3:

replace `pp3'=exp(-`aa'*(-ln(p3))^`gg')
replace `pp3'=0 if p3==0
replace `pp2'=exp(-`aa'*(-ln(p3+p2))^`gg')-`pp3'
replace `pp2'=0 if p2==0
replace `pp1'=1-`pp2'-`pp3'

replace `qq3'=exp(-`aa'*(-ln(q3))^`gg')
replace `qq3'=0 if q3==0
replace `qq2'=exp(-`aa'*(-ln(q3+q2))^`gg')-`qq3'
replace `qq2'=0 if q2==0
replace `qq1'=1-`qq2'-`qq3'

* GENERATE DIFFERENCE VARIABLES FROM TRANSFORMED PROBABILITIES:

replace `dd3'=`pp3'-`qq3'
replace `dd2'=`pp2'-`qq2'
replace `dd1'=`pp1'-`qq1'

* START LOOP OVER HALTON VARIABLES:

foreach v of varlist `hlist' {

* GENERATE LIKELIHOOD FOR EACH ROW:

replace `r'= exp(`mu'+ `eta'*`v')
replace `z'=(`dd1'*(x1^(`r'))+`dd2'*(x2^(`r'))+`dd3'*(x3^(`r')))/`sig'
replace  `p'= (1-`w')*normal(yy*`z')+`w'/2

* TAKE PRODUCT WITHIN EACH SUBJECT, AND PLACE THIS IN LAST ROW FOR EACH SUBJECT:
```



```
by i: replace `pp' = exp(sum(ln(max(`p',1e-12))))
replace `pp'=. if last~=1

* ADD COLUMN OF PRODUCTS TO THE CUMULATIVE SUM BUILDING UP BETWEEN LOOPS:

replace `ppp'=`ppp'+`pp'

* USE SIMILAR PROCEDURE TO CREATE VARIABLE FOR USE IN COMPUTATION OF
* POSTERIOR MEAN OF RISK ATTITUDE

replace `rpp'=`r'*`pp'
replace `rpp'=. if last~=1
replace `rppp'=`rppp'+`rpp'
}

* END OF LOOP

* FIND MEANS OF VARIABLES GENERATED BY LOOP:

replace `ppp'=`ppp'/draws
replace `rppp'=`rppp'/draws

replace lppp=ln(`ppp') if last==1
replace rhat=`rppp'/`ppp'
by i: replace rhat=rhat[_N] if rhat==.

* GENERATE SUBJECT-CONTRIBUTIONS TO LOG-LIKELIHOOD, AND DECLARE AS MAXIMAND

mlsum `logl'=ln(`ppp') if last==1

* USE PUTMATA TO MAKE SOME GLOBAL VARIABLES AVAILABLE OUTSIDE THE PROGRAM
* THE ddd_hat'S ARE NEEDED TO OBTAIN CLOSENESS TO INDIFFERENCE
* lppp IS NEEDED TO PERFORM NON-NESTED TESTS
* rhat IS POSTERIOR RISK ATTITUDE

replace dd1_hat=`dd1'
replace dd2_hat=`dd2'
replace dd3_hat=`dd3'

putmata lppp, replace
putmata rhat, replace

putmata dd1_hat, replace
putmata dd2_hat, replace
putmata dd3_hat, replace

}

end

* END OF FECHNER LIKELIHOOD EVALUATION PROGRAM

* LIKELIHOOD EVALUATION PROGRAM FOR RP STARTS HERE:

program define rp

* DECLARE VARIABLES, PARAMETERS, AND VARIABLE LIST FOR HALTON DRAWS

args todo b logl
tempvar pp1 pp2 pp3 qq1 qq2 qq3 dd1 dd2 dd3 r w astar z p pp ppp rpp rppp
tempname mu eta sig  w0 w1 aa gg
local hlist h1*

* EXTRACT SCALAR PARAMETERS FROM VECTOR b

mleval `mu' = `b', eq(1) scalar
mleval `eta' = `b', eq(2) scalar
mleval `sig' = `b', eq(3) scalar
mleval `w0'=`b', eq(4) scalar
mleval `w1'=`b', eq(5) scalar
mleval `aa'=`b', eq(6) scalar
mleval `gg'=`b', eq(7) scalar
```

```
* INITIALISE VARIABLES

quietly{
gen double 'pp1'=.
gen double 'pp2'=.
gen double 'pp3'=.
gen double 'qq1'=.
gen double 'qq2'=.
gen double 'qq3'=.
gen double 'dd1'=.
gen double 'dd2'=.
gen double 'dd3'=.
gen double 'w'='w0'*exp('w1'*tau)
gen double 'r'=.
gen double 'z'=.
gen double 'p'=.
gen double 'pp'=.
gen double 'rpp'=.
gen double 'ppp'=0
gen double 'rppp'=0
gen double 'astar'=0

* TRANSFORM TRUE PROBABILITIES (p1,p2,p3,q1,q2,q3) USING PRELEC WEIGHTING FUNCTION.
* TRANSFORMED PROBABILITIES ARE pp1,pp2,pp3,qq1,qq2,qq3:

replace 'pp3'=exp(-'aa'*(-ln(p3))^'gg')
replace 'pp3'=0 if p3==0
replace 'pp2'=exp(-'aa'*(-ln(p3+p2))^'gg')-'pp3'
replace 'pp2'=0 if p2==0
replace 'pp1'=1-'pp2'-'pp3'

replace 'qq3'=exp(-'aa'*(-ln(q3))^'gg')
replace 'qq3'=0 if q3==0
replace 'qq2'=exp(-'aa'*(-ln(q3+q2))^'gg')-'qq3'
replace 'qq2'=0 if q2==0
replace 'qq1'=1-'qq2'-'qq3'

* GENERATE DIFFERENCE VARIABLES FROM TRANSFORMED PROBABILITIES:

replace 'dd3'='pp3'-'qq3'
replace 'dd2'='pp2'-'qq2'
replace 'dd3'=1 if dom==1
replace 'dd2'=-2 if dom==1
replace 'dd1'='pp1'-'qq1'

* START LOOP OVER HALTON VARIABLES:

foreach v of varlist 'hlist' {

* GENERATE LIKELIHOOD FOR EACH ROW:

replace 'r'= 'mu'+ 'eta'*'v'
replace 'astar'=ln(-'dd2'/'dd3')/ln(2)
replace 'z'=('r'-ln('astar'))/'sig'
replace  'p'= (1-'w')*((1-dom)*normal(yy*'z')+dom*y)+'w'/2

* TAKE PRODUCT WITHIN EACH SUBJECT, AND PLACE THIS IN LAST ROW FOR EACH SUBJECT:

by i: replace 'pp' = exp(sum(ln(max('p',1e-12))))
replace 'pp'=. if last~=1

* ADD COLUMN OF PRODUCTS TO THE CUMULATIVE SUM BUILDING UP BETWEEN LOOPS:

replace 'ppp'='ppp'+'pp'

* USE SIMILAR PROCEDURE TO CREATE VARIABLE FOR USE IN COMPUTATION OF
* POSTERIOR MEAN OF RISK ATTITUDE

replace 'rpp'='r'*'pp'
replace 'rpp'=. if last~=1
replace 'rppp'='rppp'+'rpp'
}
```



```
* END OF LOOP

* FIND MEANS OF VARIABLES GENERATED BY LOOP:

replace 'ppp'='ppp'/draws
replace 'rppp'='rppp'/draws
replace lppp=ln('ppp') if last==1

* GENERATE SUBJECT-CONTRIBUTIONS TO LOG-LIKELIHOOD, AND DECLARE AS MAXIMAND

mlsum 'logl'=ln('ppp') if last==1

* USE PUTMATA TO MAKE GLOBAL VARIABLE lppp AVAILABLE OUTSIDE THE PROGRAM
* lppp IS NEEDED TO PERFORM NON-NESTED TESTS

putmata lppp, replace
}

end

* END OF RP LIKELIHOOD EVALUATION PROGRAM

* SIMULATION STARTS HERE

clear
set more off

* READ DATA SET CONTAINING SUBJECT NUMBER (i) TASK NUMBER (t), AND
* PROBABALITIES DEFINING LOTTERIES (p1-p3;q1-q3)

use "design.dta", clear

* SET RANDOM NUMBER SEED; RECAST VARIABLES TO CORRECT TYPE

set seed 91611143
recast int i t
recast double p* q*

* GENERATE SCALARS N AND T

summ i
scalar N=r(max)
summ t
scalar T=r(max)

* SET TRUE VALUES OF PARAMETERS FOR SIMULATION

scalar mu=-0.88
scalar eta=0.2
scalar sig_fecbner=0.05
scalar sig_rp=0.15
scalar w0=0.06
scalar w1=-0.01
scalar aa=0.9
scalar gg=0.8

* SIMULATE POSITION OF PROBLEM IN SEQUENCE (tau)
* REMEMBER THAT 50 PROBLEMS ARE SET IN TWO SITTINGS

by i: gen  int d50=_n>50
bysort i d50: egen tau_d=rank(uniform())
gen int tau=50*d50+tau_d

* SIMULATE RISK ATTITUDE FOR FECHNER MODEL;
* REMEMBER THAT THIS NEEDS TO BE FIXED WITHIN A SUBJECT

by i: generate double r_fechner=exp(mu+eta*(invnorm(uniform()))) if _n==1
by i: replace r_fechner=r_fechner[1] if r_fechner==.

* SIMULATE RISK ATTITUDE FOR RP MODEL;
* FIRST SIMULATE SUBJECT MEAN m FROM LOGNORMAL;
* THEN SIMULATE RISK ATTITUDE AS LOGNORMAL WITH MEAN PARAMETER m
```

```
by i: generate double  m=mu+eta*(invnorm(uniform())) if _n==1
by i: replace m=m[1] if m==.
gen double r_rp=exp(m+sig_rp*invnorm(uniform()))

* GENERATE TRANSFORMED PROBABILITIES FROM TRUE PROBABILITIES USING
* PRELEC WEIGHTING FUNCTION; TRANSFORMED PROBABILITIES ARE (pp1-pp3; qq1-qq3)

gen double pp3=exp(-aa*(-ln(p3))^gg)
replace pp3=0 if p3==0
gen double pp2=exp(-aa*(-ln(p3+p2))^gg)-pp3
replace pp2=0 if p2==0
gen double pp1=1-pp2-pp3

gen double qq3=exp(-aa*(-ln(q3))^gg)
replace qq3=0 if q3==0
gen double qq2=exp(-aa*(-ln(q3+q2))^gg)-qq3
replace qq2=0 if q2==0
gen double qq1=1-qq2-qq3

* GENERATE 3 (CONSTANT) VARIABLES REPRESENTING THE THREE MONEY AMOUNTS

gen double x1=0
gen double x2=1
gen double x3=2

* COMPUTE VALUATION OF THE TWO LOTTERIES UNDER FECHNER:

gen double vp_fechner=pp1*(x1^r_fechner)+pp2*(x2^r_fechner)+pp3*(x3^r_fechner)
gen double vq_fechner=qq1*(x1^r_fechner)+qq2*(x2^r_fechner)+qq3*(x3^r_fechner)

* COMPUTE VALUATION OF THE TWO LOTTERIES UNDER RP:

gen double  vp_rp=pp1*(x1^r_rp)+pp2*(x2^r_rp)+pp3*(x3^r_rp)
gen double vq_rp=qq1*(x1^r_rp)+qq2*(x2^r_rp)+qq3*(x3^r_rp)

* GENERATE TREMBLE PROBABILITY (AS DECAYING FUNCTION OF tau)
* THEN SIMULATE TREMBLE INDICATOR (trem)

gen double w=w0*exp(w1*tau)
gen int trem=(uniform()<w)

* SIMULATE CHOICE FROM FECHNER MODEL; EXPRESS AS BOTH y AND yy:

gen int y_fechner=(1-trem)*((vp_fechner-vq_fechner ///
+sig_fechner*invnorm(uniform()))>0)+trem*(uniform()>0.5)
gen int yy_fechner=2*y_fechner-1

* SIMULATE CHOICE FROM RP MODEL; EXPRESS AS BOTH y AND yy:

gen int y_rp=(1-trem)*((vp_rp-vq_rp)>0)+trem*(uniform()>0.5)
gen int yy_rp=2*y_rp-1

* END OF SIMULATION

* ESTIMATION STARTS HERE:

* GENERATE INDICATOR FOR DOMINANCE PROBLEMS (dom):

gen double d1=p1-q1
gen double d2=p2-q2
gen double d3=p3-q3
gen int dom = (d3>=0)*((d3+d2)>=0)

* GENERATE INDICTATORS FOR FIRST AND LAST OBSERVATIONS FOR EACH SUBJECT:

by i: generate int  first=1 if _n==1
by i: generate int last=1 if _n==_N

* GENERATE HALTON DRAWS (31 COLUMNS) USING mdraws; CREATE VARIABLE LIST

mat p=[3]
mdraws if first==1 , neq(1) dr(31) prefix(h) primes(p) burn(3)
```



```
scalar draws=r(n_draws)
local hlist h1*

* ENSURE THAT HALTON DRAWS ARE IN DOUBLE PRECISION

recast double h1*

* IMPOSE PANEL STRUCTURE ON HALTON DRAWS: T ROWS ALL SAME WITHIN SUBJECT:

quietly{
foreach v of varlist 'hlist' {
by i: replace 'v'='v'[1] if 'v'==.
replace 'v'=invnorm('v')
}
}

* INITALISE VARIABLES REPRESENTING ESTIMATED DIFFERENCES OF
* TRANSFORMED PROBABILITIES

gen double dd1_hat=.
gen double dd2_hat=.
gen double dd3_hat=.

* INITIALISE VARIABLES REPRESENTING: PER-SUBJECT LOG-LIKELIHOOD;
* POSTERIOR RISK ATTITUDE

gen double lppp=.
gen double rhat=.

* INITIALISE CHOICE VARIABLES (y AND yy)

gen int y=.
gen int yy=.

* ASSIGN FECHNER CHOICE DATA TO y AND yy

replace yy=yy_fechner
replace y=y_fechner

* SET STARTING VALUES OF FECHNER MODEL:

mat start=(-0.68, 0.12, 0.10, 0.04, -0.005, 1.0, 1.0)

* ESTIMATE FECHNER MODEL ON FECHNER DATA USING ML

ml model d0 fechner /mu /eta  /sig /w0  /w1  /aa /gg
ml init start, copy
ml max

* EXTRACT GLOBAL VARIABLES FROM INSIDE FECHNER PROGRAM

drop lppp rhat dd1_hat dd2_hat dd3_hat
getmata lppp
getmata rhat
getmata dd1_hat dd2_hat dd3_hat

* RENAME VARIABLE CONTAINING PER-SUBJECT LOG-LIKELIHOOD

rename lppp lp_fechner

* GENERATE VALUATION DIFFERENTIAL

gen double diff=dd1_hat*(x1^rhat)+dd2_hat*(x2^rhat)+dd3_hat*(x3^rhat)

* RE-INITIALISE VARIABLE CONTAINING PER-SUBJECT LOG-LIKELIHOOD

gen lppp=.

* ESTIMATE RP MODEL ON FECHNER DATA USING ML

ml model d0 rp /mu /eta  /sig /w0  /w1  /aa /gg
ml init start, copy
ml max
```

```
* EXTRACT VARIABLE CONTAINING PER-SUBJECT LOG-LIKELIHOOD FROM INSIDE RP PROGRAM

drop lppp
getmata lppp
rename lppp lp_rp

* CARRY OUT VUONG'S NON-NESTED TEST (FOR MODELS ESTIMATED WITH FECHNER DATA):

gen vuong1=  lp_fechner - lp_rp
summ vuong1
scalar vuong=(sqrt(r(N))*r(mean))/(r(sd))
scalar list vuong

* CARRY OUT CLARKE'S NONPARAMETRIC NON-NESTED TESTS
* (FOR MODELS ESTIMATED WITH FECHNER DATA):

signrank lp_fechner = lp_rp
signtest lp_fechner = lp_rp

* NOW SWITCH TO RP DATA

* RE-INITIALISE VARIABLE CONTAINING PER-SUBJECT LOG-LIKELIHOOD

gen double lppp=.

* ASSIGN RP CHOICE DATA TO y AND yy

replace yy=yy_rp
replace y=y_rp

* ESTIMATE FECHNER MODEL ON RP DATA USING ML

ml model d0 fechner /mu /eta  /sig /w0  /w1  /aa /gg
ml init start, copy
ml max

* EXTRACT VARIABLE CONTAINING PER-SUBJECT LOG-LIKELIHOOD
* FROM INSIDE FECHNER PROGRAM

drop lppp lp_fechner
getmata lppp
rename lppp lp_fechner

* RE-INITIALISE VARIABLE CONTAINING PER-SUBJECT LOG-LIKELIHOOD

gen double lppp=.

* ESTIMATE RP MODEL ON RP DATA USING ML

ml model d0 rp /mu /eta  /sig /w0  /w1  /aa /gg
ml init start, copy
ml max

* EXTRACT VARIABLE CONTAINING PER-SUBJECT LOG-LIKELIHOOD FROM INSIDE RP PROGRAM

drop lppp lp_rp
getmata lppp
rename lppp lp_rp

* CARRY OUT VUONG'S NON-NESTED TEST (FOR MODELS ESTIMATED WITH RP DATA):

replace vuong1=  lp_fechner - lp_rp
summ vuong1
scalar vuong=(sqrt(r(N))*r(mean))/(r(sd))
scalar list vuong

* CARRY OUT CLARKE'S NONPARAMETRIC NON-NESTED TESTS
* (FOR MODELS ESTIMATED WITH RP DATA):

signrank lp_rp = lp_fechner
signtest lp_rp = lp_fechner
```



```
* FINALLY, SIMULATE DECISION TIMES
* EXPLANATORY VARIABLES FOR DECISION TIME NEED TO BE GENERATED

* GENERATE COMPLEXITY LEVELS OF CHOICE PROBLEMS;
* VARIABLE "complex" IS NUMBER OF OUTCOMES IN SIMPLER LOTTERY

gen complex_p=(p1>0)+(p2>0)+(p3>0)
gen complex_q=(q1>0)+(q2>0)+(q3>0)
gen complex=min(complex_p, complex_q)
gen complex2=complex==2
gen complex3=complex==3

* GENERATE LOG OF EXPECTED VALUE OF SAFE LOTTERY

gen logev=ln(q1*0+q2*1+q3*2)

* GENERATE ABSOLUTE VALUATION DIFFERENTIAL, AND ITS SQUARE AND CUBE

gen abs_diff=abs(diff)
gen abs_diff2=abs_diff^2
gen abs_diff3=abs_diff^3

* GENERATE MEASURE OF OBJECTIVE DIFFERENCE BETWEEN LOTTERIES

gen obj_diff=d1^2+d2^2+d3^2

* SIMULATE SUBJECT-SPECIFIC RANDOM EFFECT (dt1) FOR DECISION TIME:

by i: generate double  dt1=1.624+0.386*(invnorm(uniform())) if _n==1
by i: replace dt1=dt1[1] if dt1==.

* SIMULATE LOG OF DECISION TIME:

gen dt2=dt1+0.267*complex2+0.388*complex3-0.0019*(tau_d-1) ///
-0.0032*(tau-1)+0.028*logev-5.251*abs_diff+10.944*abs_diff2 ///
-7.339*abs_diff3+0.157*obj_diff+0.616*invnorm(uniform())

* GENERATE DECISION TIME:

gen dt=exp(dt2)
```

### 13.3.3 The simulated data

The simulated data set is saved in the file **risky_choice_sim**. Below, we show tabulations of the two simulated choice variables. We see that, in both cases, choices are roughly equally divided between "safe" and "risky". This is clearly a desirable feature of a choice dataset, which is consistent with a good experimental design. However, it must be said that, for reasons to be explained fully in Chapter 14, this is not a sufficient condition for a good design. We also need to look at other features of the data, such as the variation over subjects in the number of risky choices made. This we do in Figure 13.3. We see that in both cases there is a healthy between-subject spread, providing further evidence of a good design.

```
. tab y_fechner

  y_fechner |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      2,909       48.48       48.48
          1 |      3,091       51.52      100.00
------------+-----------------------------------
      Total |      6,000      100.00
```

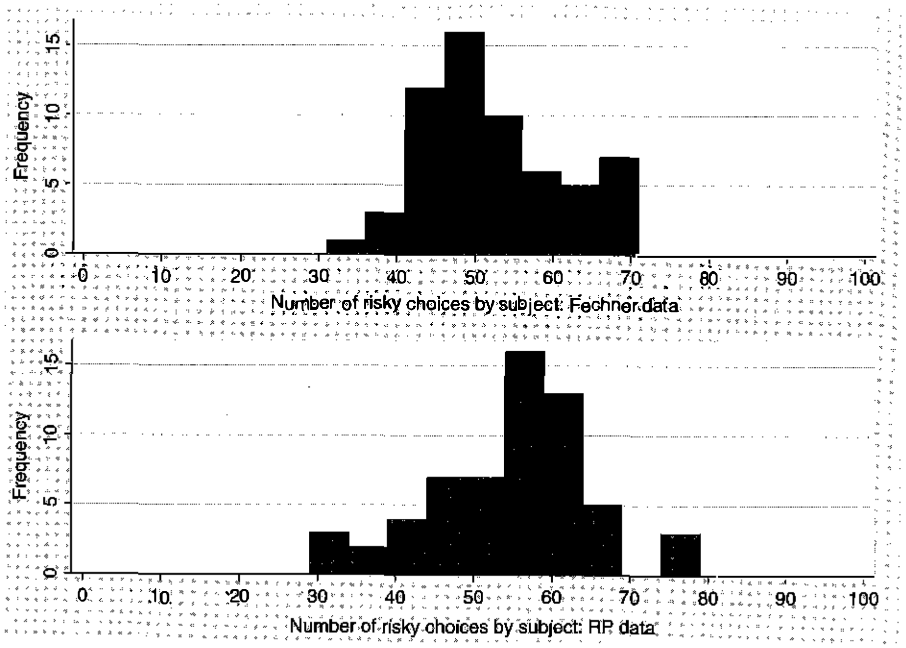



Figure 13.3: Distribution of number of risky or dominating choices made over the 60 subjects, from the Fechner data (upper pane), and the RP data (lower pane)

```
. tab y_rp

       y_rp |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      2,728       45.47       45.47
          1 |      3,272       54.53      100.00
------------+-----------------------------------
      Total |      6,000      100.00
```

### 13.3.4 *Output from estimation routines*

The two simulated sets of data have been subjected to estimation by both models, Fechner and RP. This gives four sets of estimates: Fechner estimation on Fechner data; RP estimation on Fechner data; Fechner estimation on RP data; RP estimation on RP data.

In order to give a feel for the estimation process, we present below the output from estimation of the Fechner model on Fechner data. In the following subsection, we shall collect together the results from all estimations.

```
. * SET STARTING VALUES OF FECHNER MODEL:

. mat start=(-0.68, 0.12, 0.10, 0.04, -0.005, 1.0, 1.0)

. ml model d0 fechner /mu /eta  /sig /w0  /w1  /aa /gg
```



```
. ml init start, copy

. ml max

initial:       log likelihood = -2225.4922
rescale:       log likelihood = -2225.4922
rescale eq:    log likelihood = -2179.1364
Iteration 0:   log likelihood = -2179.1364  (not concave)
Iteration 1:   log likelihood = -1966.5268
Iteration 2:   log likelihood =  -1954.092
Iteration 3:   log likelihood = -1947.7712
Iteration 4:   log likelihood = -1947.2029
Iteration 5:   log likelihood = -1947.1964
Iteration 6:   log likelihood = -1947.1964

                                                  Number of obs   =       6000
                                                  Wald chi2(0)    =          .
Log likelihood = -1947.1964                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
mu           |
       _cons |  -.8666559   .0505792   -17.13   0.000    -.9657893   -.7675224
-------------+----------------------------------------------------------------
eta          |
       _cons |   .1846398   .0164323    11.24   0.000     .1524332    .2168464
-------------+----------------------------------------------------------------
sig          |
       _cons |   .0514486   .0027959    18.40   0.000     .0459686    .0569285
-------------+----------------------------------------------------------------
w0           |
       _cons |   .0615979   .0165882     3.71   0.000     .0290857    .0941101
-------------+----------------------------------------------------------------
w1           |
       _cons |  -.0091638   .0050211    -1.83   0.068     -.019005    .0006774
-------------+----------------------------------------------------------------
aa           |
       _cons |   .9149991   .0436093    20.98   0.000     .8295264    1.000472
-------------+----------------------------------------------------------------
gg           |
       _cons |   .8044301   .0085684    93.88   0.000     .7876363     .821224
------------------------------------------------------------------------------
```

We see that convergence is achieved in six iterations. On a 3.2 GHz computer, this process of convergence takes between one and two minutes.

## 13.4 Results and Post-estimation

### 13.4.1 *The model estimates*

Two models have each been estimated on two data sets. All four sets of estimates are presented in Table 13.1, alongside the true parameter values used in the simulations. The first important observation is that the estimates are all close to the true parameter values when the correct model is estimated. This is, of course, expected, and is simply an indication that the routine has been coded correctly. When the incorrect model is estimated, some of the parameters are poorly estimated. For example, it appears to be the case that the mean of the risk attitude distribution ($\mu$) is underestimated (i.e. subjects are judged to be too risk-averse) when the RP model is applied

| | Fechner data | | | RP data | | |
|---|---|---|---|---|---|---|
| | True | Fechner est. | RP est. | True | Fechner est. | RP est. |
| $\mu$ | −0.88 | −0.886(0.051) | −0.954(0.054) | −0.88 | −0.808(0.042) | −0.889(0.035) |
| $\eta$ | 0.20 | 0.185(0.016) | 0.198(0.020) | 0.20 | 0.226(0.014) | 0.223(0.013) |
| $\sigma$ | 0.05 | 0.051(0.003) | 0.284(0.013) | 0.15 | 0.028(0.002) | 0.156(0.007) |
| $\omega_0$ | 0.06 | 0.062(0.017) | 0.121(0.019) | 0.06 | 0.050(0.011) | 0.048(0.010) |
| $\omega_1$ | −0.01 | −0.009(0.005) | −0.008(0.003) | −0.01 | −0.003(0.004) | −0.007(0.004) |
| $\beta$ | 0.90 | 0.915(0.044) | 0.861(0.045) | 0.90 | 0.956(0.036) | 0.902(0.029) |
| $\gamma$ | 0.80 | 0.804(0.009) | 0.825(0.010) | 0.80 | 0.781(0.006) | 0.786(0.007) |
| $n$ | | 60 | 60 | | 60 | 60 |
| $T$ | | 100 | 100 | | 100 | 100 |
| $LogL$ | | −1947.20 | −2071.02 | | −1361.25 | −1288.13 |
| Vuong | | +8.61 | | | −6.19 | |

Table 13.1: MLEs of parameters of Fechner and RP models estimated using Fechner and RP data

to Fechner data, and the same parameter is overestimated (i.e. subjects are judged to be too risk-seeking) when the Fechner model is applied to RP data. Another example is that the tremble probability appears to be seriously overestimated when the RP model is applied to Fechner data.

The biases that arise when the wrong model is estimated, and the desire to identify which is the "correct" error story, leads us to search for methods for adjudicating between the two models. We start by noting that, for each data set, the correct model results in a considerably higher maximised log-likelihood. One obvious approach is to apply a test that formalises this comparison. Since the two models are non-nested, the appropriate test is the Vuong (1989) test. We will also consider non-parametric analogues of the Vuong test.

### 13.4.2 Vuong's non-nested likelihood ratio test

The Vuong (1989) test has been applied to the problem of testing between competing error specifications in the context of risky choice by Loomes et al. (2002), and others.

Consider any two non-nested models 1 and 2, containing the same number of unknown parameters. Let $\hat{f}_i$ be the estimated probability of observing the $T$ actual choices made by subject $i$, on the assumption that model 1 is the true model. Let $\hat{g}_i$ be the estimated probability of observing the same $T$ choices on the assumption that model 2 is the true model. The Vuong test is based on the quantity $D$, defined by:

$$D = n^{-1/2} \sum_{i=1}^{n} \ln\left(\frac{\hat{f}_i}{\hat{g}_i}\right) \tag{13.30}$$



$D$ defined in (13.30) is similar to the log-likelihood ratio of the two models, but since the models are non-nested, it can be of either sign (if it were the case that model 1 nests model 2, $D$ would always be positive). To implement the test, we need to estimate the variance of $D$. An appropriate variance estimator is:

$$\hat{V} = n^{-1} \sum_{i=1}^{n} \left( \left[\ln\left(\frac{\hat{f}_i}{\hat{g}_i}\right)\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{\hat{f}_i}{\hat{g}_i}\right)\right]^2 \right) \tag{13.31}$$

The Vuong test statistic is then:

$$Z = \frac{D}{\sqrt{\hat{V}}} \tag{13.32}$$

As proved by Vuong (1989), the statistic $Z$ defined in (13.32) has a limiting standard normal distribution under the hypothesis that the two models are equivalent. A significantly positive value of $Z$ indicates that model 1 is closer to the true data generating process than model 2, while a significantly negative value of $Z$ indicates the converse.

In order to apply the test, we need to obtain the quantities $\hat{f}_i$ and $\hat{g}_i$ for $i = 1, \ldots, n$. In both the Fechner and RP models, these quantities are given by formulae of the form:

$$\int_0^\infty \prod_{t=1}^{T} \hat{p}_{it}(r) f(r; \hat{\mu}, \hat{\eta}) dr \tag{13.33}$$

where hats indicate evaluation at the MLE.

The two variables $\ln(\hat{f}_i)$ and $\ln(\hat{g}_i)$ have been computed as "`lp_fechner`" and "`lp_rp`" in the program. The Vuong test statistic is based on the mean of the difference between these two variables.

Vuong test statistics are presented in the final row of Table 13.1. Recall that the "null distribution" for the test is standard normal. We see that when the Fechner model is the true model, the Vuong test favours it strongly, with a significantly positive statistic of 8.61. Likewise, when the RP model is the true model, the Voung test favours it strongly, with a test statistic of −6.19.

### 13.4.3 Clarke's non-parametric non-nested test

Clarke (2003) proposed a non-parametric alternative to the Vuong test introduced in the last subsection. While the Vuong Test essentially analyses the mean of the differences between the two log-likelihoods, the non-parametric test analyses the median of the same differences. There are two non-parametric tests that can be used for this purpose, both of which were introduced in Section 3.7.3: the signed-rank test, and the sign test. The former ranks the absolute differences, and then compares the ranks between positive and negative differences; the latter simply counts the number of positive and negative differences. As pointed out in Section 3.7.3, the signed-rank test relies on an assumption of symmetry in the distribution of differences, while the sign test relies on no such assumption. According to Clarke (2003), there is no reason to expect the symmetry assumption to hold in differences of log-likelihoods of competing models, and so the sign test is preferred.

Here, we shall conduct both tests. When applied to the data generated from a Random Preference model, the output from the two test commands are:

```
. signrank lp_rp = lp_fechner

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |       44        1574         915
    negative |       16         256         915
        zero |        0           0           0
-------------+---------------------------------
         all |       60        1830        1830

unadjusted variance    18452.50
adjustment for ties        0.00
adjustment for zeros       0.00
                     ----------
adjusted variance      18452.50

Ho: lp_rp = lp_fechner
             z =   4.851
    Prob > |z| =   0.0000

. signtest lp_rp = lp_fechner

Sign test

        sign |    observed    expected
-------------+------------------------
    positive |          44          30
    negative |          16          30
        zero |           0           0
-------------+------------------------
         all |          60          60

One-sided tests:
  Ho: median of lp_rp - lp_fechner = 0 vs.
  Ha: median of lp_rp - lp_fechner > 0
      Pr(#positive >= 44) =
         Binomial(n = 60, x >= 44, p = 0.5) =  0.0002

  Ho: median of lp_rp - lp_fechner = 0 vs.
  Ha: median of lp_rp - lp_fechner < 0
      Pr(#negative >= 16) =
         Binomial(n = 60, x >= 16, p = 0.5) =  0.9999

Two-sided test:
  Ho: median of lp_rp - lp_fechner = 0 vs.
  Ha: median of lp_rp - lp_fechner != 0
      Pr(#positive >= 44 or #negative >= 44) =
         min(1, 2*Binomial(n = 60, x >= 44, p = 0.5)) =  0.0004
```

Unsurprisingly, and in agreement with the Vuong test, both of the non-parametric non-nested tests result in a strong preference for the Random Preference model (i.e. for the true model), with p-values of 0.0000 for the signrank test, and 0.0004 for the sign test. Similar conclusions are drawn when the true model is Fechner: both tests result in strong evidence favouring the true model.

While there are clearly no surprises in the results discussed in this (and the previous) sub-section, it has nevertheless been useful to convey an idea of the variety of tests that are available for adjudicating between non-nested models.



### *13.4.4 Obtaining individual risk attitudes*

In the estimation of $\mu$ and $\eta$, the model has revealed the distribution of risk attitudes over the population. For certain purposes, it is very useful to have a measure of risk attitude for each individual subject. As explained earlier, it is more appropriate to use the Fechner model for this purpose, being as this model assumes that each individual has their own unique risk attitude parameter which remains the same over time.

Having estimated the Fechner model, we have appled Bayes rule to obtain the posterior expectation of each subject's risk attitude, conditional on their $T$ choices, using formula (13.29). The result of implementing this formula in the program is the variable "`rhat`". The distribution of rhat over the sample of 60 subjects is shown in Figure 13.4.

### *13.4.5 Obtaining closeness to indifference*

Finally, using the estimates of risk attitude for each subject, we are able to estimate the valuation differential perceived by each subject in each choice problem.

This requires substitution of the posterior estimate of risk attitude, $\hat{r}_i$ defined in (13.29), into the (transformed) valuation differential formula defined in (13.9):

$$\hat{\hat{\nabla}}_{it}(\hat{r}_i) = \hat{\hat{d}}_{2t} + \hat{\hat{d}}_{3t}2^{\hat{r}_i} \tag{13.34}$$

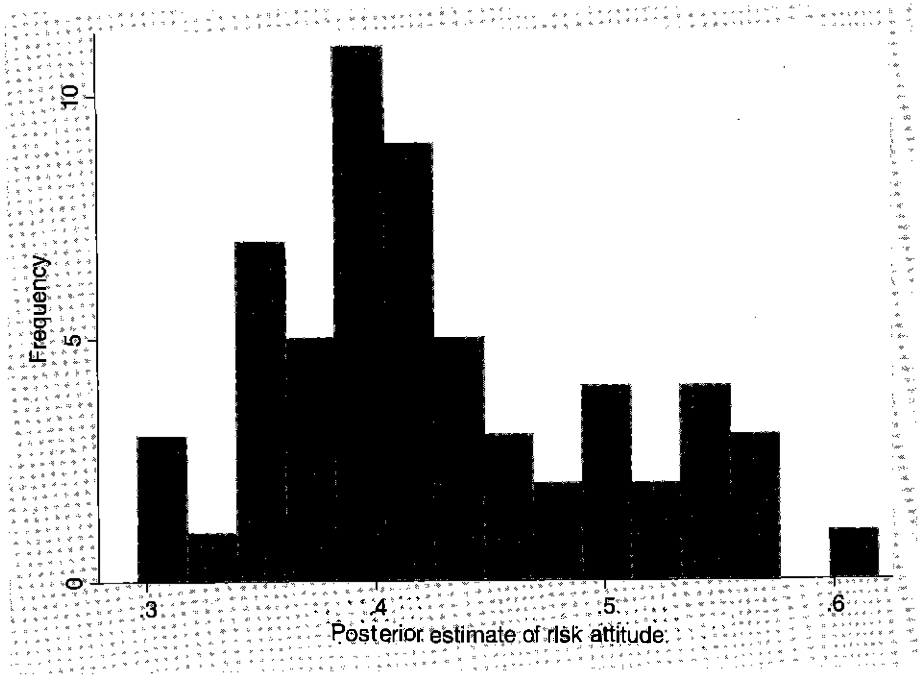



Figure 13.4: Distribution of posterior estimates of risk attitudes over individual subjects

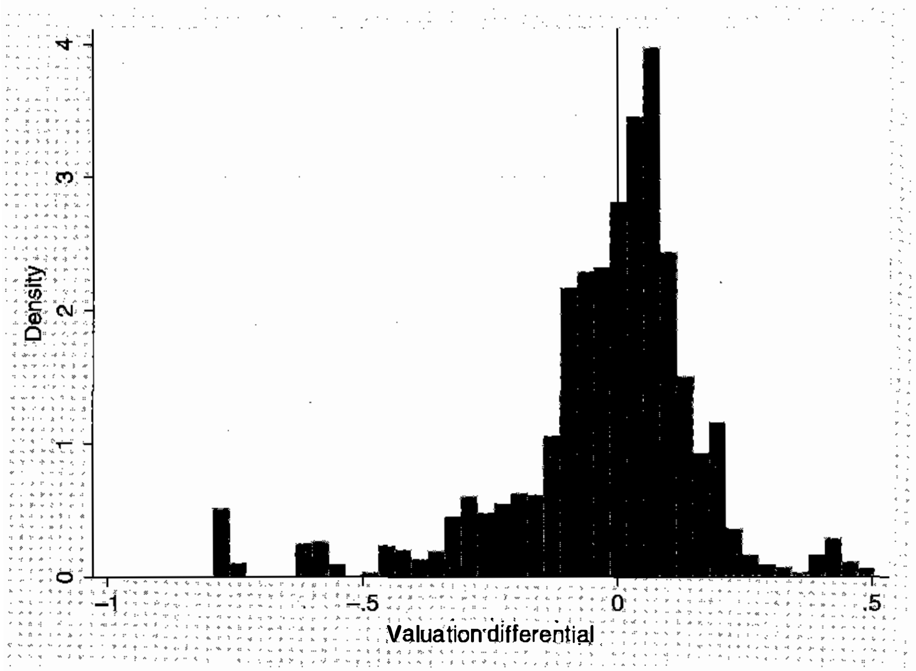



Figure 13.5: Distribution of the valuation differential

Equation (13.34) is generated in the program as the variable "diff". Figure 13.5 shows a histogram of this variable. We see that this distribution is broadly symmetric with mode close to zero. This simply tells us that for the majority of choice problems, subjects are reasonably close to indifference.

The absolute value of the valuation differential, $|\hat{\nabla}_{it}|$, is the variable that was used as the explanatory variable representing "closeness-to-indifference" in the model of decision times in Chapter 5.

### *13.4.6 Simulation of decision times*

The final part of the STATA code is a simulation of decision times for the simulated choices. These decision times are simulated from a data generating process (DGP) based on results from Moffatt's (2005*b*) analysis of real decision time data. The DGP is, in fact, as follows:

$$\begin{aligned}\log(decision\ time_{it}) &= 1.624 + 0.267complex2_t + 0.388complex3_t \\ &\quad - 0.0019(\tau_{it}^d - 1) - 0.0032(\tau_{it} - 1) + 0.028log(EV_t) \\ &\quad - 5.251|\hat{\Delta}_{it}| + 10.944|\hat{\Delta}_{it}|^2 - 7.339|\hat{\Delta}_{it}|^3 \\ &\quad + 0.157\Delta_t^o + u_i + \epsilon_{it} \\ i = 1\ldots,60 \quad t &= 1\ldots,100\ var(u_i) = 0.386^2 \quad var(\epsilon_{it}) = 0.616^2\end{aligned} \tag{13.35}$$



The first two explanatory variables in (13.35) are dummy variables indicating the level of complexity of problem $t$, according to the rule defined in Section 5.4: complexity is simply the number of outcomes in the simpler lottery. The excluded complexity level is the least complex, level 1. The third and fourth explanatory variables represent the position of a problem in the sequence: $\tau_{it}^d$ is the position of problem $t$ within the day on which the problem was solved, so $\tau^d$ ranges from 1 through 50; $\tau_{it}$ is in contrast the position of problem $t$ in the complete sequence of problems faced by subject $i$, and therefore ranges from 1 through 100. The fifth explanatory variable represents the logarithm of the expected value of the simpler of the two lotteries, which is used to represent the financial incentive associated with each problem. The next three explanatory variables are the closeness to indifference of subject $i$ in choice problem $t$, defined in (13.34) above, and its square and cube. The purpose of these three variables is to allow closeness to indifference to have a non-linear effect on effort, as explained in Section 5.4. The final explanatory variable is the measure of the objective difference between the two lotteries, defined in Section 5.3.

## 13.5 Summary and Further Reading

The central theme of this chapter has been estimation of risky choice models using repeated data on lottery choices by subjects. There is a large literature on the estimation of individual risk attitudes from lottery choice data, apparently starting with Binswanger (1980).

The models estimated here, and the methods used to estimate them, bear similarities to a number of studies, including Harless & Camerer (1994), Hey & Orme (1994), Loomes et al. (2002), Harrison & Rutström (2009), Conte et al. (2011) and Von Gaudecker et al. (2011). The last two of these studies, like us, use MSL with Halton draws. A difference from our own estimation is that they both have more than one dimension of heterogeneity. Conte et al. (2011) assume that both risk aversion and probability weighting vary continuously over subjects; Von Gaudecker et al. (2011) assume between-subject variation in four different parameters. This requires multiple sets of Halton draws.

Technical issues relating to the estimation of the tremble parameter have been explained by Moffatt & Peters (2001).

## Exercises

1. Some risky choice experiments allow subjects to express indifference. Extend the likelihood function to deal with indifference. Hint: treat the data as ordinal with three outcomes (see Chapter 7) and introduce a threshold parameter $\kappa$.
2. How would you go about computing a posterior tremble probability for each individual decision?
3. Starmer (2000) noted the growing evidence in the literature that choices under risk tend to change with experience, and he poses the question (p.376): "[are]

individuals discovering expected utility preferences?" How could the models developed in this chapter be extended to allow a test of this hypothesis? Think in particular of allowing the weighting parameters to depend on experience.

4. Consider a population of subjects of whom a proportion p follow EU, and the remaining $(1-p)$ follow RD, with weighting parameters varying continuously between them. Construct the likelihood function for the resulting mixture model. You may find it useful to consult Conte et al. (2011).

# Chapter 14
# Optimal Design in Binary Choice Experiments

## 14.1 Introduction

Chapter 2 was on the general topic of experimental design in experimental economics. The key themes covered in that chapter were randomisation techniques and the use of power analysis to find the required sample size for treatment tests. This chapter returns to the topic of experimental design. However, it focuses on design in binary choice models. It appears at this stage of the book for the simple reason that material in earlier chapters is prerequisite for an understanding of the techniques used in this chapter.

One area in which there is a particularly pressing need for guidance on optimal designs is in risky choice experiments. Many such experiments have been conducted, with a variety of objectives including the measurement of risk attitude and testing different theories of choice under risk. As explained in Chapter 13, these experiments follow a standard format, in which subjects are presented with a sequence of binary choice problems in which it is usually the case that one of the alternatives is "safer" and the other is "riskier". See, for example, Holt & Laury (2002). Each choice problem is defined by the possible outcomes, and corresponding probabilities, in each alternative. When it comes to choosing the outcomes and probabilities, most experimenters have followed an informal approach. The following quote from Hey & di Cagno (1990, p.286) is typical of the attitude to the design problem held by most researchers in the area:

> We decided on 60 questions as we thought that this would be the most we could reasonably ask within the subjects' attention span ... The choice of questions was not so easy ... We tried to get a mixture so that the slope of the line joining the pair of gambles varied considerably: from 1/7 to 7. The idea behind this was that we would then be able to distinguish between very risk-averse people and not-very-risk averse people, but we were rather groping in the dark.

The "slope" referred to in this quote is that of the line connecting two lotteries in the Marschak-Machina triangle, and is loosely related to the "threshold risk aversion" measure, $r^*$, used in the context of the Random Preference (RP) model in Chapter 13. Although, as evidenced in the above quote, care is often taken by the experimenter in setting the choice problems in such a way as to meet the objectives of the study, it is fair to say that formal techniques have rarely been applied to this experimental design problem.

One would be justified in asking why experimenters rely on dichotomous choice problems at all, if they are concerned with precision in estimation. In theory, a certainty equivalent of a single lottery has more informational content than an indication of preference of one lottery over another. Following this reasoning through, it might be asserted that the optimal design is one which elicits certainty equivalents for a sequence of suitably chosen lotteries. However, as hinted in Section 6.6.2, we firmly believe that true preferences are more likely to be revealed by choices than by valuations. Others evidently share this belief: most of the work on the estimation of risk attitudes is based on data on choices between lotteries.[1]

Given the wide acceptance of dichotomous choice as a means of eliciting preferences, it is important for researchers to have guidance on appropriate design. In particular, it is desirable to have a clear framework for choosing the parameters of the lottery pairs. Such a framework comes from the statistical literature on optimal experimental design. The purpose of this chapter is to address the question of how optimal experimental design theory can be applied to the design of economic experiments, with risky choice experiments as the central focus.

In the statistical literature, the concept of an optimal experimental design is usually taken to mean a design that gives maximal precision in the estimation of the parameters of interest. The determinant of Fisher's information matrix is often chosen as the criterion for optimal design. Maximising this quantity is equivalent to minimising the volume of the "confidence ellipsoid" surrounding the point estimates, and hence there is a strongly intuitive sense in which this criterion maximises precision in estimation. A design for which the determinant of the information matrix is maximised is referred to as a "D-optimal design".

The problem of D-optimal design in a linear model is simple, at least in principle, because the criterion does not involve the model's parameters. The optimal design can therefore be found without knowledge of the parameters. In contrast, in non-linear models, in which we are more interested since we are working in the context of binary data, the determinant of the information matrix is a function of the parameters of the model. Therefore, strictly speaking, the values of the parameters need to be known in order for the D-optimal design to be found. This problem is known as the "chicken and egg problem". At first sight, this may appear as quite a serious problem: our ultimate objective is to estimate a set of parameters with as much precision as possible, but in order to establish the method for achieving this, we require to know the values of these parameters! However, on further reflection, we become less concerned because, although the parameter values are unknown, estimates of the parameters are usually available from previous studies, and it is considered acceptable to develop the optimal design on the basis of these estimates.

[1] See, for example, Hey & Orme (1994), Loomes et al. (2002), Harrison & Rutström (2009), Conte et al. (2011) and Von Gaudecker et al. (2011).



It is in this spirit that we develop optimal designs in the context of risky choice experiments. In fact, we take estimates obtained and reported in earlier chapters within the book, and proceed with the optimal design exercise on the basis of these estimates.

Section 14.2 sets out the rudiments of optimal design theory, in particular defining the concept of D-optimality, and then applies this concept to linear, probit and logit models. Section 14.3 specifies the parametric model of risky choice to which the optimal design methodology is to be applied. In Section 14.4, we apply the principle of D-optimal design introduced in Section 14.2 to the risky choice model developed in Section 14.3, hence obtaining a set of choice problems that is optimal for a particular subject whose risk attitude parameters are known. Section 14.5 is concerned with the way in which the design should be adjusted if the model contains a "tremble parameter". Section 14.6 extends the optimal design to a situation in which the experiment involves multiple subjects, taking account of subject heterogeneity.

## 14.2 Rudiments of Experimental Design Theory

### *14.2.1 The principle of D-optimal design*

Consider a model in which the scalar dependent variable is $y$, the single explanatory variable is $x$, and the probability, or probability density, associated with a particular observation $(y_i, x_i)$ is $f(y_i|x_i;\theta)$, where $\theta$ is a $k \times 1$ vector of parameters. Assume that there are a total of $n$ independent observations. The log-likelihood function for this model is:

$$\text{LogL} = \sum_{i=1}^{n} \ln f(y_i|x_i;\theta) \tag{14.1}$$

The maximum likelihood estimate (MLE) of the parameter vector $\theta$ is the value that maximises LogL. The information matrix is given by:

$$I = E\left(\frac{\partial^2 \text{LogL}}{\partial\theta\partial\theta'}\right) = E\left(\frac{\partial \text{LogL}}{\partial\theta}\frac{\partial \text{LogL}}{\partial\theta'}\right) \tag{14.2}$$

The variance of the MLE is given by the inverse of the information matrix (14.2). Hence standard errors of individual estimates are obtained from the square roots of the diagonal elements of $I^{-1}$.

The principle of D-optimal design is simply to select values of $x_i$, subject to the specified constraints, that maximise the determinant of the information matrix. This is equivalent to minimising the volume of the "confidence ellipsoid" of the parameters contained in $\theta$, that is, estimating the entire set of parameters with maximal overall precision. When $\theta$ contains only one element, that is, when the model has only one parameter, meeting the D-optimal design criterion is equivalent simply to minimising the width of the confidence interval associated with that parameter.

Clearly, the information matrix and its determinant increase with the sample size $n$. Often, when we are comparing designs, we need to adjust for the number of observations, so we divide the information matrix by $n$ to obtain the "per-observation information matrix": $i = \frac{1}{n}I$.

### 14.2.2 Simple linear regression

Consider the simple (normal) regression model:

$$\begin{aligned} &y_i = \theta_1 + \theta_2 x_i + \epsilon_i \qquad i = 1, \ldots, n \\ &\epsilon_i \sim N(0,1) \\ &-1 \le x_i \le +1 \quad \forall i \end{aligned} \tag{14.3}$$

Assume that the investigator has control over the values taken by the explanatory variable $x_i$, subject only to a lower and an upper bound, which we assume without loss of generality to be $-1$ and 1. Note that the error term is assumed to be normally distributed, and, for the sake of further simplicity, to have unit variance. Given these assumptions concerning the error term, we may construct the log-likelihood function for this model as:

$$\text{LogL} = \sum_{i=1}^{n} \left[k - (y_i - \theta_1 - \theta_2 x_i)^2\right] \tag{14.4}$$

where $k$ is a constant. It is easily verified that in this model the **MLE**s of the two parameters are the same as the estimates from a least squares regression of $y$ on $x$. Differentiating (14.4) twice with respect to the two parameters $\theta_1$ and $\theta_2$ we find the information matrix to be:

$$I = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \tag{14.5}$$

So the variance of the **MLE** vector is:

$$V\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = I^{-1} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \tag{14.6}$$

Standard errors are obtained as the square roots of diagonal elements of $V$.

To obtain the D-Optimal design, we need to choose values of $x_i$ that maximise the determinant of $I$. It may be verified that this determinant may be written as:

$$\det(I) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (x_i - x_j)^2 \tag{14.7}$$

From (14.7) it is clear that the differences between different $x$-values must be as great as possible. For this, half of the values must be set to the maximum allowed, and the other half to the minimum. In experimental design jargon, we are choosing all design points to be at the "corners of the design space".



An important feature of (14.7) is that it does not involve the parameters of the model, $\theta_1$ and $\theta_2$, with the implication that, in the context of the linear regression model, the parameter values do not need to be known in order for the D-optimal design to be found. As we shall discover in the next sub-sections, this result does not carry over to non-linear models such as probit and logit.

### 14.2.3 Simple probit and simple logit

Now consider a binary data setting, in which an underlying continuous variable $y^*$ depends on $x$ according to:

$$\begin{aligned} &y_i^* = \theta_1 + \theta_2 x_i + \epsilon_i \qquad i = 1, \ldots, n \\ &\epsilon_i \sim N(0,1) \end{aligned} \tag{14.8}$$

but all that is observed is whether $y^*$ is positive or negative. That is, we observe $y$ where:

$$y_i = \begin{cases} 1 & \text{if} \quad y_i^* > 0 \\ -1 & \text{if} \quad y_i^* \le 0 \end{cases} \tag{14.9}$$

Equation (14.9) is of course the well known probit model, which was introduced in Chapter 6.[2] The log-likelihood function for the probit model is:

$$\text{LogL} = \sum_{i=1}^{n} \ln \Phi\left[y_i \times (\theta_1 + \theta_2 x_i)\right] \tag{14.10}$$

The information matrix for the model (14.10) may be derived as:

$$I = \begin{pmatrix} \sum w_i & \sum w_i x_i \\ \sum w_i x_i & \sum w_i x_i^2 \end{pmatrix} \tag{14.11}$$

where:

$$w_i = \frac{\left[\phi(\theta_i + \theta_2 x_i)\right]^2}{\Phi(\theta_1 + \theta_2 x_i)\left[1 - \Phi(\theta_1 + \theta_2 x_i)\right]} \tag{14.12}$$

The determinant of the information matrix (14.11) may be written as:

$$\det(I) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_i w_j (x_i - x_j)^2 \tag{14.13}$$

Again, $\det(I)$ is maximised with just two design points. However, $\det(I)$ is weighted by the $w_i$s. These weights are maximised when $\Phi(\theta_1 + \theta_2 x_i) = 0.5$, that is, when the

[2] There is a slight difference from the treatment in Chapter 6. There we denoted as $y$ the binary variable taking values 1 or 0, and as $yy$ the recoded variable taking values 1 or $-1$. Here, for the sake of expositional simplicity we are defining $y$ to be the variable with values 1 or $-1$.

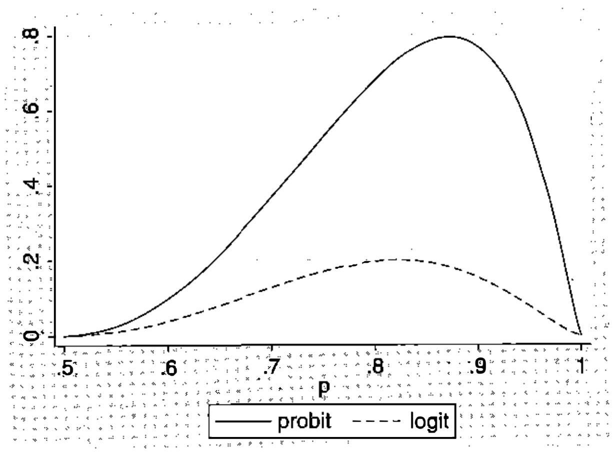



Figure 14.1: Determinant of information matrix against percentile of larger design point; probit and logit

probabilities of the two outcomes are equalised. So, the desire to have design points as far from each other as possible, occupying the "corners of the design space", is countered by the desire to have design points giving rise to perfect indifference the requirement of "utility balance" (Huber & Zwerina, 1996).

In Figure 14.1, the solid curve shows det($I$) against the percentile of the upper design point. The design is symmetric, so the lower design point is an equal distance from the centre. We see, as expected, that when both design points are in the centre (percentile=0.50) the information is zero. The intuition here is that if all design points are in the centre of the distribution, all that can be observed is which side of the centre each observation lies, and the spread of the distribution cannot be identified. We also see that when both design points are the maximum distance from the centre (percentile=1.0), the information is zero again. Again this is intuitive: if all individuals are given such extreme problems that their choice can be predicted with certainty, the choice data will be of no value. The most important feature of the solid line in Figure 14.1 is the maximum at 0.87. This implies that the design points that maximise det($I$) are the 13th and 87th percentiles of the underlying response function. This is the D-optimal design for the probit model.

If the required number of design points is odd, the optimal design is to place one design point exactly in the centre, and to divide the remaining points equally between the 13th and 87th percentiles.

A well known alternative to probit for the modelling of binary data is the logit model, also introduced in Chapter 6, defined by:

$$P(y_i = 1) = \frac{\exp(\theta_1 + \theta_2 x_i)}{1 + \exp(\theta_1 + \theta_2 x_i)} \equiv P_i \tag{14.14}$$



The information matrix for (14.14) has the same form as (14.11) above, but with weights given by:

$$w_i = P_i(1 - P_i) \tag{14.15}$$

With (14.15) in (14.13), it is found, again numerically, that the design points that maximize det($I$) are the 18th and 82nd percentiles of the underlying response function.[3] The broken line in Figure 14.1 shows det($I$) against the percentile of the upper design point for the logit model.

Note that in order to find these optimal design points, unlike in the setting of the linear regression model, the parameters of the underlying distribution (i.e. $\theta_1$ and $\theta_2$) must be known in advance, since obviously these are needed in order to recover a point on the distribution from knowledge of its percentile. As mentioned in Section 14.1, this is sometimes referred to as the "chicken and egg" problem.

## 14.3 The Random Preference (RP) Model Revisited

We will assume the random preference (RP) model which was first introduced in Section 12.6 and then considered again in more detail in Section 13.2.4.

We are concerned with the optimal design of an experiment in which a typical task requires a choice between two lotteries. We will assume the same setting as used in Chapters 12 and 13. We assume that each choice problem involves three outcomes, \$0, \$10 and \$20, which we normalise as 0, 1, and 2 respectively. The two lotteries can then be defined in terms of vectors of probabilities:

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \qquad \mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$$

where the three probabilities in a vector are the probabilities of receiving the three outcomes 0, 1 and 2 respectively. It is usually the case that one of the two lotteries is riskier, and the other is safer, and when this is not the case, one of the lotteries stochastically dominates the other. $\mathbf{p}$ will always be the riskier, or dominating lottery. $\mathbf{q}$ will always be the safer, or dominated lottery.

Again we assume the *power utility function*:

$$U(x) = x^r \qquad r > 0 \tag{14.16}$$

With this parameterisation, the Coefficient of Relative Risk Aversion is given by $1 - r$. That is, the higher is the power parameter, $r$, the less risk averse is the individual.

[3] We are often led to believe there are no major differences between probit and logit. For example, according to Greene (2008, p.774), "in most applications, the choice between these two seems not to make much difference". Against this background, it perhaps seems surprising that the optimal design points under probit are five percentiles further into the tails than under logit.

We will assume Expected Utility maximisation (EU). As in Chapter 13, we define, for a given choice problem, a "threshold" attitude to risk value, $r^*$, such that (given EU):

$$r < r^* \implies \mathbf{q} \text{ is chosen} \tag{14.17}$$

One important reason why we chose the power utility function (14.16) is so that $r^*$ can be expressed in closed form, as follows:[4]

$$r^* \equiv \frac{\ln\left(-\frac{q_2-p_2}{q_3-p_3}\right)}{\ln(2)} \tag{14.18}$$

Consider a particular subject. The central assumption of the Random Preference (RP) model is that the subject's risk-attitude varies between tasks according to the following distribution:

$$\ln(r) \sim N(m, \sigma^2) \tag{14.19}$$

The lognormal distribution is being assumed for $r$ in order to ensure that $r$ is always positive, as is required in the definition of power utility.

Bringing together (14.17)–(14.19), the probability of the subject choosing the safer lottery is:

$$P(S) = P(r < r^*) = \Phi\left(\frac{\ln r^* - m}{\sigma}\right) = \Phi\left[-\frac{m}{\sigma} + \left(\frac{1}{\sigma}\right)\ln r^*\right] \tag{14.20}$$

Equation (14.20) is a simple probit model with explanatory variable $\ln(r^*)$, where $r^*$ is the threshold risk attitude (under EU) for the choice problem, computed directly using (14.18).

Equation (14.20) is what we might describe as a "structural form" probit model, since it is written in terms of the structural parameters, $m$ and $\sigma$. The corresponding reduced form model would be:

$$P(S) = \Phi[\theta_1 + \theta_2 \ln r^*] \tag{14.21}$$

The experimental design theory outlined in Section 14.2 above prescribes a rule for choosing values of $r^*$ which will al ow the reduced form parameters, $\theta_1$ and $\theta_2$ in (14.21), to be estimated with maximal precision. However, we are interested not in $\theta_1$ and $\theta_2$, but in the structural form parameters, $m$ and $\sigma$, appearing in (14.20). A technica question arising at this point is therefore whether the values of $r^*$ which are optimal for the estimation of $\theta_1$ and $\theta_2$ are also optimal for the estimation of $m$ and $\sigma$. The answer is yes. To demonstrate this, we define vectors of reduced form and structural form parameters as follows:

$$\theta = \begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} \qquad \beta = \begin{pmatrix}m\\ \sigma\end{pmatrix} \tag{14.22}$$

[4] If the other popular version of CRRA, $U(x) = \frac{x^{1-r}}{1-r}$, is assumed, threshold values of $r$ cannot be found in closed form.



and we note that the relationship between $\theta$ and $\beta$ may be represented in the following two ways:

$$\theta = \begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \begin{pmatrix}-\frac{m}{\sigma}\\ \frac{1}{\sigma}\end{pmatrix} \qquad \beta = \begin{pmatrix}m\\ \sigma\end{pmatrix} = \begin{pmatrix}-\frac{\theta_1}{\theta_2}\\ \frac{1}{\theta_2}\end{pmatrix} \tag{14.23}$$

Differentiation of the second of these relationships yields:

$$\mathbf{D} = \frac{\partial \beta}{\partial \theta'} = \begin{pmatrix}\frac{\partial m}{\partial \theta_1} & \frac{\partial m}{\partial \theta_2}\\ \frac{\partial \sigma}{\partial \theta_1} & \frac{\partial \sigma}{\partial \theta_2}\end{pmatrix} = \begin{pmatrix}-\frac{1}{\theta_2} & \frac{\theta_2}{\theta_2^2}\\ 0 & -\frac{1}{\theta_2^2}\end{pmatrix} \tag{14.24}$$

The information matrix with respect to $\beta$ may be written, to close approximation, in terms of that with respect to $\theta$ as follows:

$$\mathbf{I}(\beta) \approx \mathbf{D}\mathbf{I}(\theta)\mathbf{D}' \tag{14.25}$$

Note that the formula used here is closely related to that underlying the "delta method", explained in Section 6.5.2.

Since $\mathbf{D}$ in (14.24) is a square non-singular matrix, we can find the determinant of (14.25) to be:

$$\begin{aligned}\det\left(\mathbf{I}(\beta)\right) &\approx \det(\mathbf{D}) \times \det\left(\mathbf{I}(\theta)\right) \times \det(\mathbf{D}')\\ &= \det(\mathbf{D}\mathbf{D}') \times \det\left(\mathbf{I}(\theta)\right) = \delta \det\left(\mathbf{I}(\theta)\right) \quad \text{where} \quad \delta > 0\end{aligned} \tag{14.26}$$

It is known that $\delta > 0$ because it is the determinant of the positive definite matrix $\mathbf{D}\mathbf{D}'$. Hence, the values of $r^*$ which maximise $\det\left(\mathbf{I}(\beta)\right)$ are the same as those which maximise $\mathbf{I}(\theta)$.

We have been considering the probit model, since we commenced with the assumption of log-normality in (14.19). We could instead have started from the assumption of a log-logistic distribution. If we define:

$$U = \frac{\ln(r) - m}{\sigma} \tag{14.27}$$

and assume that $U$ follows a logistic distribution, with distribution function:

$$P(U < u) = \frac{\exp(u)}{1 + \exp(u)} \tag{14.28}$$

then the probability of the safe lottery being chosen is:

$$P(S) = P(r < r^*) = P\left[U < \frac{\ln(r) - m}{\sigma}\right] = \frac{\exp\left[-\frac{m}{\sigma} + \left(\frac{1}{\sigma}\right)\ln(r^*)\right]}{1 + \exp\left[-\frac{m}{\sigma} + \left(\frac{1}{\sigma}\right)\ln(r^*)\right]} \tag{14.29}$$

Equation (14.29) is the binary logit model, with explanatory variable $\ln(r^*)$.

## 14.4 Application of D-optimal Design Theory to a Risky Choice Experiment

We next consider how we might apply the optimal design theory outlined in Section 14.2 in the estimation of the parameters of the theoretical models outlined in Section 14.3.

A point that has already been made is that, in order to apply these optimal design rules, the values of the parameters must be known. We will therefore assume that the model has been estimated on a previous data set and estimates of the parameters are available. An obvious data set to use for this purpose is the "Random Preference" data analysed in Chapter 13. When the RP model (i.e. the "correct" model) was estimated on this data set, we obtained the results presented in the final column of Table 13.1. The estimates that are required here are: $\hat{\mu} = -0.89$; $\hat{\eta} = 0.22$; $\hat{\sigma} = 0.16$. On the basis of these estimates, we are approaching the problem of optimal design on the assumption that individuals' "mean log risk attitude" $m$ varies over the population according to:

$$m \sim N(-0.89, 0.22^2) \tag{14.30}$$

and, an individual with "mean log risk attitude" $m$, has risk attidude varying between tasks according to:

$$\ln(r) \sim N(m, 0.16^2) \tag{14.31}$$

We start by considering what the optimal set of choice problems is for a particular subject whose mean risk attitude is "typical", having $m = -0.89$, so that their median power parameter is $e^{-0.89} = 0.41$. This subject's risk attitude parameter varies between tasks according to:

$$\ln(r) \sim N(-0.89, 0.16^2) \tag{14.32}$$

In Section 14.2, we obtained the result that, for the probit model, the D-optimal design consists of two design points: the 13th and 87th percentiles of the underlying response function. Moreover, it was noted that, if an odd number of design points is required, one should be placed at the 50th percentile. Hence, if a third design point is required, we choose the median of the distribution.

The 13th, 50th and 87th percentiles of $r$, whose distribution is defined in (14.32) are, in fact:

$$0.343, 0.411, 0.492$$

These are the three required values of $r^*$ for the subject under analysis. To verify that these are the required values, we can compute the probability of the subject choosing the safe alternative at each of these values of $r^*$, using (14.20) and (14.32):

$$P(S|r^* = 0.343) = \Phi\left[\frac{\ln(0.343) - (-0.89)}{0.16}\right] = \Phi(-1.13) = 0.13$$

$$P(S|r^* = 0.411) = \Phi\left[\frac{\ln(0.411) - (-0.89)}{0.16}\right] = \Phi(0) = 0.50$$



$$P(S|r^* = 0.492) = \Phi\left[\frac{\ln(0.492) - (-0.89)}{0.16}\right] = \Phi(1.13) = 0.87 \tag{14.33}$$

As required the three probabilities computed in (14.33) are in agreement with the three required percentage points.

All that remains is to reverse-engineer choice problems with these values of $r^*$. Of course, for any given value of $r^*$, there is an infinity of possible choice problems. Let us (for the moment) restrict these possibilities by requiring that the safer lottery is a certainty of the middle outcome ($10), and the riskier lottery involves only the lowest ($0) and highest ($20) outcomes. When the choice problem is of this form, (14.18) becomes:

$$r^* = -\frac{\ln(p_3)}{\ln(2)} \tag{14.34}$$

where $p_3$ is the probability of the highest outcome under the risky lottery. Inverting (14.34), we obtain:

$$p_3 = 2^{-r^*} \tag{14.35}$$

Equation (14.35) gives the value of $p_3$ that is required in order for the choice problem to have threshold risk attitude $r^*$, assuming EU. Applying (14.35) to our three values of $r^*$ gives rise to the three choice problems presented in Figure 14.2.

## 14.5 Optimal Design in the Presence of a Tremble Parameter

For reasons explained in Chapter 13, it is often useful to assume that there is a probability $\omega$ that the subject will lose concentration, and choose randomly

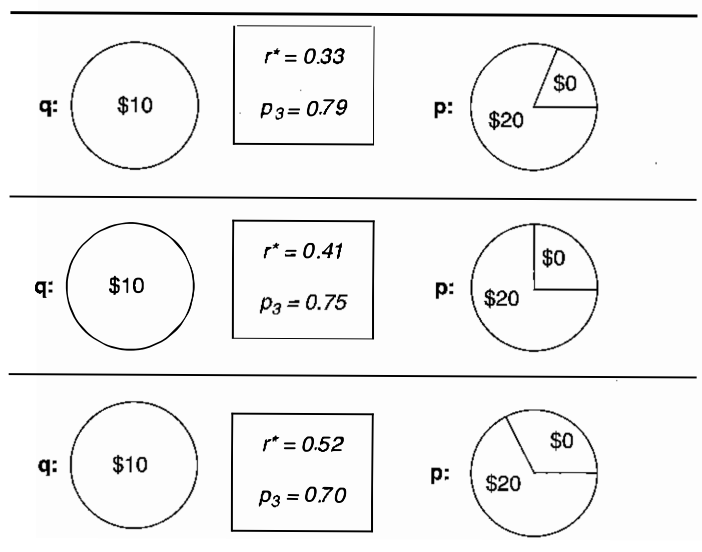



Figure 14.2: Optimal design for subject with $\ln(r) \sim N(-0.89, 0.16^2)$

between the two alternatives. $\omega$ is known as the tremble probability, or tremble parameter.

Introducing a tremble parameter to the model of Section 14.3, the probability of choosing the safe alternative (for a particular individual with mean log risk attitude $m$) becomes:

$$P(S) = (1-\omega)\Phi\left[-\frac{m}{\sigma} + \left(\frac{1}{\sigma}\right)\ln r^*\right] + \frac{\omega}{2} \tag{14.36}$$

and the likelihood function is constructed accordingly. If we were to find the $3 \times 3$ information matrix for $(m, \sigma, \omega)$, we would find that its determinant increases steeply when we add choice problems with *extreme* design points, for example those shown in Figure 14.3.

The first choice problem shown in Figure 14.3 is an example of a dominance problem, because the **p**-lottery stochastically dominates the **q**-lottery. The threshold risk attitude for this problem is 0.00, implying that *any* permissible risk attitude (i.e. any $r > 0$) will result in a choice of the **p**-lottery over the **q**-lottery. If a subject indicates a preference for **q**, we can be sure that they are "trembling"; hence the usefulness of this sort of problem in the estimation of the tremble parameter. The second choice problem shown in Figure 14.3 is located at the other extreme. Here the threshold risk attitude is 3.00, indicating that a subject would need to be extremely risk-loving in order to have a preference for the **p**-lottery. If a subject indicates a preference for **p** in this case, we can be almost certain that they are trembling, especially if their other decisions indicate that their risk attitude is somewhere in the normal range.

Even though it is very unlikely that the tremble parameter is the focus of any experimental analysis, it is important that it is estimated with satisfactory precision if it is present in a model, since imprecision in the estimation of one parameter has a detrimental effect on the precision with which other parameters are estimated. To ensure this, it is important to include in the design at least some problems of the type shown in Figure 14.3.

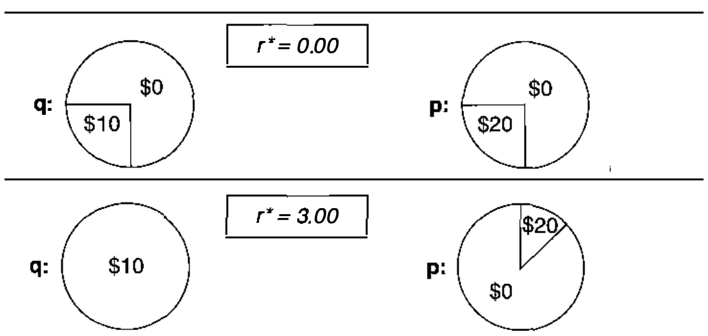



Figure 14.3: Examples of choice problems with extreme design points



## 14.6 Optimal Design for a Sample of Subjects

In Section 14.4, we constructed the optimal design for a particular subject with a particular known value of $m$ (i.e. their "personal average" risk attitude). In practice, all subjects have different values of $m$, and the investigator does not know the values of $m$ for each individual. Given this, how should a design be constructed so that it is optimal for a sample of subjects?

One possible approach, implemented by Chaudhuri & Mykland (1993, 1995), is to perform an "interactive experiment". In such an experiment, subjects' choices are continually monitored, and all choices made up to a particular stage in the experiment are used in constructing a design which is locally optimal (for the individual subject) for the next stage. Such an approach clearly has the potential advantage of maximal efficiency. However, we also see a problem with this approach: if subjects realise that their responses have an influence on the parameters of subsequent choice problems, they might deliberately choose less preferred alternatives in an attempt to "steer" the experiment towards more desirable choice problems later on. In short, there is a violation of incentive compatibility, because there are reasons to believe that true preferences are not being revealed.[5] For this reason, we restrict attention to the problem of choosing a design that is the same for all subjects, and is constructed before the start of the experiment.

In order to construct such a design, the investigator must have some idea of the form and the extent of between-subject heterogeneity. A natural assumption for heterogeneity is:

$$m \sim N(\mu, \eta^2) . \tag{14.37}$$

Assumption (14.37) leads to the random effects probit model:

$$\begin{aligned} \ln(r_{it}) &= m_i + \epsilon_{it} \\ \epsilon_{it} &\sim N(0, \sigma^2) \\ m_{it} &\sim N(\mu, \eta^2) . \end{aligned} \tag{14.38}$$

There are three parameters to be estimated in (14.38): the between mean ($\mu$), the between standard deviation ($\eta$), and the within standard deviation ($\sigma$). The likelihood function for this model is derived in Chapter 9.

Rather than attempt to find the D-optimal design for this model, we will describe a simpler approach here. We once again use the previously obtained estimates of the between parameters, $\mu$ and $\eta$, as well as the within parameter, $\sigma$. We will select quantiles of the between-subject distribution, and construct the optimal design for a subject at each quantile. For example, if we take the deciles of $N(\mu, \eta^2)$, which is $N(-0.89, 0.22^2)$, this gives nine values for $m$. For each of these values of $m$,

[5] The "Prior Incentive System (Prince)" of Johnson et al. (2014) provides an ingenious means of designing an interactive experiment in such a way as to remove any incentive for such strategic choices.

we find the three optimal design points (lower, middle, upper), using the procedure outlined in Section 14.4. For each design point, we would find the corresponding choice problem using the procedure demonstrated at the end of Section 14.4. This will result in 27 different choice problems.

For the purpose of identifying the tremble parameter, as explained in Section 14.5, we would add two "extreme" problems to the design, one from either extreme. The two problems presented in Figure 14.3 would be suitable for this purpose. In total, we would therefore have 29 choice problems, to be presented to *all* subjects in the sample.

If a larger number of tasks is desired in the experiment, a "foldover" may be implemented, as recommended by Louviere et al. (2000). In the present context, a foldover simply means the same 29 problems presented with the left and right lotteries interchanged. This would give 58 choice problems, perhaps a reasonable number of tasks for a typical experiment.

### *14.6.1 Choice of design used in Chapter 13*

In Chapter 13, a particular experimental design was assumed, in which there were 50 distinct risky choice problems. This set of choice problems is presented in Appendix C. Here, we shall briefly outline the design used to generate these 50 choice problems.

Recall that each choice problem is assigned a measure of complexity, defined simply as the number of outcomes appearing in the simpler of the two lotteries. This variable was found to be a key determinant of decision time in the analysis of Chapter 5; decision time was found to be an increasing, concave, function of complexity. In order to invesigate the impact of complexity as we did in Chapter 5, it is necessary for complexity to vary across choice problems. Furthermore, since we held a prior belief that decision time is a non-linear function of complexity, in order to capture this non-linearity in estimation, two levels of complexity would not be sufficient. For this reason, we require the design to include problems of all three complexity levels, 1, 2 and 3.

Referring again to the table in Appendix C, the first 17 problems are all complexity level 1. For each of these, the safer lottery is a certain outcome of \$10. Hence these 17 problems are chosen using a procedure similar to that described in Section 14.6. The 17 choice problems all have different values of $r^*$ ranging from 0.074 to 3.32. Think of these as quantiles of an assumed prior distribution of $m$ (mean log risk attitude) across the population. For each of these selected values of $r^*$, formula (14.35) is used to obtain the the required choice problem.

Problems 18–34 are of complexity level 2, and problems 35–46 are of complexity level 3. For these two groups, once again a range of values for $r*$ is chosen, and a problem is "reverse-engineered" for each of the chosen values of $r*$.

The last four choice problems in the design, 47–50, are dominance problems. Hence their $r*$ value is zero. Three of these are of complexity level 2; the fourth has complexity level 3.



## 14.7 Summary and Further Reading

The main objectives of the chapter have been to set out the principles of optimal design theory, and then to demonstrate how these principles can be applied to a particular problem in experimental economics.

While a vast literature exists on the problem of optimal experimental design, most of that work has been applied to linear models (Silvey, 1980; Fedorov, 1972) and most of the seminal papers were highly theoretical. Optimal designs in non-linear models have been covered by Ford et al. (1992) and Atkinson (1996). A key issue is the "chicken-and-egg" problem which means that it is important to have a reasonably good idea of the parameters of interest in order to construct the optimal design.

While it is important to bear in mind the objectives of the study in the process of finding the optimal design, it can be a mistake to focus exclusively on a particular objective. In the context of a risky choice experiment, one basic requirement of a design is that it gives a healthy mix of safe and risky choices. In order to ensure such a mix, consideration must be given to the distribution of risk attitude over the population, and to the manner in which risk attitude varies for a given individual. The broad features of the design must be decided with these factors in mind, in accordance with the rules derived in Section 14.4, even if the objective of the experiment is something other than a study of the distribution of risk attitudes. If, for example, the objective of the study is to test departures from EU, a design with the required broad features (i.e. giving a healthy mix of responses) would be adjusted in the way required to maximise precision with respect to the additional (e.g. probability weighting) parameters. Steps in this direction have been taken by Moffatt (2007). See also Müller & de Leon (1996), who considered the problem of finding the optimal design when the objective is discriminating between two non-EU models: subjective EUT and regret theory.

While the "healthy mix" of responses is a necessary condition for a good design, it is not a sufficient condition. Consider, for example, a situation in which all of the choice problems are of the type presented in Figure 14.3. These are labelled "extreme" problems in that there is a very high chance of a particular alternative being chosen. If all problems are of this type, it would be easy to ensure a mix of responses, simply by dividing the problems equally between the two extremes. However, the resulting choice data would be of little use in estimating the parameters (with the possible exception of the tremble). For a different example, consider a design in which all problems are such that subjects are very close to indifference between the alternatives. This design would lead to an approximately even mix of safe and risky choices, but, again, the data would not be useful for estimation: the basic problem here is that while the location parameter of risk attitude could be estimated by such choices close to indifference, the spread parameter would not be identified. For identification of the latter, some problems located a modest distance away from indifference are required. Readers interested in these issues are referred to Kanninen (1993) and Huber & Zwerina (1996).

## Exercises

In Section 6.2.5, we reported on the estimation of a probit model of the responder's decision in an ultimatum game. The results were (d=1 means "accept"):

$$\hat{P}(d = 1) = \Phi(-3.855 + 0.144y) \tag{14.39}$$

where $y$ is the amount of the offer and $d = 1$ indicates acceptance of the offer.

A version of the Strategy Method which is sometimes used is the *closed-ended referendum*, in which a particular offer is suggested to each responder, and they indicate whether or not they would accept the suggested offer.

1. If you were to ask half of all respondents whether they would accept one (low) offer, and the other half of respondents whether they would accept a (high) offer, what would the low and the high offers be, if you wanted a D-optimal design? Use the probit results as the prior.
2. Results from logit estimation on the same data are:

$$P(d = 1) = \frac{\exp(-6.623 + 0.247y)}{1 + \exp(-0.623 + 0.247y)}. \tag{14.40}$$

Compute the two optimal offers using the logit estimates as the prior.

# Chapter 15
# Social Preference Models

## 15.1 Introduction

At various stages of the book, we have considered the determinants of dictator game giving, including in Chapter 11 where we used the panel hurdle model as a means of identifying such determinants. In this chapter, we go a step further and estimate the parameters of a utility function underlying dictator game giving. That is, we focus on the *structure* of the preferences underlying behaviour. In this context, it has become conventional to assume that utility is an increasing function of two arguments: own payoff ($x_1$) and other's payoff ($x_2$).

In order to estimate the parameters of such a utility function, it is obviously essential for the experimental design to include variation in the endowment. However, it is also desirable for the prices of the two "goods" to vary. That is, the price of "giving" and the price of "keeping" should be varied in order to allow estimation of parameters representing, for example, the degree of substitutability or complementarity between these two "goods".

Behaviour in dictator games becomes particularly interesting when endowments are earned. Here, we find a fairly clear division of the population between types who believe that the final allocation should be closely related to the amounts earned, and types who strive for equality regardless of amounts earned. Hence, in this situation in which endowments are earned, the finite mixture modelling framework is called for in order to separate out these different types.

In Section 15.2, we will first use an existing data set (Andreoni & Miller, 2002) to estimate the parameters of a constant elasticity of substitution (CES) utility function. In Section 15.3 we consider the problem of estimating the parameters of a utility function in the context of a model that treats zero observations as binding non-negativity constraints. In Section 15.4 we consider finite mixture models similar to those of Cappelen et al. (2007), which are again also based on the assumption that agents maximise a utility function, but in which agents are assumed to differ in the type of "fairness ideal" that governs their behaviour. Finally, in Section 15.5 we consider methods for estimating the parameters of the utility function when the data that is available is on subjects' *choices* between hypothetical allocations. Here, the appropriate framework for estimation is discrete choice modelling, as used by Engelmann & Strobel (2004).

## 15.2 Estimation of Preference Parameters from Dictator Game Data

### 15.2.1 The framework

The experimental setting considered here and in Section 15.3 is that of Andreoni & Miller (2002). In this setting, each individual is given an endowment ($m$) which they are required to allocate between "self" and "other", with both of these "goods" having a "price". For example, if "giving to self" has price $\frac{1}{2}$, the amount actually received by "self" will be twice the amount allocated; if "giving to other" has price $\frac{1}{3}$, the amount actually received by "other" will be three times the amount allocated.

We define the following variables:

$x_1$ = amount *received* by self
$x_2$ = amount *received* by other
$m$ = endowment
$p_1$ = price of $x_1$ (i.e. for each unit of the endowment that you direct to yourself, you receive $1/p_1$ units).
$p_2$ = price of $x_2$ (i.e. for each unit of the endowment that you direct to the other player, they receive $1/p_2$ units).

An important point to stress at this stage is that, although $x_1$ and $x_2$ will be the two arguments of the dictator's utility function, they are *not* decision variables. The decision variables are, in fact:

$p_1x_1$ = amount *directed* to self
$p_2x_2$ = amount *directed* to other

Of course, these two decision variables are not both free variables. They are constrained by the budget constraint:

$$p_1x_1 + p_2x_2 \leq m \tag{15.1}$$

We will normally refer to the amount directed to other, $p_2x_2$ as the single decision variable, and recognise that, because the budget constraint will always be binding, the other decision variable is determined as $p_1x_1 = m - p_2x_2$.

It is also useful to define "budget shares": $w_1 = \frac{p_1x_1}{m}$; $w_2 = \frac{p_2x_2}{m}$.

### 15.2.2 The Andreoni-Miller data

Andreoni & Miller's (2002) data is contained in the file **garp**. The variables are as defined in Section 15.2.1. Here we shall provide a description of the data, and report on some exploratory analysis thereof.



There were 176 subjects in the experiment. Each subject was faced with a sequence of decision problems in the form of budgets. Each budget had a different combination of endowment ($m$), price of keeping ($p_1$), and price of giving ($p_2$). These combinations are shown in Table 15.1. The task subjects are required to perform for each budget is to decide how much of the endowment ($m$) to keep for themselves ($p_1x_1$), and how much to send to the other player ($p_2x_2$). The decision problems were presented in a random order to each subject. Subjects were told that, when all decisions had been made, one of the decision problems would be chosen at random and carried out with another randomly chosen subject as the recipient.

Budgets 1–8 were faced by all 176 subjects. Budgets 9–11 were only faced by 34 of the subjects. The final column of Table 15.1 shows the average amount sent to the other player for each budget. Clearly, there is a good amount of variation in this outcome.

Note from Table 15.1 that for three of the problems, 7, 8, and 9, the two prices are both one, implying that these tasks correspond to the standard dictator game. From the second and sixth columns, we see that in these three tasks, average giving is between 17 and 24% of the endowment, in close agreement with previous dictator game experiments (Camerer, 2003).

Figure 15.1 shows a (jittered) scatter of amount received by other against amount received by self. This further highlights the wide variation in the amount of giving. This is partly a result of the richness of the design (with 11 different budget constraints), and also a high apparent variation in preference for giving. As expected, there is a higher concentration of points in the lower-right region of the plot, reflecting an overall bias towards giving-to-self (with 42% of the observations on giving-to-other being zero). The "jitter" option has been used for the scatter so that clusters of observations at particular points (e.g. on the horizontal axis) are easily discernible.

| Budget | $m$ | $p_1$ | $p_2$ | Observations | Mean amount sent to other |
|---|---|---|---|---|---|
| 1 | 40 | 0.33 | 1 | 176 | 8.02 |
| 2 | 40 | 1 | 0.33 | 176 | 12.81 |
| 3 | 60 | 0.5 | 1 | 176 | 12.67 |
| 4 | 60 | 1 | 0.5 | 176 | 19.40 |
| 5 | 75 | 0.5 | 1 | 176 | 15.51 |
| 6 | 75 | 1 | 0.5 | 176 | 22.68 |
| 7 | 60 | 1 | 1 | 176 | 14.55 |
| 8 | 100 | 1 | 1 | 176 | 23.03 |
| 9 | 80 | 1 | 1 | 34 | 13.5 |
| 10 | 40 | 0.25 | 1 | 34 | 3.41 |
| 11 | 40 | 1 | 0.25 | 34 | 14.76 |

Table 15.1: Andreoni & Miller's (2002) design
Notes: There are 11 different budgets, each with a different combination of endowment ($m$), price of keeping ($p_1$), and price of giving ($p_2$). In budgets 1–8, all 176 subjects participated; in budgets 9–11, only 34 participated. Final column shows mean amount sent to other.

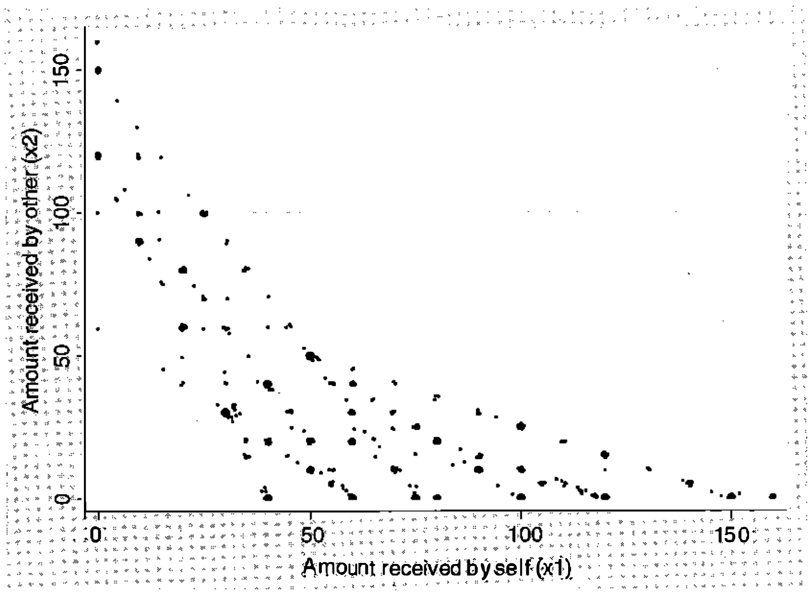



Figure 15.1: Jittered scatter of data in $(x_1, x_2)$ space

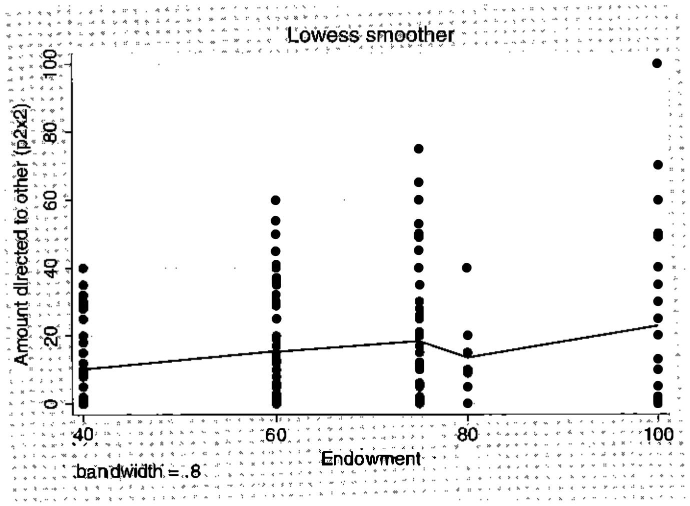



Figure 15.2: Amount directed to other against endowment

Figure 15.2 shows a scatter of amount directed to other against endowment, with smoother. The positive relationship evident here simply indicates that giving is a "normal good".

Figure 15.3 shows smoothers of amount *received* by other against price of giving to other (left pane) and against price of giving to self (right pane). The



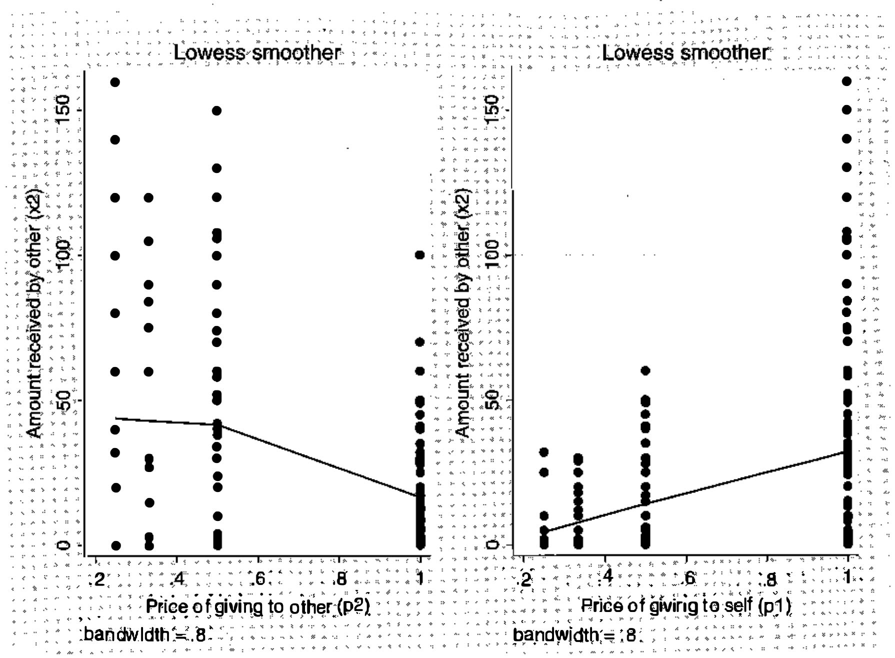



Figure 15.3: Amount received by other against price of giving to other (left pane), and against price of giving to self (right pane), with smoothers

downward-sloping curve seen in the left-hand graph is consistent with the "law of demand"; the upward-sloping curve seen in the right-hand graph is consistent with the two "goods" (amount received by other and amount received by self) being substitutes.

These results may be confirmed using a linear regression. Regression of amount received by other on the two prices, with clustering at the subject level, gives the following results. The effects of the two prices are very strong, and with signs as expected on the basis of the two graphs in Figure 15.3.

```
. regress x2 p2 p1, vce(cluster i)

Linear regression                                      Number of obs =    1510
                                                       F( 2,    175) =   61.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1847
                                                       Root MSE      =  28.661

                                (Std. Err. adjusted for 176 clusters in i)
------------------------------------------------------------------------------
             |               Robust
          x2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          p2 |  -39.00726   4.934956    -7.90   0.000    -48.74695   -29.26757
          p1 |   14.47704   1.664276     8.7[illegible]   0.000      11.1924    17.76167
       _cons |   43.95138   4.663821     9.42   0.000     34.74681    53.15596
------------------------------------------------------------------------------
```

Next we add income (i.e. the endowment) into the regression. Income is seen to have a strongly positive effect, confirming that giving to other is a "normal good". The interpretation of its coefficient (0.265) is that when the dictator's endowment rises by one unit, *ceteris paribus*, amount received by other will rise by around one quarter of one unit. However, a consequence of adding income to the regression is that the effect of the price of giving to "self" is no longer significant. This is partly a result of the strong positive correlation between $m$ and $p_1$ essentially causing $p_1$ to take the role of "proxy" for $m$ in the model in which the latter is excluded.

```
. regress x2 p2 p1 m, vce(cluster i)

Linear regression                                      Number of obs =     1510
                                                       F(  3,   175) =    61.25
                                                       Prob > F      =   0.0000
                                                       R-squared     =   0.1976
                                                       Root MSE      =   28.441

                                  (Std. Err. adjusted for 176 clusters in i)
------------------------------------------------------------------------------
             |               Robust
          x2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          p2 |  -52.12677   5.063235   -10.30   0.000    -62.11964   -42.13391
          p1 |   1.357528   1.783083     0.76   0.447    -2.161587    4.876643
           m |    .265248   .0277023     9.57   0.000     .2105744    .3199216
       _cons |   47.92717   4.707122    10.18   0.000     38.63713     57.2172
------------------------------------------------------------------------------
```

It is obvious from Figure 15.1, and also from all previous analyses of dictator game data within this book, that there is an accumulation of zero observations in giving to other. Around 42% of this sample consist of observations with giving equal to zero. The linear regressions just performed do not take account of this accumulation of zero observations. A model which does take account of this feature of the data is the Tobit model, explained in detail in Section 6.6.3. We next estimate a Tobit model of giving on the two prices and income, again with cluster-robust standard errors. The results are as follows.

```
  tobit x2 p2 p1 m, vce(cluster i) ll(0)

Tobit regression                                  Number of obs   =       1510
                                                  F(   3,   1507) =      54.33
                                                  Prob > F        =     0.0000
Log pseudolikelihood =  -5027.146                 Pseudo R2       =     0.0256

                                  (Std. Err. adjusted for 176 clusters in i)
------------------------------------------------------------------------------
             |               Robust
          x2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          p2 |   -67.1347   7.049639    -9.52   0.000    -80.96285   -53.30656
          p1 |    10.8052   3.910197     2.76   0.006     3.135191     18.4752
           m |   .3322818   .0380964     8.72   0.000     .2575541    .4070095
       _cons |   34.41715   6.122105     5.62   0.000      22.4084     46.4259
-------------+----------------------------------------------------------------
      /sigma |   42.59774    2.46888                      37.75494    47.44055
------------------------------------------------------------------------------
  Obs. summary:        628  left-censored observations at x2<=0
                       882     uncensored observations
                         0 right-censored observations
```



We see that the coefficient estimates from Tobit are considerably larger in magnitude than the corresponding OLS estimates. Most strikingly, the Tobit coefficient of price of amount received by self ($p_1$) is 10.81, eight times larger than the OLS estimate of the same parameter, which is 1.36. Furthermore, the Tobit coefficient is strongly significant ($p = 0.006$) compared to a complete lack of significance under OLS ($p = 0.447$). This emphatically confirms the importance of dealing with zero censoring when analysing data sets of this type. This will be the focus in Section 15.3.

Of course, we can go one step further and estimate the random effects Tobit model. The results are:

```
. xtset i t
. xttobit x2 p2 p1 m, ll(0)

Random-effects tobit regression                 Number of obs      =      1510
Group variable: i                               Number of groups   =       176

Random effects u_i ~ Gaussian                   Obs per group: min =         8
                                                               avg =       8.6
                                                               max =        11

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(3)       =    605.11
Log likelihood  = -4663.2072                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          x2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          p2 |  -75.14353   4.942489   -15.20   0.000    -84.83063   -65.45643
          p1 |   9.896787   5.060785     1.96   0.051    -.0221691    19.81574
           m |   .3672872   .0639333     5.74   0.000     .2419803    .4925941
       _cons |   32.68706   6.512942     5.02   0.000     19.92193     45.4522
-------------+----------------------------------------------------------------
    /sigma_u |    44.0585   3.276081    13.45   0.000      37.6375     50.4795
    /sigma_e |   28.67666   .7433699    38.58   0.000     27.21968    30.13364
-------------+----------------------------------------------------------------
         rho |   .7024244   .0320737                      .6367994    .7620325
------------------------------------------------------------------------------
```

The importance of between-subject heterogeneity is clearly seen from the large and significant estimate of $\sigma_u$, of 44.06. The estimates are different as well: some of the slope estimates, particularly that of price of giving to other, take even larger values as a result of allowing for the heterogeneity.

### *15.2.3 Estimating the parameters of a CES utility function*

In this section we will use the data set introduced in the previous sub-section to estimate the parameters of a utility function for altruism.

Following Andreoni & Miller (2002) and others, we will assume the constant elasticity of substitution (CES) utility function:

$$U(x_1, x_2) = \left[\alpha x_1^{\rho} + (1-\alpha)x_2^{\rho}\right]^{\frac{1}{\rho}} \quad 0 \leq \alpha \leq 1 \quad -\infty \leq \rho \leq 1 \qquad (15.2)$$

The CES utility function (15.2) is used in many areas of economics. In the present setting, the parameter $\alpha$ indicates selfishness, while the parameter $\rho$ indicates willingness to trade off equity and efficiency in response to price changes. Values of $\rho$ less than zero indicate a concern for equality in payoffs; values of $\rho$ between zero and one indicate a focus on efficiency. The *elasticity of substitution*, usually denoted $\sigma$, may be deduced directly from $\rho$ using:

$$\sigma = \frac{1}{1-\rho} \tag{15.3}$$

$\sigma$ is clearly an increasing function of $\rho$, with values of $\rho$ between zero and one (indicating a focus on efficiency) being associated with values of $\sigma$ between one and $+\infty$.

A useful way of interpreting the elasticity of substitution, $\sigma$, is in terms of the curvature of the indifference curves. The larger is $\sigma$, the less curved the indifference curves become. As $\sigma$ approaches $+\infty$ the indifference curves become downward-sloping straight lines, implying that the two goods are perfect substitutes, and that all that matters is the total payoff. At the other extreme, if $\sigma$ approaches its lower limit of zero, the indifference curves become L-shaped, implying perfect complements, and that all that matters is equality of payoffs. The intermediate case is when $\sigma = 1$, implying Cobb-Douglas preferences: $U = x_1^{\alpha} x_2^{1-\alpha}$.

Maximising (15.2) subject to the budget constraint (15.1), we arrive at the "Marshallian demand function" for own pay-off:

$$w_1 = \frac{p_1^{\frac{\rho}{\rho-1}}}{p_1^{\frac{\rho}{\rho-1}} + \left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{\rho-1}} p_2^{\frac{\rho}{\rho-1}}} + \epsilon \tag{15.4}$$

where $w_1$ is, as previously mentioned, the share of the total allocation that is allocated to "self"; that is: $w_1 = \frac{p_1 x_1}{m}$. Note that a stochastic term ($\epsilon$) has been appended in (15.4) in order to turn the deterministic budget-share equation into an estimable model. The equation for the second budget share $w_2$ could easily be deduced from the deterministic part of (15.4) because $w_2 = 1 - w_1$. However, we only need one of the two equations to estimate the two parameters. We will use (15.4).

Non-linear least squares is required to estimate the two parameters in (15.4). The principle underlying non-linear least squares is exactly the same as ordinary least squares. If the sample is of size $n$ and the data set consists of the three variables $w_i, p_{1i}, p_{2i}, i = 1, \ldots, n$, the problem is to minimise the following sum of squares with respect to the two parameters $\alpha$ and $\rho$:

$$\sum_{i=1}^{n} \left[ w_{1i} - \frac{p_{1i}^{\frac{\rho}{\rho-1}}}{p_{1i}^{\frac{\rho}{\rho-1}} + \left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{\rho-1}} p_2^{\frac{\rho}{\rho-1}}} \right]^2 \tag{15.5}$$



The reason why non-linear least squares is required is that (15.4) is a non-linear function of the two parameters, and there is therefore no closed form expression for the solution to the minimisation problem, as there is when the model is linear. Instead, a numerical routine is used to locate the solution.

The STATA command which carries out non-linear least squares is `nl`. As with the regress command, it is possible to use the `vce(cluster i)` option to obtain cluster-robust standard errors. Another option we use here is "`initial`", in which starting values for the non-linear optimisation are provided. This option turns out to be essential – in the absence of starting values, estimation is not performed – although it is not necessary for the starting values to be particularly close to the solution.

The `nl` command with the two options just discussed, together with the results when applied to the Andreoni & Miller (2002) data, are:

```
. nl (w1 = (p1^({rho}/({rho}-1)))/((p1^({rho}/({rho}-1)))   ///
> +((({aa}/(1-{aa}))^(1/({rho}-1)))*(p2^({rho}/({rho}-1))))),  ///
> initial(rho 0.0 aa 0.5) vce(cluster i)
(obs = 1510)

Iteration 0:  residual SS =  122.2299
Iteration 1:  residual SS =  115.4766
Iteration 2:  residual SS =  115.4615
Iteration 3:  residual SS =  115.4615
Iteration 4:  residual SS =  115.4615

Nonlinear regression                              Number of obs =        1510
                                                  R-squared     =      0.8804
                                                  Adj R-squared =      0.8798
                                                  Root MSE      =    .2767056
                                                  Res. dev.     =    403.0932

                              (Std. Err. adjusted for 176 clusters in i)
------------------------------------------------------------------------------
             |               Robust
          w1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        /rho |    .272248    .0479813     5.67   0.000     .1775515    .3669445
         /aa |   .6918387    .0150264    46.04   0.000     .6621824     .721495
------------------------------------------------------------------------------

. nlcom sigma: 1/(1- _b[rho:_cons])

       sigma:  1/(1- _b[rho:_cons])

------------------------------------------------------------------------------
          w1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sigma |   1.374095    .0905952    15.17   0.000     1.196531    1.551658
------------------------------------------------------------------------------
```

Following the `nl` command, we apply the `nlcom` command to deduce an estimate of the elasticity of substitution, $\sigma$, using the formula (15.3). We see that this estimate is 1.37 and the confidence interval provides evidence that the elasticity of substitution is larger than one. This indicates that the subjects in this sample, overall, attach somewhat more importance to efficiency than to equality in payoffs. The estimate of $\alpha$ is 0.692. This may be interpreted as the proportion of the allocation that the individual would take for themselves in a situation of equal prices. Being significantly greater than 0.5 (again based on the confidence interval), this estimate indicates that subjects are relatively selfish.

## 15.3 A Model of Altruism with Binding Non-negativity Constraints

### 15.3.1 Background

In the previous section, we demonstrated how the parameters of a standard utility function can be estimated using data from a dictator game experiment. There, we did not take any account of zero-censoring, that is, the accumulation of observations at zero in the giving variable. In Section 15.2.2, in which we investigated the determinants of giving using ad hoc estimation methods, we found that disregarding the zero censoring had major consequences on estimation: Tobit estimates were very different from OLS estimates. From that result we concluded that it is important to deal with zero censoring.

The Tobit model is a useful framework for dealing with zero censoring. However, it is an *ad hoc* approach, since under normal circumstances it has no theoretical basis. In this section, the objective will be to incorporate the zero observations within the theoretical model, that is, within the constrained utility maximisation problem.

The model developed here is similar to that of Wales & Woodland (1983), who estimated the parameters of a utility function over three consumption goods.

### 15.3.2 The model

The variables are as defined in Section 15.2.1. Here we commence by assuming a Stone-Geary utility function:

$$U(x_1, x_2) = a_1 \ln(x_1 - b_1) + a_2 \ln(x_2 - b_2) \tag{15.6}$$

For identification of the parameters, it will be necessary to impose the normalisations:

$$a_1 + a_2 = 1 \qquad b_1 + b_2 = 0 \tag{15.7}$$

$b_1$ and $b_2$ are "subsistence levels" of $x_1$ and $x_2$. However, one of them is allowed to take a negative value, meaning that indifference curves cross axes. This is necessary in order to explain zero observations in $x_1$ or $x_2$. A Stone-Geary indifference map with $b_1 = 1$ and $b_2 = -1$ is illustrated in Figure 15.4.

The parameters $a_1$ and $a_2$ represent the proportion of *supernumerary income*, that is, income remaining when subsistence requirements have been satisfied, that is spent on each good. Hence these parameters may be interpreted as indicating the extent to which each good is a "luxury".

The constrained optimisation problem may be stated as:

$$\max_{x_1,x_2} U(x_1, x_2) \quad \text{s.t.} \quad p_1x_1 + p_2x_2 \le m, \quad 0 \le p_1x_1 \le m \tag{15.8}$$



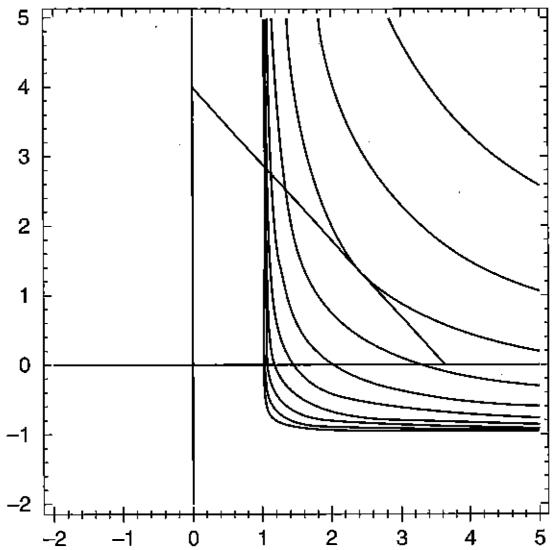



Figure 15.4: Stone-Geary indifference map with one subsistence level negative and one positive

The distinguishing feature of this model, apparent from (15.8), is that the utility function is being maximised subject to non-negativity constraints on both goods, in addition to a budget constraint.

The Lagrangian function is:

$$L = a_1 \ln(x_1 - b_1) + a_2 \ln(x_2 - b_2) + \lambda(m - p_1x_1 - p_2x_2) + \mu_1 p_1x_1 + \mu_2(m - p_1x_1) \tag{15.9}$$

whence the complementary slackness conditions are:

$$\begin{aligned} \frac{a_1}{p_1(x_1 - b_1)} &< \frac{a_2}{p_2(x_2 - b_2)} && \text{if } p_1x_1 = 0 \\ \frac{a_1}{p_1(x_1 - b_1)} &= \frac{a_2}{p_2(x_2 - b_2)} && \text{if } 0 < p_1x_1 < m \\ \frac{a_1}{p_1(x_1 - b_1)} &> \frac{a_2}{p_2(x_2 - b_2)} && \text{if } p_1x_1 = m \end{aligned} \tag{15.10}$$

Let us focus on "good 1". Combining the complementary slackness conditions (15.10) with the (binding) budget constraint ($p_1x_1 + p_2x_2 = m$), we obtain the well-known linear expenditure system (with non-negativity constraints):

$$p_1x_1 = \begin{cases} p_1b_1 + a_1(m - p_1b_1 - p_2b_2) & \text{if } 0 < \text{RHS} < m \\ 0 & \text{if } \text{RHS} \le 0 \\ m & \text{if } \text{RHS} \ge m \end{cases} \tag{15.11}$$

where "RHS" denotes the right-hand side of the first equation. The first equation of (15.11) has the following interpretation: expenditure on good 1 is subsistence expenditure plus a proportion $a_1$ of supernumerary income.

Applying the normalisations $a_1 + a_2 = 1$, $b_1 + b_2 = 0$, and after some rearranging, the three conditions (15.11) may be written:

$$\begin{aligned} b_1 &< \frac{-a_1 m}{(1-a_1)p_1 + a_1 p_2} && \text{if } p_1 x_1 = 0 \\ b_1 &= \frac{p_1 x_1 - a_1 m}{(1-a_1)p_1 + a_1 p_2} && \text{if } 0 < p_1 x_1 < m \\ b_1 &> \frac{(1-a_1)m}{(1-a_1)p_1 + a_1 p_2} && \text{if } p_1 x_1 = m \end{aligned} \tag{15.12}$$

We assume that the "subsistence level" for amount received by self, $b_1$, varies (within subjects) according to $b_1 \sim N(\gamma, \sigma^2)$. Henceforth, subscripts will be attached to each variable, $i$ for subject, and $t$ for task. The single-observation likelihood contributions for the three different regimes are (the reason for the $i$ subscript on $\gamma$ will become apparent shortly):

$$\begin{aligned} (p_1x_1)_{it} = 0: &\quad \Phi\left(\frac{\frac{-a_1 m_{it}}{(1-a_1)p_{1,it}+a_1p_{2,it}} - \gamma_i}{\sigma}\right) \\ 0 < (p_1x_1)_{it} < m_{it}: &\quad \frac{1}{\sigma}\phi\left(\frac{\frac{(p_1x_1)_{it}-a_1 m_{it}}{(1-a_1)p_{1,it}+a_1p_{2,it}} - \gamma_i}{\sigma}\right) \\ (p_1x_1)_{it} = m_{it}: &\quad 1-\Phi\left(\frac{\frac{(1-a_1)m_{it}}{(1-a_1)p_{1,it}+a_1p_{2,it}} - \gamma_i}{\sigma}\right) \end{aligned} \tag{15.13}$$

The single-observation likelihood contribution (conditional on $\gamma_i$) can then be defined as:

$$\begin{aligned} (f_{it}|\gamma_i) =& I\left((p_1x_1)_{it} = 0\right)\Phi\left(\frac{\frac{-a_1 m_{it}}{(1-a_1)p_{1,it}+a_1p_{2,it}} - \gamma_i}{\sigma}\right) \\ &+ I(0 < (p_1x_1)_{it} < m_{it})\frac{1}{\sigma}\phi\left(\frac{\frac{(p_1x_1)_{it}-a_1 m_{it}}{(1-a_1)p_{1,it}+a_1p_{2,it}} - \gamma_i}{\sigma}\right) \\ &+ I((p_1x_1)_{it} = m_{it})1-\Phi\left(\frac{\frac{(1-a_1)m_{it}}{(1-a_1)p_{1,it}+a_1p_{2,it}} - \gamma_i}{\sigma}\right) \end{aligned} \tag{15.14}$$

where $I(.)$ is the indicator function.

Between-subject heterogeneity may be introduced by allowing the mean parameter $\gamma$ to vary between subjects, hence the $i$ subscript on $\gamma$ in (15.13) and (15.14).



We will assume:

$$\gamma \sim N(\mu, \eta^2) \tag{15.15}$$

The subject likelihood contribution is then:

$$L_i = \int_{-\infty}^{\infty} \prod_{t=1}^{T} (f_{it}|\gamma) f(\gamma; \mu, \eta) d\gamma \tag{15.16}$$

where $(f_{it}|\gamma)$ is defined in (15.14) and $f(.; \mu, \eta)$ is the density function associated with the normal distribution (15.15). There are four parameters to be estimated: $a_1, \mu, \eta, \sigma$.

As usual, the integral (15.16) is evaluated by means of Halton draws. Specifically the integral will be replaced by the following mean over $R$ draws:

$$L_i = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}(f_{it}|\gamma_{r,i}) \tag{15.17}$$

where $\gamma_{r,i}$ is the $r$th Halton draw, transformed to be a realisation from the normal distribution (15.15).

Finally, having estimated the parameters, we shall obtain posterior expected subsistence levels for each subject, using the formula:

$$\hat{\gamma}_i = \hat{E}\left[\gamma_i|(p_1x_1)_{i1} \dots (p_1x_1)_{iT}\right] = \frac{\frac{1}{R}\sum_{r=1}^{R}\gamma_{r,i}\prod_{t=1}^{T}(\hat{f}_{it}|\gamma_{r,i})}{\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}(\hat{f}_{it}|\gamma_{r,i})} \tag{15.18}$$

where hats indicate that parameters have been replaced by **MLEs**.

### *15.3.3 Estimation*

The experimental setting is once again that of Andreoni & Miller (2002). The experiment was described in Section 15.2.2, and the data is contained in the file **garp**.

There are 234 subjects, each of whom performs up to 11 allocation tasks. The tasks, listed in Table 15.1, differ in the amount of the endowment, the price of self-allocation, and the price of other-allocation. The various plots shown in Section 15.2.2 were intended to give a feel for the data.

The STATA code is presented below. The likelihood evaluation program is named "sg". There are four parameters: `a1`, `mu`, `s_u` and `s_e`, which respectively correspond to the parameters appearing in the theoretical model of Section 15.3.2: $a_1, \mu, \eta, \sigma$. Note also that we obtain, for each subject, the posterior mean of $b_1$. This posterior mean is labelled "`b1_post`".

In the STATA code, we adopt the following names for the parameters and the other components of the log-likelihood function.

| Component of LogL | STATA name |
|---|---|
| $a_1$ | a1 |
| $\mu$ | mu |
| $\sigma_u$ | s_u |
| $\sigma_\epsilon$ | s_e |
| $\frac{p_1x_1-a_1m}{(1-a_1)p_1+a_1p_2}$ | w |
| $\frac{\frac{p_1x_1-a_1m}{(1-a_1)p_1+a_1p_2}-\gamma}{\sigma}$ | z |
| $I(p_1x_1=0)$ | d0 |
| $I(0<p_1x_1<m)$ | d_int |
| $I(p_1x_1=m)$ | dm |
| $f_{it,r}=I(p_1x_1=0)\Phi\left(\frac{\frac{-a_1m}{(1-a_1)p_1+a_1p_2}-\gamma_r}{\sigma}\right)$ $+I(0<p_1x_1<m)\phi\left(\frac{\frac{p_1x_1-a_1m}{(1-a_1)p_1+a_1p_2}-\gamma_r}{\sigma}\right)$ $+I(p_1x_1=m)\Phi\left(\frac{\frac{(1-a_1)m}{(1-a_1)p_1+a_1p_2}-\gamma_r}{\sigma}\right)$ | f |
| $\prod_{t=1}^{T} f_{it,r}$ | ff |
| $\gamma_r\times\prod_{t=1}^{T} f_{it,r}$ | gff |
| $\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T} f_{it,r}\right)$ | fff |
| $\frac{1}{R}\sum_{r=1}^{R}\left(\gamma_r\times\prod_{t=1}^{T} f_{it,r}\right)$ | gfff |
| $R$ | draws |
| $\ln L_i$ | lnfff |

The STATA code is presented below, with limited annotations. Readers requiring more detail should refer back to Chapter 10 in which similar estimation methods were explained in considerably more detail.

```
* LIKELIHOOD EVALUATION PROGRAM STARTS HERE

program define sg

args todo b logl
tempvar  gamma y d0 d_int dm w z p0 p_int pm f ff fff gff gfff
tempname a1 mu s_u s_e

local hlist h1*

mleval `a1'=`b', eq(1) scalar
mleval `mu' = `b', eq(2) scalar
mleval `s_u' = `b', eq(3) scalar
mleval `s_e' = `b', eq(4) scalar

quietly gen double `d0'=.
quietly gen double `d_int'=.
quietly gen double `dm'=.
quietly gen double `gamma'=.

quietly gen double `w'=.
quietly gen double `z'=.

quietly gen double `p0'=.
quietly gen double `p_int'=.
quietly gen double `pm'=.

quietly gen double `f'=.
quietly gen double `ff'=.
```



```
quietly gen double `gff'=.
quietly gen double `fff'=0
quietly gen double `gfff'=0

quietly{

replace `d0'=x1<=0
replace `d_int'=((p1*x1)>0)&((p1*x1)<m)
replace `dm'=p1*x1>=m

foreach v of varlist `hlist' {
replace `gamma'= `mu'+ `s_u'*`v'

replace `w'=(p1*x1-`a1'*m)/((1-`a1')*p1+`a1'*p2)
replace `z'=(`w'-`gamma')/(`s_e')
replace `p0'=normal(`z')
replace `p_int'=(1/`s_e')*normalden(`z')
replace `pm'=1-normal(`z')

replace  `f'= `d0'*`p0'+`d_int'*`p_int'+`dm'*`pm'
by i: replace `ff' = exp(sum(ln(max(`f',0.000000001))))
replace `gff'=`gamma'*`ff'
replace `ff'=. if last~=1
replace `gff'=. if last~=1
replace `fff'=`fff'+`ff'
replace `gfff'=`gfff'+`gff'

}

replace `fff'=`fff'/draws
replace `gfff'=`gfff'/draws

mlsum `logl'=ln(`fff') if last==1
}

quietly  replace g_post=`gfff'/`fff'

quietly{
putmata g_post, replace
}

end

* END OF LIKELIHOOD EVALUATION PROGRAM

*READ DATA

use "garp.dta", clear

*INITIALISE POSTERIOR MEAN OF b1

gen double g_post=.

drop if x1==.

bysort i: generate first=1 if _n==1 bysort i: generate last=1 if
_n==_N

* GENERATE HALTON SEQUENCES

mat p=[3] mdraws if first==1, neq(1) dr(32) prefix(h) primes(p)

scalar draws=r(n_draws)

local hlist h1*

quietly {
foreach v of varlist `hlist'{
```

```
by i: replace 'v'='v'[1] if 'v'==.
replace 'v'=invnorm('v')
}
}

* SET STARTING VALUES

mat start=(.7,5,10,10)

* RUN ML

ml model d0 sg /a1 /mu /s_u /s_e ml init start, copy ml max,
difficult

* EXTRACT POSTERIOR MEAN OF b1

drop g_post getmata g_post
```

The STATA output from running the above program is:

```
. ml max, difficult

                                          Number of obs   =       1510
                                          Wald chi2(0)    =          .
Log likelihood = -4402.2506               Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
a1           |
       _cons |   .5558002   .0176082    31.56   0.000     .5212889    .5903116
-------------+----------------------------------------------------------------
mu           |
       _cons |   29.77309    3.08699     9.64   0.000      23.7227    35.82348
-------------+----------------------------------------------------------------
s_u          |
       _cons |   33.99634   2.496788    13.62   0.000     29.10273    38.88995
-------------+----------------------------------------------------------------
s_e          |
       _cons |   25.14556   .6879349    36.55   0.000     23.79724    26.49389
------------------------------------------------------------------------------
```

### 15.3.4 Results

The results obtained above are collected in Table 15.2. The positive estimate of $\mu$ indicates that a typical subject has a positive subsistence level for "self", of around 30 units. However the high estimate of $\eta$ also indicates that the between-subject variation in subsistence level is high (to be confirmed in Figure 15.5 below). The estimate of $a_1$ is 0.55 indicating that 55% of supernumerary income will be spent on "self" and 45% on "other". The message appears to be that individuals are willing to share almost equally, *once their own "subsistence needs" have been fully met.*

Having estimated the parameters, we have obtained the posterior mean of "self-subsistence level" for each subject. The distribution of this quantity over the 176 subjects is shown in Figure 15.5. This indicates a large amount of heterogeneity. Subjects range from having self-subsistence levels close to zero, implying that they would under certain circumstances allocate zero to themsleves, to having self-subsistence levels close to 100 – the latter subjects would under no circumstances allocate zero to themselves.

Also evident from Figure 15.5 is the multi-modality. There are at least two clear modes: one around zero; the other around 80. This suggests that, in attitude towards



| Parameter | Estimate (asy.se) |
|---|---|
| $a_1$ | 0.5558(0.0176) |
| $\mu$ | 29.7731(3.0870) |
| $\eta$ | 33.9963(2.4968) |
| $\sigma$ | 25.1456(0.6879) |
| | |
| LogL | −4402.25 |
| $n$ | 176 |
| $T$(mean of) | 8.58 |

Table 15.2: Maximum likelihood estimates of parameters in the altruism model

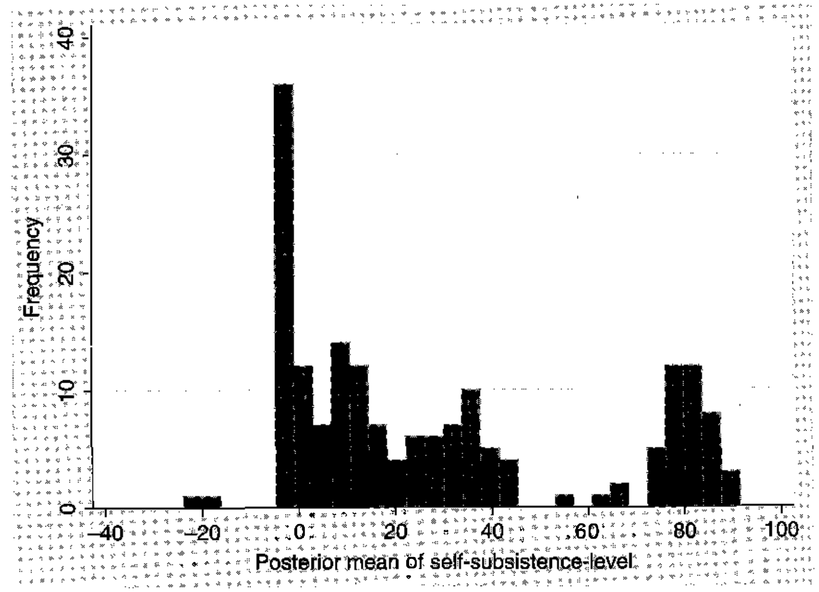



Figure 15.5: A Histogram of the posterior mean of self-allocation subsistence level

giving, the population divides into discrete "types", e.g. "unselfish" and "highly selfish". This finding provides clear motivation for the finite mixture modelling approach, which is the topic of the next section.

## 15.4 Finite Mixture Models of Altruism

This section presents another example of the use of the finite mixture framework introduced in Chapter 8. The application is to another version of the dictator game. The example closely follows the contribution of Cappelen et al. (2007).

### 15.4.1 The experiment

Consider a dictator game in which the distribution phase is preceded by a production phase. There are two players, player 1 and player 2. Each player is given an endowment of 30 at the start. Each player ($i$) is randomly assigned a rate of return (or an *ability level*), $a_i$. $a_i$ can be either 2 or 4. If $a_i$ is 2, any investment made by player $i$ is doubled; if $a_i$ is 4, any investment made by player $i$ is quadrupled. Players are then asked how much they would like to invest in the production phase. They must choose one of the amounts: 0, 10, 20. Let $q_i$ be the amount invested by player $i$. The total contribution by player $i$ is then $a_i q_i$.

The distribution phase follows. Subjects are paired randomly. Each is given information on the other's rate of return ($a$), investment level ($q$) and total contribution ($aq$).

Total earnings of the pair of players is given by:

$$X = a_1 q_1 + a_2 q_2 \tag{15.19}$$

Each player is asked how much of the total earnings they would like to have for themselves, on the understanding that the other player receives the remainder. Let $y_i$ be the amount that player $i$ chooses to take for himself. A coin is tossed to determine which player's suggested distribution of total earnings is put into effect.

In Section 15.4.2 we look at the data. Then we turn to the construction of finite mixture models for the determination of $y_i$.

### 15.4.2 The data

A data set that has been simulated in such a way as to resemble the data of Cappelen et al. (2007) is contained in the file **fairness_sim**. There are 190 observations, each containing a pair of players (player 1 and player 2).

Figure 15.6 contains two graphs of the data. The first shows how the amount claimed by player 1 is related to the total earnings of both players. When total earnings are low, most observations are on or close to the 45-degree line, implying a tendency to claim most of the total amount. When total earnings are higher, observations are further below the 45-degree line, implying more willingness to give. The second graph in Figure 15.6 shows a frequency histogram of the amount claimed by player 1 as a proportion of total earnings. The most important feature of this histogram is the high incidence of subjects claiming the whole amount for themselves. This feature of the data will be explained in terms of upper censoring.

### 15.4.3 A mixture model of fairness

The theoretical model developed here is similar to that used by Cappelen et al. (2007). Individuals wish to maximise their own pay-off, but at the same time they have concerns of fairness.



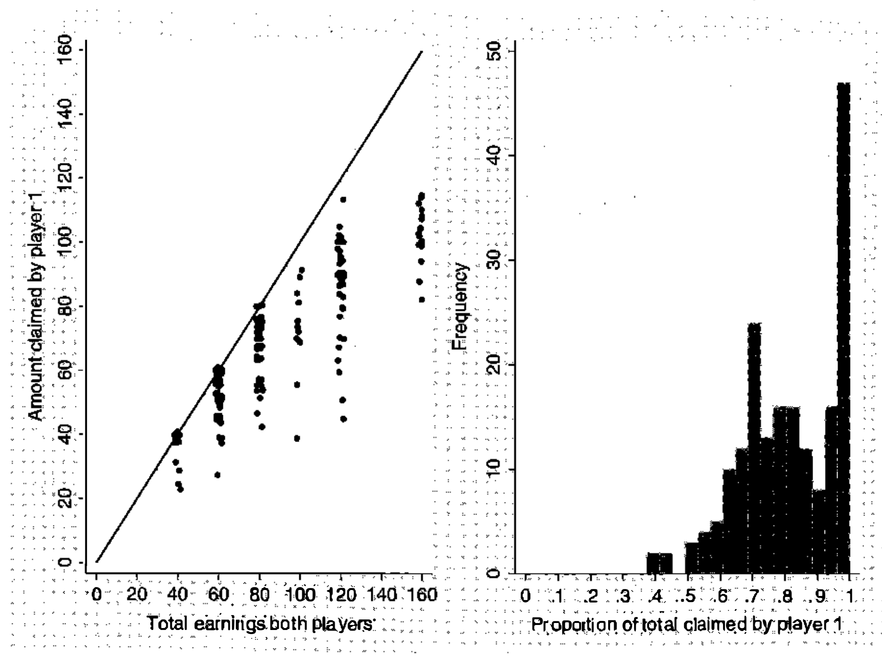



Figure 15.6: Left pane: amount claimed by player 1 against total earnings by both players; right pane: frequency histogram of proportion of total claimed by player 1

Let us assume that an individual's utility function is:

$$U^k = \alpha y - \frac{\beta}{2}(y - m^k)^2 \tag{15.20}$$

where: $k$ indicates which of three "fairness ideals" is being used (see below); $m^k$ is the value of $y$ that would be required to attain fairness ideal $k$; $\alpha$ is a positive parameter representing the importance the individual attaches to their own pay-off; $\beta$ is a positive parameter representing the importance the individual attaches to fairness.

The utility maximizing choice of $y$ is found by differentiating (15.20) and setting equal to zero:

$$\frac{\partial U^k}{\partial y} = \alpha - \beta(y - m^k) = 0 \tag{15.21}$$

$$\implies y = m^k + \frac{\alpha}{\beta}$$

So, the optimal choice of $y$ is simply the value of $y$ required to attain the fairness ideal, plus a fixed positive "selfishness premium" $\alpha/\beta$. Obviously the parameters $\alpha$ and $\beta$ are not separately identifiable; let us refer to the "selfishness premium" as $\delta$, so:

$$y = m^k + \delta \tag{15.22}$$

### 15.4.4 The "fairness ideals"

We assume that all subjects are motivated by one of three possible fairness ideals. Therefore there are three different subject types (player $i$ is the player under consideration; player $j$ is the other player):

| | | |
|---|---|---|
| Type 1 (Egalitarian): | $m^E = \frac{X}{2}$ | (i.e. wants to share total earnings equally, with no regard to who has generated the earnings) |
| Type 2 (Libertarian): | $m^L = a_i q_i$ | (i.e. believes that players should be paid according to what they have earned themselves) |
| Type 3 (Liberal Egalitarian): | $m^{LE} = \frac{q_i}{q_i+q_j} X$ | (i.e. believes that players should be paid in proportion to the amount they have invested ($q$), but that their rate of return ($a$) should not matter). |

### 15.4.5 The econometric model

Applying (15.22) to each of the three fairness ideals defined in Section 15.4.4, and appending an error term, we arrive at the following econometric model:

Type 1 (Egalitarian):

$$y_i = \frac{X_i}{2} + \delta + \epsilon_{1,j} \tag{15.23}$$

Type 2 (Libertarian):

$$y_i = a_i q_i + \delta + \epsilon_{2,j} \tag{15.24}$$

Type 3 (Liberal Egalitarian):

$$y_i = \left(\frac{q_i}{q_i + q_j}\right) X_i + \delta + \epsilon_{3,j} \tag{15.25}$$

For the sake of simplicity, we assume that all three errors have the same variance:

$$V(\epsilon_{1,j}) = V(\epsilon_{2,j}) = V(\epsilon_{3,j}) = \sigma^2 \tag{15.26}$$

Let the mixing proportions be: $p_1, p_2, p_3$. The likelihood contribution for subject $i$ is:



$$p_1 \frac{1}{\sigma} \phi\left(\frac{y_i - \frac{X_i}{2} - \delta}{\sigma}\right) + p_2 \frac{1}{\sigma} \phi\left(\frac{y_i - a_i q_i - \delta}{\sigma}\right) + (1 - p_1 - p_2) \frac{1}{\sigma} \phi\left(\frac{y_i - \left(\frac{q_i}{q_i+q_j}\right) X_i - \delta}{\sigma}\right) \tag{15.27}$$

The parameters to be estimated are: $\delta$, $\sigma$, $p_1$, $p_2$.

Recall that there are a number of subjects for whom $y_i = X_i$; i.e. they demand all of the total earnings for themselves, and wish to give nothing to the other player. It is best to treat these observations as right-censored.

Upper censored observations have the following likelihood contribution:

$$p_1 \Phi\left(\frac{\frac{X_i}{2} + \delta - X_i}{\sigma}\right) + p_2 \Phi\left(\frac{a_i q_i + \delta - X_i}{\sigma}\right) + (1 - p_1 - p_2) \Phi\left(\frac{\left(\frac{q_i}{q_i+q_j}\right) X_i + \delta - X_i}{\sigma}\right) \tag{15.28}$$

### 15.4.6 The program and results

The following code contains the likelihood evaluation program for the model both without and with the allowance for right censoring mentioned at the end of Section 15.4.5. The two models are then estimated. The program is:

```
*PROGRAM WITHOUT DEALING WITH UPPER CENSORING

prog drop _all program define fairness1
args lnf d sig p1 p2
tempvar f1 f2 f3

quietly gen double `f1'=(1/`sig')*normalden((y-`d'-me)/`sig')
quietly gen double `f2'=(1/`sig')*normalden((y-`d'-ml)/`sig')
quietly gen double `f3'=(1/`sig')*normalden((y-`d'-mle)/`sig')

quietly replace `lnf'=ln(`p1'*`f1'+`p2'*`f2'+(1-`p1'-`p2')*`f3')

quietly replace postp1=(`p1'*`f1')/(`p1'*`f1'+`p2'*`f2'+(1-`p1'-`p2')*`f3')
quietly replace postp2=(`p2'*`f2')/(`p1'*`f1'+`p2'*`f2'+(1-`p1'-`p2')*`f3')
quietly replace postp3=((1-`p1'-`p2')*`f3')/(`p1'*`f1'+`p2'*`f2'+(1-`p1'-`p2')*`f3')

quietly putmata postp1, replace
quietly putmata postp2, replace
quietly putmata postp3, replace

end

*PROGRAM DEALING WITH UPPER CENSORING

program define fairness2
args lnf d sig p1 p2
tempvar y f1 f2 f3
```

```
quietly gen double ‘f1’=(1/‘sig’)*normalden((y-‘d’-me)/‘sig’) if y<x
quietly replace  ‘f1’=normal((‘d’+me-x)/‘sig’) if y==x
quietly gen double ‘f2’=(1/‘sig’)*normalden((y-‘d’-ml)/‘sig’) if y<x
quietly replace  ‘f2’=normal((‘d’+ml-x)/‘sig’) if y==x
quietly gen double ‘f3’=(1/‘sig’)*normalden((y-‘d’-mle)/‘sig’) if y<x
quietly replace  ‘f3’=normal((‘d’+mle-x)/‘sig’) if y==x

quietly replace ‘lnf’=ln(‘p1’*‘f1’+‘p2’*‘f2’+(1-‘p1’-‘p2’)*‘f3’)

quietly replace postp1=(‘p1’*‘f1’)/(‘p1’*‘f1’+‘p2’*‘f2’+(1-‘p1’-‘p2’)*‘f3’)
quietly replace postp2=(‘p2’*‘f2’)/(‘p1’*‘f1’+‘p2’*‘f2’+(1-‘p1’-‘p2’)*‘f3’)
quietly replace postp3=((1-‘p1’-‘p2’)*‘f3’)/(‘p1’*‘f1’+‘p2’*‘f2’+(1-‘p1’-‘p2’)*‘f3’)

quietly putmata postp1, replace
quietly putmata postp2, replace
quietly putmata postp3, replace

end

* READ DATA

use fairness_sim, clear

*GENERATE VARIABLES REPRESENTING FAIRNESS IDEALS (me, ml, mle)

gen me=x/2 gen ml=a1q1 gen mle=(q1/(q1+q2))*x

* INITIALISE VARIABLES REPRESENTING POSTERIOR TYPE PROBABILITIES

gen postp1=. gen postp2=. gen postp3=.

* SET STARTING VALUES

mat start=(20,10,.5,.2)

* ESTIMATE MODEL WITHOUT DEALING WITH CENSORING

ml model lf fairness1 /d /sig /p1 /p2 ml init start, copy ml
maximize nlcom p3: 1- _b[p1:_cons]- _b[p2:_cons]

* ESTIMATE MODEL DEALING WITH CENSORING

ml model lf fairness2 /d /sig /p1 /p2 ml init start, copy ml
maximize nlcom p3: 1- _b[p1:_cons]- _b[p2:_cons]

* EXTRACT AND PLOT POSTERIOR TYPE PROBABILITIES

drop postp1 postp2 postp3

getmata postp1 getmata postp2 getmata postp3

label variable postp1 "probability egalitarian" label variable
postp2 "probability libertarian" label variable postp3 "probability
lib-egalitarian"

scatter postp2 postp1
```

Results from the two models are presented in Table 15.3. Consider first the results from the model without right censoring. We see that type 1 (egalitarian) is the dominant type, with 75% of the population being of this type. The remainder of the population divide roughly evenly between type 2 (libertarian) and type 3 (liberal egalitarian). We also see that the "selfishness premium" is estimated to be 24.01, meaning that a typical subject will claim 24 units more than what their "fairness norm" would dictate.

Turning to the model with right censoring, we see that dealing with censoring has reduced the estimate of the proportion of egalitarians, with a corresponding



| | Without censoring | With censoring |
|---|---|---|
| $\delta$ | 24.012(0.710) | 25.969(0.871) |
| $\sigma$ | 8.654(0.520) | 10.267(0.717) |
| $p_1$ | 0.751(0.078) | 0.724(0.091) |
| $p_2$ | 0.137(0.073) | 0.181(0.090) |
| $p_3$ | 0.112(0.097) | 0.095(0.107) |
| $n$ | 190 | 190 |
| $LogL$ | −702.53 | −604.50 |

Table 15.3: MLEs from finite mixture models without and with right censoring
Note: the estimate of $p_3$ has been obtained using the delta method.

increase in the estimate of the proportion of libertarians. We also note that the estimate of the selfishness premium has also risen. All of these differences are explained in terms of upper-censored observations representing constrained selfishness. By building in the censoring to the model, more widespread selfishness (e.g. libertarianism; higher selfishness premium) is revealed.

Finally, the posterior probabilities (from the model with right censoring) are plotted in Figure 15.7. The probabilities of the first two types are measured on the axes, meaning that the probability of the third type (liberal egalitarian) is represented by the distance from the downward-sloping line.

The majority of observations being in the bottom-right region of the triangle is consistent with our conclusion that the majority of the population are egalitarian. A small number of subjects have a high probability of being libertarian. No subjects appear to have a high probability of being liberal egalitarian (although recall that we predict 9% of the population to be liberal egalitarian).

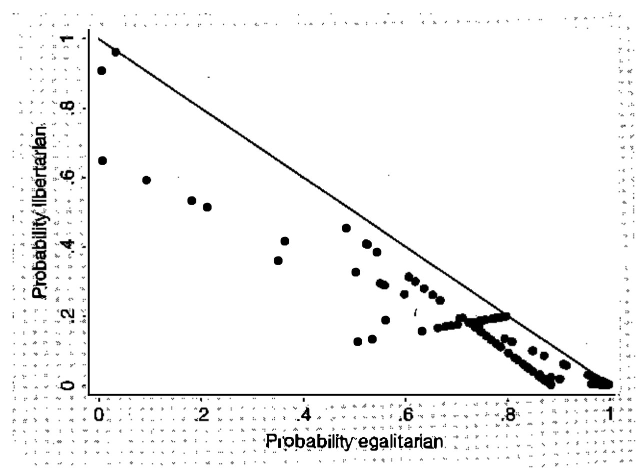



Figure 15.7: Posterior type probabilities in fairness experiment

## 15.5 Estimation of Social Preference Parameters Using Discrete Choice Models

### 15.5.1 The setting

In this section, we once again pursue the goal of estimating the parameters of a utility function involving own payoff and other's payoff. However, the experimental approach is very different. Here, we are asking subjects to choose between hypothetical allocations. The approach is very similar to that of Engelmann & Strobel (2004).

The task faced by a subject is to choose between three different (hypothetical) allocations, A, B and C, say, of the type shown in Table 15.4. Importantly, the subject is always given the identity of "Person 2".

The three allocations in the example of Table 15.4 all have both attractions and drawbacks, and it is reasonable to expect people to divide between the three when asked to choose between them. A selfish individual would prefer allocation A, since, given their assumed identity as "Person 2", allocation A gives them a pay-off of 8 which is higher than the pay-off they would recieve under B or C (6 and 7 respectively).

However, as we have seen many times in this book, not all individuals are selfish, and some are highly altruistic. An individual who is "inequity-averse" is likely to choose allocation B, since equality is perfect under this allocation, with all three persons receiving the same pay-off.

In fact, there are different types of inequity aversion. It is useful to make a distinction between "disadvantageous" and "advantageous" inequality – a distinction highlighted by Fehr & Schmidt (1999). To see the difference, we continue with the example in Table 15.4, and consider a situation in which allocation B is not available, and individuals are required to choose between A and C. Both A and C are unequal allocations. An individual who is averse to "disadvantageous" inequality is likely to choose A over C, since C is an allocation in which they (Person 2) are *disadvantaged* relative to Person 1. An individual who is averse to "advantageous" inequality is likely to choose C over A, since A is an allocation in which they (Person 2) are *advantaged* relative to Person 3.

Some individuals are neither self-interested, nor overly concerned about inequity, but instead are motivated by *efficiency* considerations. Such individuals simply look for the allocation that is most efficient in terms of generating the highest total pay-off for the group. We see that when faced with the allocations in Table

| Allocation | A | B | C |
|---|---|---|---|
| Person 1 | 8 | 6 | 10 |
| Person 2 | 8 | 6 | 7 |
| Person 3 | 4 | 6 | 7 |
| Total | 20 | 18 | 24 |

Table 15.4: An example of three hypothetical allocations. The survey respondent is given the identity of "Person 2"



15.4, such an individual would choose allocation C since this gives the highest total pay-off of 24.

Of course, it is unlikely that individuals are motivated by only one of the concerns described above. It is more likely that all of the concerns matter, perhaps to varying degrees. Hence we set out to estimate a utility function involving all of the concerns, and applying to all individuals. The parameter estimates will convey the importance of each concern.

### 15.5.2 Formalising the criteria for choosing between allocations

Let $x_{jk}$ be the payoff to person $k$ in allocation $j$. Each allocation has the following attributes.

1. Efficiency: $EFF_j = \sum_{k=1}^{3} x_{jk}$

Efficiency is simply the sum of the payoffs over all persons, regardless of the distribution. For the example in Table 15.4, the efficiency attributes for the three allocations are:

$$EFF_A = 20;\ EFF_B = 18;\ EFF_C = 24$$

2. Minimax: $MM_j = min(x_{jk}, k = 1, 2, 3)$

Minimax is the smallest payoff in the allocation. If an individual uses this as a criterion, they are displaying an extreme form of inequality aversion since they are only ever concerned with the welfare of the worst-off person. For the example in Table 15.4, the minimax attributes for the three allocations are:

$$MM_A = 4;\ MM_B = 6;\ MM_C = 7$$

3. Self: $SELF_j = x_{j2}$

"Self" is the decision-maker's own payoff. Remember that the decision-maker is assumed to take the role of Person 2; hence the definition of "self" as the payoff of Person 2 in the allocation. An individual using "self" as their criterion is clearly a selfish individual since they are concerned with their own welfare and have no regard for the welfare of any other individual. For the example in Table 15.4, the "self" attributes for the three allocations are:

$$SELF_A = 8;\ SELF_B = 6;\ SELF_C = 7$$

The final two attributes that we consider derive from the well known "Fehr-Schmidt utility function" (Fehr & Schmidt, 1999). If there are $n$ persons in total, this utility function, for person $i$, is:

$$u_i = x_i - \alpha_i \frac{\sum_{k \neq i} \max(x_k - x_i, 0)}{n-1} - \beta_i \frac{\sum_{k \neq i} \max(x_i - x_k, 0)}{n-1} \qquad (15.29)$$

The interpretation of (15.29) is as follows. Individual i's utility is given by their own payoff, penalised by the presence of two different types of inequality. The second term on the right-hand side is the term that adjusts for "disadvantageous inequality"; that is, inequality resulting from others receiving *higher* payoffs than themselves. The final term is the term that adjusts for "advantageous inequality"; that is, inequality resulting from others receiving *lower* payoffs than themselves. It is usually assumed that both types of inequality are undesirable, so the two parameters $\alpha_i$ and $\beta_i$, respectively referred to as individual $i$'s coefficients of disadvantageous and advantageous inequity aversion, are both expected to be positive. It is also assumed that individuals care more about disadvantageous inequity than about advantageous inequity, so that $\alpha_i > \beta_i$. The denominators of the two inequity terms are present simply to ensure that the measures to not rise with the number of persons in the "economy".

In the case $n = 3$, the two measures of inequality just described become:

4a. (absence of) Disadvantageous inequality: $FSD_j = -\frac{1}{2}\sum_{k\neq 2} max(x_{jk}-x_{j2}, 0)$
4b. (absence of) Advantageous inequalty: $FSA_j = -\frac{1}{2}\sum_{k\neq 2} max(x_{j2} - x_{jk}, 0)$

FSD and FSA are the attributes in which we are interested. In these acronyms, "FS" stands for Fehr-Schmidt, and "D" and "A" for disadvantageous and advantageous respectively. Defining them with a negative sign allows them to be treated as positive attributes, that is, quantities that an inequity-averse individual would seek to maximise.

For the example in Table 15.4, the attributes FSD and FSA for the three allocations are:

$$FSD_A = 0; \quad FSD_B = 0; \quad FSD_C = -\frac{3}{2}$$
$$FSA_A = -2; \quad FSA_B = 0; \quad FSA_C = 0$$

### 15.5.3 Data

Simulated data closely resembling Engelmann and Strobel's data is contained in the file **ES_sim**. Note that there are three rows for each subject, one row for each allocation (this is known as a "long" data set). The attributes are named as in Section 15.5.2. $y$ is a binary variable indicating which of the three allocations is chosen (1 if chosen; 0 if not chosen).

### 15.5.4 The conditional logit model (CLM)

Let us henceforth use $i$ to index an individual in the data set. Each individual chooses one from $J = 3$ possible allocations. Suppose that individual $i$'s utility from choosing allocation $j$ is given by:

$$U_{ij} = \alpha_1 FSD_{ij} + \alpha_2 + FSA_{ij} + \alpha_3 + EFF_{ij} + \alpha_4 MM_{ij} + \epsilon_{ij} = z'_{ij}\alpha + \epsilon_{ij} \quad (15.30)$$



Note that the attribute variables have both $i$ and $j$ subscripts, indicating that different individuals face allocations with different sets of attributes. Note also that there is no intercept in (15.30). This is because an intercept parameter would not be identified, because, as is well known, adding a constant to a utility function does not alter implied behaviour. For convenience, we have collected the list of attributes together in the vector $z_{ij}$ and the associated parameters into the vector $\alpha$. The term $z'_{ij}\alpha$ may be referred to as the deterministic component of utility, and $\epsilon_{ij}$ as the random component.

We then assume that each individual chooses the allocation that yields the highest utility. Formally, the *observed* decision variable is $y_{ij}$, and:

$$\begin{aligned} y_{ij} &= 1 \quad \text{if} \quad U_{ij} = max\,(U_{i1}, U_{i2} \ldots U_{iJ}) \\ y_{ij} &= 0 \quad \text{otherwise} \end{aligned} \quad (15.31)$$

We then need to consider what is the probability that individual $i$ will choose allocation $j$. This depends on what is assumed about the distribution of the random component of utility.

$$\begin{aligned} y_{ij} = 1 &\Leftrightarrow z'_{ij}\alpha + \epsilon_{ij} > z'_{ik}\alpha + \epsilon_{ik} \quad \forall k \neq j \\ &\Leftrightarrow \epsilon_{ik} - \epsilon_{ij} < z'_{ij}\alpha - z'_{ik}\alpha \quad \forall k \neq j \end{aligned} \quad (15.32)$$

For convenience, it is assumed that the $\epsilon_{ij}$s are independent and identically distributed (i.i.d.) with type I extreme value distribution (also known as the Gumbel distribution), defined by the density function:

$$f(\epsilon) = \exp\,(-\epsilon - \exp(-\epsilon)) \quad -\infty < \epsilon < \infty \quad (15.33)$$

and the distribution function:

$$F(\epsilon) = \exp\,(-\exp(-\epsilon)) \quad -\infty < \epsilon < \infty \quad (15.34)$$

It can be shown (Maddala, 1983) that if the $\epsilon_{ij}$s are i.i.d. and have the distribution defined in (15.33) and (15.34), then the probability of the event defined in (15.32), that is, the probability of allocation $j$ being chosen by individual $i$, is given by:

$$P(y_{ij} = 1) = \frac{exp(z'_{ij}\alpha)}{\sum_{k=1}^{J} exp(z'_{ik}\alpha)} \quad (15.35)$$

The model defined by (15.35) is known as the *conditional logit model.* The likelihood contribution associated with individual $i$ is given by:

$$L_i(\alpha) = \frac{\sum_{k=1}^{J} y_{ik} exp(z'_{ik}\alpha)}{\sum_{k=1}^{J} exp(z'_{ik}\alpha)} \quad (15.36)$$

from which the sample log-likelihood is obtained as:

$$LogL(\alpha) = \sum_{i=1}^{n} \ln L_i(\alpha) \quad (15.37)$$

### 15.5.5 Results

The parameters contained in the vector $\alpha$ in (15.30) are estimated by maximisation of the likelihood function defined in (15.36) and (15.37) using the following STATA command:

```
. asclogit y FSD FSA EFF MM, case( i) alternatives(j) noconstant
```

The as at the beginning of the command name stands for "alternative-specific".

The output looks like this:

```
Iteration 0:   log likelihood = -317.10088
Iteration 1:   log likelihood = -308.55197
Iteration 2:   log likelihood = -308.51212
Iteration 3:   log likelihood = -308.51212

Alternative-specific conditional logit         Number of obs      =       990
Case variable: i                               Number of cases    =       330

Alternative variable: t                        Alts per case: min =         3
                                                              avg =       3.0
                                                              max =         3

                                                  Wald chi2(4)    =     80.96
Log likelihood = -308.51212                       Prob > chi2     =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
t            |
         FSD |   .3267221    .1405881     2.32   0.020     .0511745    .6022697
         FSA |   .3447768    .1688655     2.04   0.041     .0138065    .6757472
         EFF |   .1879009    .0714842     2.63   0.009     .0477943    .3280074
          MM |   .0804075    .0895162     0.90   0.369    -.0950409     .255856
------------------------------------------------------------------------------
```

The two inequity aversion attributes have been included, as well as the efficiency and minimax attributes. It appears (from this simulated data set) that subjects display both types of inequity aversion: both FSD and FSA have significantly positive effects on utility. Efficiency appears to be even more important: the coefficient of EFF is strongly significantly positive. The minimax attribute (MM) does not appear to be important.

Recall that another attribute was SELF defined as payoff to self. It turns out that the experimental design is such that when SELF is added to the above model, the problem of perfect multicollinearity arises resulting in a failure to estimate the effect of this variable. This is why SELF has been excluded.

### 15.5.6 The effect of subject characteristics

It is reasonable to expect individuals to value different criteria differently. In previous chapters, we have allowed for this type of difference between subjects by building *unobserved heterogeneity* into models. Here we will demonstrate an alternative approach: *observed heterogeneity*. Observed heterogeneity refers to the situation in which differences between subjects can be explained by differences in subject characteristics, perhaps the most obvious being gender.



The way in which subject characteristics are introduced in the CLM is through *interactions* with the attribute variables. Let the variable $male_i$ be a dummy variable taking the value 1 if subject $i$ is male. An important point is that subject characteristics have only an $i$ subscript, as distinct from the attributes which have both an $i$ and a $j$ subscript.

Let us extend the utility function (15.30) with two additional terms:

$$U_{ij} = \alpha_1 FSD_{ij} + \alpha_2 FSD_{ij} \star \text{male}_i + \alpha_3 FSA_{ij} + \alpha_4 FSA_{ij} \star \text{male}_i + \alpha_5 EFF_{ij} + \alpha_6 MM_{ij} + \epsilon_{ij} \qquad (15.38)$$

The two additional terms are interactions between the dummy variable male and the two inequity attributes. When these two variables are included in the model, the results are as follows:

```
Alternative-specific conditional logit         Number of obs      =       990
Case variable: i                               Number of cases    =       330

Alternative variable: j                        Alts per case: min =         3
                                                              avg =       3.0
                                                              max =         3

                                                  Wald chi2(6)    =     85.42
Log likelihood = -299.6794                        Prob > chi2     =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
j            |
         FSD |   .1907648    .1552983     1.23   0.219    -.1136143     .495144
    male_FSD |   .2535549    .1281861     1.98   0.048     .0023147     .504795
         FSA |   .5649655    .1879811     3.01   0.003     .1965293    .9334017
    male_FSA |  -.5760542     .192775    -2.99   0.003    -.9538863   -.1982221
         EFF |   .1606768    .0741216     2.17   0.030     .0154012    .3059525
          MM |   .1170375     .091562     1.28   0.201    -.0624207    .2964958
------------------------------------------------------------------------------
```

Here (not forgetting that the data is simulated), we see a very interesting result: the significantly positive coefficient on the interaction male_FSD indicates that males exhibit more aversion to disadvantageous inequity than do females; the significantly negative coefficient on male_FSA indicates that females exhibit more aversion to advantageous inequity than do males.

## 15.6 Summary and Further Reading

The theme of this chapter has been parametric estimation of utility functions for altruism. The utility functions have usually had two arguments, own payoff and other's payoff. Andreoni & Miller's (2002) data played a central role in this chapter. The focus of Andreoni & Miller's (2002) article was, in fact, the testing of the axioms of revealed preference, which may be seen as preliminary to the sort of analysis performed in this chapter. If subjects adhere to the axioms of revealed preference (which Andreoni & Miller (2002) find that the vast majority of subjects do), then it can be inferred that their choices could have been generated by a monotonic, continuous, and convex utility function, which is essentially the starting point of this chapter. In Section 15.2 we showed exactly how to estimate the parameters of a CES utility function for altruism.

Another key finding of Andreoni & Miller (2002) was heterogeneity in altruism. In this chapter, heterogeneity has been addressed in many different ways. Firstly, early in Section 15.2, we applied the random effects Tobit model to the data on giving, with zero giving being explained by censoring. This sort of approach to dealing with zeros has been used in the literature, for example by Fisman et al. (2007). However, we have classified such methods as *ad hoc*. This provided the motivation for Section 15.3 where we developed a model in which accumulations of zero observations in the giving data are embraced by treating zeros as "corner solutions" to the subject's constrained optimisation problem. This is an approach introduced in the context of a quadratic utility function by Wales & Woodland (1983) and amended to the Stone-Geary utility function by Moffatt (1991). To our knowledge, this is the first application of this type of model to experimental data. In the version of the model developed here, between-subject variability was applied to particular parameters representing attitude to giving, and significant heterogeneity was found.

Another attempt to deal with heterogeneity was in Section 15.4 in which we developed a finite mixture model. This model was very closely related to the model of Cappelen et al. (2007), who were concerned with separation of subjects between different fairness ideals. It should be said that the method of estimation used here was very different from that used by Cappelen et al. (2007). In fact, an array of econometric modelling approaches are available for estimating this type of model, each with advantages and drawbacks, and the interested reader is referred to Conte & Moffatt (2014).

Yet another reference to heterogeneity was in Section 15.5 in which choice data was used to estimate the parameters of utility functions. Here, we allowed for *observed heterogeneity*, by forming interaction variables that combined the attributes of the alternatives with subject characteristics. Hence we demonstrated a way of detecting whether certain types of subject attach more importance than others to particular attributes.

The choice modelling approach described in Section 15.5 was particularly useful for estimating the parameters of the well-known Fehr & Schmidt (1999) utility function. These parameters are known as inequity aversion parameters, and they form the basis of many theoretical models of social preferences. Because the indifference curves of the Fehr & Schmidt (1999) utility function are piece-wise linear, the solution to the constrained optimisation problem will always be a corner solution (divide equally, give nothing, or give everything). Because of this, direct estimation of the parameters of this utility function is not possible using the approaches followed in Sections 15.2 and 15.3 of this chapter. However, by treating the inequity measures as attributes in a choice model, as we have seen in Section 15.5, it becomes possible to obtain estimates of these parameters.

A number of other "two-good" utility functions for altruism appear in the literature, and have gained popularity. These include "Equity, Reciprocity and Competition (ERC)" (Bolten & Ockenfels, 2000) and "warm glow" (Andreoni, 1988).

A potentially useful extension to the type of models estimated in this chapter is to allow the parameters of the utility function to vary by treatment. For example Jakiela (2013) considers a data set similar to that of Andreoni & Miller (2002), but including a "take treatment", and finds that certain CES parameters are different under this treatment.



# Exercises

1. Let $x_1$ be own payoff, and let $x_2$ be other's payoff. Consider the constant elasticity of substitution (CES) utility function analysed in Section 15.2:

$$U(x_1, x_2) = \left[\alpha x_1^{\rho} + (1-\alpha)x_2^{\rho}\right]^{\frac{1}{\rho}} \quad 0 \leq \alpha \leq 1 \quad -\infty \leq \rho \leq 1$$

Maximise $U(x_1, x_2)$ subject to the budget constraint:

$$p_1x_1 + p_2x_2 \leq m \ .$$

to obtain the "Marshallian demand fuction" for own pay-off:

$$w_1 = \frac{p_1^{\frac{\rho}{\rho-1}}}{p_1^{\frac{\rho}{\rho-1}} + \left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{\rho-1}} p_2^{\frac{\rho}{\rho-1}}}$$

where $w_1$ is the share of the total allocation that is allocated to "self"; that is: $w_1 = \frac{p_1x_1}{m}$.

2. When analysing Andreoni & Miller's (2002) data in Section 15.2, it was possible to estimate both parameters of the CES utility function, by virtue of the design including variation in both endowment and prices. Consider the more standard design in which only endowment varies but the two prices are fixed at unity. Which of the two parameters, if either, would it then be possible to estimate with the resulting data?

3. a. Suppose that that an individual chooses one from $J$ possible alternatives, and that their utility from choosing alternative $j$ is given by:

$$U_j = z'_j\alpha + \epsilon_j \tag{15.39}$$

where $z_j$ is a vector of attributes of alternative $j$, $\alpha$ is a corresponding vector of parameters, and $\epsilon_j$, $j = 1, \ldots, J$ are i.i.d. with type I extreme value distribution defined by the c.d.f.:

$$F(\epsilon) = \exp\left(-\exp(-\epsilon)\right) \quad -\infty < \epsilon < \infty. \tag{15.40}$$

If the individual chooses the alternative with the highest utility, prove that the probability of choosing alternative $j$ is given (as in (15.35) in Section 15.5.4) by:

$$P(y_{ij}=1)=\frac{exp(z'_{ij}\alpha)}{\sum_{k=1}^{J} exp(z'_{ik}\alpha)} \tag{15.41}$$

b. Verify that if the parameter vector $\alpha$ appearing in (15.39) contains an intercept $\alpha_0$, the choice probability (15.41) is invariant to the value of $\alpha_0$, and the parameter $\alpha_0$ is therefore not identified.

4. In the context of the discrete choice model developed in Section 15.5, verify that $FS_D + FS_A + SELF = EFF$. Hence explain why the four variables cannot appear together in the same model.

# Chapter 16
# Repeated Games and Quantal Response Models

## 16.1 Introduction

This is the first of three chapters to analyse data from interactive games. It starts off by considering the setting of a repeated game characterised by a mixed strategy Nash Equilibrium. The central question to be addressed is how closely players adhere to the mixed strategy Nash Equilibrium. This question can be answered by virtue of observing long sequences of repeated plays by fixed pairs of players. Related questions that will be addressed include whether players are making choices randomly – another prediction of the theory. Such questions will be addressed using simple non-parametric procedures.

The chapter progresses to a treatment of a very well-known model of bounded rationality: the quantal response equilibrium (QRE) model. In practice, players do not play Nash. QRE is an econometric model built on two assumptions: that players depart from the Nash prediction in a random manner; and that each player forms *correct* beliefs about the randomness in others' decisions. Hence QRE may be thought of as a stochastic generalisation of the Nash Equilibrium. The basic version of the model contains only one parameter, $\mu$, representing the amount of "noise" in subjects' decisions. In keeping with the rest of this book, the principal objective will be to demonstrate how to estimate $\mu$ using data on subjects' choices. This estimation problem is non-standard because, while for standard models, choice probabilities can be expressed in closed form for given parameter values, for this problem the choice probabilities for any given value of $\mu$ can only be obtained as the solution of a non-linear optimisation routine. Hence this non-linear optimisation routine needs to be called within the function evaluation program.

In Section 16.3 we present the theory of QRE and describe a method for computing the choice probabilities. In Section 16.4 we estimate the QRE model on real data from the "pursue-evade" game. In Section 16.5, we extend the QRE model to allow estimation of a risk-aversion parameter in addition to the noise parameter. We conclude the chapter in Section 16.6 by applying the QRE model to contest data.

## 16.2 Analysis of Repeated Game Data

### 16.2.1 Computing a mixed strategy Nash equilibrium

As an example we shall use the "pursue-evade" game, as analysed by Rosenthal et al. (2003). Probably the best-known real-life application of this game is to penalty-taking in soccer: a goalkeeper prefers to dive to the same side as the penalty-taker kicks, and is therefore a "pursuer"; the penalty-taker prefers to kick the ball the opposite side to the one to which the goalkeeper dives, and is therefore an "evader". Here, we describe the game, and derive the mixed strategy Nash equilibrium. In the next sub-section, we apply various non-parametric tests to test compliance to this equilibrium.

The pursue-evade game as described as follows. Subjects are arranged in pairs. In each pair, one player takes the role of "pursuer", the other of "evader". The game is then played many times. In each repetition of the game, each player has two choices, left and right. If they choose differently, the evader has successfully evaded, and no money is transferred. If they choose the same, the pursuer has "found" the evader, and an amount of money is transferred from the evader to the pursuer. The amount transferred depends on where the choices coincide: if they both choose left, one unit is transferred; if they both choose right, two units are transferred.

The pay-off matrix for the pursue-evade game as just described is as follows:

| | | Evader | |
|---|---|---|---|
| | | Left | Right |
| Pursuer | Left | 1, −1 | 0, 0 |
| | Right | 0, 0 | 2, −2 |

The method for solving for a mixed-strategy Nash equilibrium will now be demonstrated. The key is to find a pair of strategies that make both players *indifferent* between the two choices. Suppose pursuer plays L with probability $p_{PL}$ and R with probability $1 - p_{PL}$, while evader plays L with probability $p_{EL}$ and R with probability $1 - p_{EL}$. Then, for pursuer, the expected values from playing L and R are (in self-explanatory notation):

$$\begin{aligned} EV_P(L) &= p_{EL} \times 1 + (1 - p_{EL}) \times 0 = p_{EL} \\ EV_P(R) &= p_{EL} \times 0 + (1 - p_{EL}) \times 2 = 2(1 - p_{EL}) \end{aligned} \tag{16.1}$$

Hence pursuer is indifferent between the two choices if:

$$p_{EL} = 2(1 - p_{EL}) \implies p_{EL} = \frac{2}{3} \tag{16.2}$$

For evader, the expected values from playing L and R are:

$$\begin{aligned} EV_E(L) &= p_{PL} \times (-1) + (1 - p_{PL}) \times 0 = -p_{PL} \\ EV_E(R) &= p_{PL} \times 0 + (1 - p_{PL}) \times (-2) = -2(1 - p_{PL}) \end{aligned} \tag{16.3}$$



Hence evader is indifferent if:

$$-p_{PL} = -2(1 - p_{PL}) \implies p_{PL} = \frac{2}{3} \tag{16.4}$$

The mixed-strategy Nash equilibrium is therefore characterised by the pair of best responses:

$$\left[\left(\frac{2}{3}L, \frac{1}{3}R\right), \left(\frac{2}{3}L, \frac{1}{3}R\right)\right] \tag{16.5}$$

Therefore, in the pursue-evade game described above, we might expect both players to randomise between L and R in such a way that in the long-run, two-thirds of their decisions are L.

### 16.2.2 Non-parametric tests on repeated game data

The data used for demonstration purposes is from Rosenthal et al. (2003). The data set is contained in the file **pursue_evade**. Although there are 40 pairs in Rosenthal et al.'s (2003) experiment, we will focus on pairs 21–34, who each play 100 repetitions of the game exactly as described in Section 16.2.1. The first few rows of the data are shown in Figure 16.1. Note that, when analysing data from experimental games such as this, the natural unit of observation is a *pair* of players in a particular period. The decision variables are "pur_L" and "eva_L", indicating respectively whether (1) or not (0) the pursuer and the evader choose Left. The variable "pay" indicates the amount transferred from the evader to the pursuer.

The first thing we might do is to check how closely each player conforms to the mixed-strategy Nash equilibrium derived in Section 16.2.1. We do this by simply observing how close their proportion of "L" choices is to the prediction of 0.67. The proportions for each player in each pair are obtained as follows:

```
. table pair, contents(mean pur_L mean eva_L)
--------------------------------------
     Pair | mean(pur_L)   mean(eva_L)
----------+---------------------------
       21 |         .43           .84
       22 |         .62           .59
       23 |         .55           .59
       24 |         .76           .78
       25 |         .59           .86
       26 |         .66           .82
       27 |         .53           .67
       28 |         .62            .7
       29 |         .67           .78
       30 |         .53           .59
       31 |         .55           .69
       32 |         .56           .69
       33 |         .42           .55
       34 |         .46           .66
--------------------------------------
```

The numbers in the table are the sample means of the choices of each player, which simply represent the proportion of times that the player chose L. We see that some of the evaders are remarkably close to the Nash prediction of 0.67. We also see that

| | pair | period | pur_L | eva_L | pay |
|---|---|---|---|---|---|
| 1 | 21 | 1 | 1 | 1 | 1 |
| 2 | 21 | 2 | 0 | 1 | 0 |
| 3 | 21 | 3 | 0 | 1 | 0 |
| 4 | 21 | 4 | 1 | 0 | 0 |
| 5 | 21 | 5 | 0 | 1 | 0 |
| 6 | 21 | 6 | 0 | 0 | 2 |
| 7 | 21 | 7 | 0 | 0 | 2 |
| 8 | 21 | 8 | 1 | 0 | 0 |
| 9 | 21 | 9 | 0 | 1 | 0 |
| 10 | 21 | 10 | 0 | 1 | 0 |
| 11 | 21 | 11 | 0 | 0 | 2 |
| 12 | 21 | 12 | 1 | 1 | 1 |
| 13 | 21 | 13 | 0 | 1 | 0 |
| 14 | 21 | 14 | 0 | 1 | 0 |
| 15 | 21 | 15 | 0 | 1 | 0 |
| 16 | 21 | 16 | 0 | 1 | 0 |
| 17 | 21 | 17 | 1 | 1 | 1 |
| 18 | 21 | 18 | 0 | 1 | 0 |

Figure 16.1: The first 18 rows of the "pursue-evade" data

pursuers are typically below the Nash prediction of 0.67 (i.e. they do not choose L frequently enough). We can obtain the overall frequencies for all pursuers and all evaders as follows:

```
. summ pur_L eva_L
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       pur_L |      1400    .5678571     .495551          0          1
       eva_L |      1400    .7007143    .4581088          0          1
```

This confirms that the evaders are closer than pursuers to the Nash prediction overall.

It is possible to perform a statistical test of the hypothesis that a particular subject is following the Nash prediction. The required test is the binomial test which was described in detail in Section 3.3. If $p$ is the true proportion of L-choices made by a subject, the null hypothesis that we require to test is $p = 0.6667$.

Since pair 31 look fairly typical, we shall use their choices to demonstrate the test:



```
. bitest pur_L=0.6667 if pair==31

    Variable |        N   Observed k   Expected k   Assumed p   Observed p
-------------+------------------------------------------------------------
       pur_L |      100        55         66.67      0.66670      0.55000

  Pr(k >= 55)            = 0.994302  (one-sided test)
  Pr(k <= 55)            = 0.009986  (one-sided test)
  Pr(k <= 55 or k >= 79) = 0.014760  (two-sided test)

. bitest eva_L=0.6667 if pair==31

    Variable |        N   Observed k   Expected k   Assumed p   Observed p
-------------+------------------------------------------------------------
       eva_L |      100        69         66.67      0.66670      0.69000

  Pr(k >= 69)            = 0.352782  (one-sided test)
  Pr(k <= 69)            = 0.723211  (one-sided test)
  Pr(k <= 64 or k >= 69) = 0.672191  (two-sided test)
```

Using the two-sided tests, we see that we have evidence ($p = 0.015$) that the pursuer's choices are not consistent with the Nash prediction. However, there is no evidence ($p = 0.672$) that the evader is departing from the Nash prediction.

Table 16.1 shows, in addition to the proportions of left-choices, the p-values from the binomial test for each player. On the evidence of this test, we see that roughly half of all players (whether pursuer or evader) have choices that are consistent with the Nash prediction.

Another theoretical prediction that can be tested is randomness in the player's sequence of choices. Clearly, any pattern in the sequence of choices could potentially be exploited by the opponent, so the optimal strategy is to produce as random a sequence as possible.

| Pair | Pur_L | Eva_L | Binomial.Pur. | Binomial.Eva. | Runs.Pur. | Runs.Eva. |
|---|---|---|---|---|---|---|
| 21 | 0.43 | 0.84 | 0.000** | 0.000** | 0.01* | 0.67 |
| 22 | 0.62 | 0.59 | 0.340 | 0.112 | 0.05 | 0.74 |
| 23 | 0.55 | 0.59 | 0.015* | 0.112 | 0.36 | 0.02* |
| 24 | 0.76 | 0.78 | 0.056 | 0.019* | 0.00** | 0.09 |
| 25 | 0.59 | 0.86 | 0.112 | 0.000** | 0.13 | 0.97 |
| 26 | 0.66 | 0.82 | 0.916 | 0.001** | 0.00** | 0.61 |
| 27 | 0.53 | 0.67 | 0.006** | 1.000 | 0.01* | 0.69 |
| 28 | 0.62 | 0.70 | 0.340 | 0.525 | 0.54 | 0.63 |
| 29 | 0.67 | 0.78 | 1.000 | 0.019* | 0.10 | 0.62 |
| 30 | 0.53 | 0.59 | 0.006** | 0.112 | 0.01* | 0.90 |
| 31 | 0.55 | 0.69 | 0.015* | 0.672 | 0.76 | 0.96 |
| 32 | 0.56 | 0.69 | 0.026* | 0.672 | 0.09 | 0.01* |
| 33 | 0.42 | 0.55 | 0.000** | 0.015* | 0.79 | 0.03* |
| 34 | 0.46 | 0.66 | 0.000** | 0.916 | 0.34 | 0.05 |

Table 16.1: Proportions of L-choices; p-values from binomial tests; p-values from runs tests

Note: * = sig ($p < 0.05$); ** = sig ($p < 0.01$).

Consider the choices of the players in pair 21:

```
Pursuer in pair 21:

1001000100010000100100101010101100010100101100100101011010111000001010 0
1000101100010011001110111110010

(62 runs)

Evader in pair 21:

1110100011011111111110111111101111011111011111011111011111101111 1101111
1101111111111111101111101111111

(29 runs)
```

In order to test whether these two sequences of 100 numbers are random, we need to consider the number of "runs", that is, the number of groups of consecutive digits that are the same.[1] The two sequences listed above contain respectively 62 and 29 runs. Note that, since the sequence is of length 100, the largest possible number of runs is 100 (with the choices continually switching between 0 and 1), and the lowest possible is 1 (with the choice never changing). A natural test of randomness would be one that asks whether the number of runs is approximately what is expected if the sequence is indeed random. If the number of runs is too high, this would imply negative serial correlation (switching too often), while too few runs would imply positive serial correlation (not switching often enough).

A test which does exactly this is the runs test. This is one non-parametric test that was not covered in Chapter 3. We shall therefore describe it briefly here. For further details the reader is referred to Siegel & Castellan (1988).

The null hypothesis implicit in this test is that the sequence of choices is random, and also that the underlying mixed strategy (i.e. the probability of L) is fixed over time. Let $m$ be the number of L-choices and $n$ be the number of R-choices, in a sequence of $N = m + n$ decisions. Let $r$ be the number of runs in the sequence. Provided both $m$ and $n$ are larger than about 20, a good approximation to the sampling distribution of $r$, under the null hypothesis of randonmess of the sequence, is:

$$r \sim N\left[\left(\frac{2mn}{N} + 1\right), \frac{2mn(2mn - N)}{N^2(N-1)}\right] \tag{16.6}$$

It follows that the null hypothesis of randonmess may be tested using the statistic:

$$z \sim \frac{r - \left(\frac{2mn}{N} + 1\right)}{\sqrt{\frac{2mn(2mn-N)}{N^2(N-1)}}} \tag{16.7}$$

Values of $z$ obtained using (16.7) are approximately standard normal under the null hypothesis.

[1] To provide an example that makes this concept quite clear: the sequence 11110001101 contains five "runs": 1111; 000; 11; 0; 1.



For the pursuer in pair 21, we have:

$$m = 43, \quad n = 57, \quad N = 100, \quad r = 62$$

$$z \sim \frac{62 - \left(\frac{2\times43\times57}{100} + 1\right)}{\sqrt{\frac{2\times43\times57(2\times43\times57-100)}{100^2(100-1)}}} = \frac{11.98}{\sqrt{23.78}} = 2.46$$

Unsurprisingly, the above test can be performed in STATA, as follows:

```
. runtest pur_L  in 1/100, t(0.5)

N(pur_L <= .5) = 57
N(pur_L >  .5) = 43
           obs = 100
       N(runs) = 62
             z = 2.46
      Prob>|z| = .01
```

The purpose of `in 1/100` is to select the appropriate rows of the data set on which to apply the test. The option `t(0.5)` specifies a threshold. This number could actually be any number strictly between 0 and 1. It is simply required in order to separate the two different values appearing in the sequence. Note that the test statistic ($z$) is returned, and it is the same as the one obtained above. Note also that a p-value is also shown. This p-value indicates that there is evidence that the sequence of choices of the pursuer in pair 21 is not random. Furthermore, the positive value of $z$ indicates that there are more runs than expected, that is, it indicates negative serial correlation. A negative value of $z$ would indicate positive serial correlation.

For the evader in pair 21, we obtain:

```
. runtest eva_L  in 1/100, t(0.5)
N(eva_L <= .5) = 16
N(eva_L >  .5) = 84
           obs = 100
       N(runs) = 29
             z = .42
      Prob>|z| = .67
```

Here, the null hypothesis is accepted, suggesting that the second sequence of numbers presented above is random.

The final two columns of Table 16.1 above provide p-values for the runs tests applied to all players. While some players appear to violate randomness, the majority appear to produce random sequences.

## 16.3 Quantal Response Equilibrium (QRE)

### *16.3.1 Theory of QRE*

The Quantal Response Equilibrium (QRE) model was developed by McKelvey & Palfrey (1995). It captures the idea that best responses are not played with certainty. It replaces the perfectly rational expectations equilibrium embodied in Nash equilibrium with an imperfect, or noisy, equilibrium. The principle of an equilibrium is maintained by assuming that players estimate expected payoffs in an unbiased way.

In Section 16.2.1, we derived the mixed strategy Nash equilibrium for the pursue-evade game, by computing expected values to players of different strategies, and then imposing indifference between strategies. To derive the QRE, we first need to apply a stochastic term to each of the expected values:

$$EV_P^*(L) = p_{EL} + \epsilon_{PL} \tag{16.8}$$
$$EV_P^*(R) = 2(1 - p_{EL}) + \epsilon_{PR} \tag{16.9}$$

$$EV_E^*(L) = -p_{PL} + \epsilon_{EL} \tag{16.10}$$
$$EV_E^*(R) = -2(1 - p_{PL}) + \epsilon_{ER} \tag{16.11}$$

For convenience, we assume that each of the stochastic terms in (16.8)–(16.11) are independently distributed with the type-I extreme value distribution with variance parameter $\mu$. The cdf is given by:

$$F(\epsilon; \mu) = \exp\left(-\exp\left(-\frac{\epsilon}{\mu}\right)\right) \tag{16.12}$$

With the distributional assumption (16.12), together with the assumption that each player selects the alternative with the higher (stochastic) expected value, it is possible to derive the following choice probabilities for the two players (for further details of this derivation, the reader is referred to Maddala, 1983):

$$p_{PL} = \frac{\exp\left(\frac{p_{EL}}{\mu}\right)}{\exp\left(\frac{p_{EL}}{\mu}\right) + \exp\left(\frac{2(1-p_{EL})}{\mu}\right)}$$
$$p_{EL} = \frac{\exp\left(-\frac{p_{PL}}{\mu}\right)}{\exp\left(-\frac{p_{PL}}{\mu}\right) + \exp\left(-\frac{2(1-p_{PL})}{\mu}\right)} \tag{16.13}$$

The key parameter in the QRE model is the "noise parameter" or "error parameter", $\mu$. From the probability formulae (16.13), the interpretation of the parameter $\mu$, as the extent of "noise" in decision-making, is clear: when $\mu$ is close to zero, the choice probabilities are close to those dictated by the mixed strategy Nash equilibrium; when $\mu$ is large, the probabilities become 0.5, meaning that players divide their choices equally between L and R.

McKelvey & Palfrey (1995) use a parameter $\lambda$, defined in such a way as to be inversely related to the level of error: $\lambda = 0$ indicates that actions consist of only error; $\lambda = \infty$ indicates that there is no error. Here, we are following the alternative parametrisation of Goeree et al. (2003), in which the parameter $\mu$ is defined as $\mu = 1/\lambda$, so $\mu = 0$ indicates no error, and $\mu = \infty$ indicates "all error".



### *16.3.2 Computing the probabilities in the QRE model*

To compute the probabilities of each player's choices, it is required to solve the above two equations (16.13) simultaneously for the two unknowns $p_{PL}$ and $p_{EL}$, for any given value of $\mu$. Since the two equations are clearly non-linear in $p_{PL}$ and $p_{EL}$, this is a numerical problem. One way to approach this is to find the values of $p_{PL}$ and $p_{EL}$ that simultaneously minimize the two quantities:

$$s_1 = \left(p_{PL} - \frac{\exp\left(\frac{p_{EL}}{\mu}\right)}{\exp\left(\frac{p_{EL}}{\mu}\right) + \exp\left(\frac{2(1-p_{EL})}{\mu}\right)}\right)^2 \tag{16.14}$$

and

$$s_2 = \left(p_{EL} - \frac{\exp\left(-\frac{p_{PL}}{\mu}\right)}{\exp\left(-\frac{p_{PL}}{\mu}\right) + \exp\left(-\frac{2(1-p_{PL})}{\mu}\right)}\right)^2 \tag{16.15}$$

Both of the quantities $s_1$ defined in (16.14) and $s_2$ defined in (16.15) have a minimum of zero, and the unique pair of values of $p_{PL}$ and $p_{EL}$ that make them both equal to zero will be the required values of $p_{PL}$ and $p_{EL}$.

There is an "optimize" command in mata that performs this sort of minimisation. The code required to implement it is as follows. Note that comment lines in MATA need to start with the marker "//" instead of "*".

```
clear mata

* START MATA FROM WITHIN STATA

mata:

// SET STARTING VALUES FOR THE TWO PROBABILITIES

start=(0.5,0.5)

// CREATE PROGRAM ("vector_min") FOR EVALUATING 2x1 VECTOR (ss)
// WHOSE ELEMENTS ARE TO BE MINIMISED

void vector_min(todo, p, ss, S, H)
    {
  external mu
                p_PL    = p[1]
                p_EL    = p[2]

  s1= (p_PL-exp(p_EL/mu)/(exp(p_EL/mu)+exp(2*(1-p_EL)/mu)))^2
  s2= (p_EL-exp((-p_PL)/mu)/(exp((-p_PL)/mu)+exp(-2*(1-p_PL)/mu)))^2

                ss = s1 \ s2
  }


// BEGIN DEFINITION OF OPTIMISATION PROBLEM,
// RETURNING S, A PROBLEM-DESCRIPTION HANDLE CONTAINING DEFAULT VALUES

 S = optimize_init()

// MODIFY DEFAULTS (THE LAST TWO SET THE LEVEL OF ACCURACY):

optimize_init_evaluator(S, &vector_min())
 optimize_init_evaluatortype(S, "gf0")
```

```
optimize_init_params(S, start)
 optimize_init_which(S,  "min" )
optimize_init_conv_ptol(S, 1e-16)
 optimize_init_conv_vtol(S, 1e-16)

// RETURN TO STATA

end

* SET VALUE OF mu

scalar mu=.7
mata: mu=st_numscalar("mu")

* PERFORM OPTIMIZATION; STORE SOLUTION IN 2x1 VECTOR p:

mata: p = optimize(S)

* EXTRACT ELEMENTS OF p

mata: st_numscalar("p_PL",p[1])
mata:  st_numscalar("p_EL", p[2])

* DISPLAY RESULT

scalar list p_PL
scalar list p_EL
```

The function to be minimised is the $2 \times 1$ vector labelled ss. The solution is the $2 \times 1$ vector $p$, from which we extract the elements $p_{PL}(\mu)$ and $p_{EL}(\mu)$. Note that we are making explicit that these probabilities are conditional on a given value of $\mu$. When this program is run (note that $\mu$ is set to 0.7 in the program) the results are:

```
. scalar list p_PL
      p_PL =  .50249676

. scalar list p_EL
      p_EL =  .66901177
```

We can run this program for a range of values of $\mu$, and plot the resulting probabilities against $\mu$. This is done in Figure 16.2. As expected both probabilities are 0.67 when $\mu \approx 0$ (the program fails when $\mu$ is set to zero, since division by zero would be required; it suffices to set $\mu$ to a small positive number). This corresponds to the mixed-strategy Nash equilibrium with zero noise. As the noise-level increases both probabilities gradually approach 0.5, as expected.

## 16.4 Estimation of the QRE Model

Having obtained the two probabilities, $p_{PL}(\mu)$ and $p_{EL}(\mu)$, by the procedure described in Section 16.3.2, it is a simple matter to construct the log-likelihood function. Let $y_{P,it} = 1$ if the proposer in pair $i$ chooses L in round $t$, and 0 otherwise. Define $y_{E,it}$ likewise for the evader in pair $i$ and round $t$. Then the log-likelihood function for the complete sample of $nT$ realisations of the game is:

$$\text{LogL}(\mu) = \sum_{i=1}^{n}\sum_{t=1}^{T} \ln\left[\left\{y_{P,it}p_{PL}(\mu) + (1 - y_{P,it})\big(1 - p_{PL}(\mu)\big)\right\}\times \left\{y_{E,it}p_{EL}(\mu) + (1 - y_{E,it})\big(1 - p_{EL}(\mu)\big)\right\}\right] \quad (16.16)$$



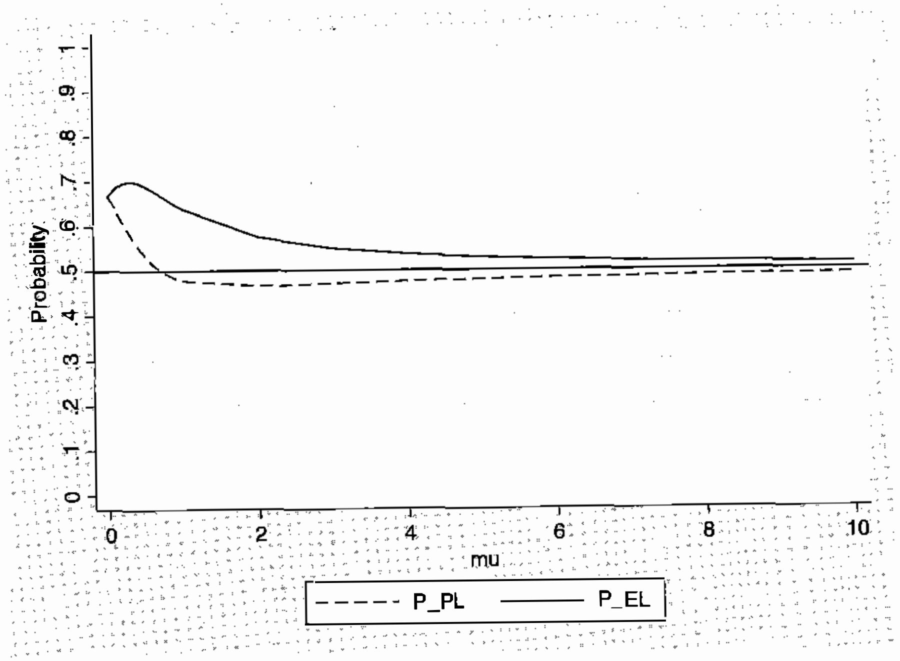


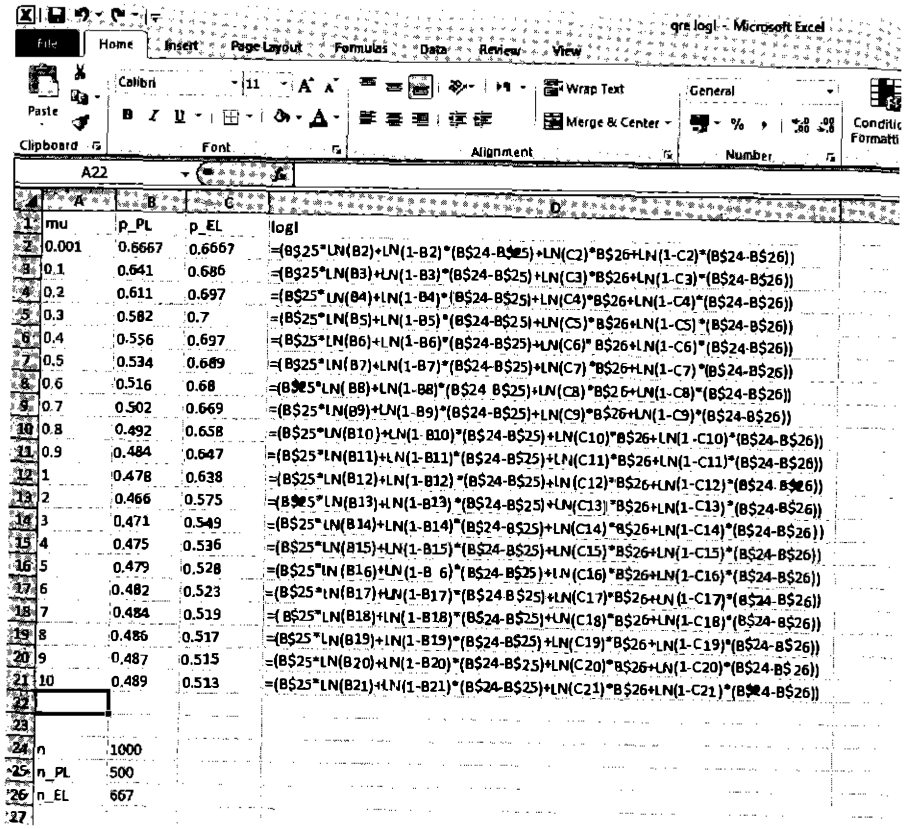

Figure 16.2: Probabilities of pursuer choosing L and evader choosing L against "noise level" ($\mu$)

In fact, it is possible to write the log-likelihood function (16.16) in a much simpler way, as a consequence of the probabilities being fixed both across pairs and across rounds of the game. Let $n_{PL}$ be the number of times, out of the total number of rounds played, $n$, that pursuers choose L, and define $n_{PR}$, $n_{EL}$ and $n_{ER}$ similarly. Then the log-likelihood function may be written as:

$$\begin{aligned}\text{LogL}(\mu) = & n_{PL}\ln p_{PL}(\mu) + n_{PR}\ln\left[1 - p_{PL}(\mu)\right] \\ & + n_{EL}\ln p_{EL}(\mu) + n_{ER}\ln\left[1 - p_{EL}(\mu)\right]\end{aligned} \quad (16.17)$$

Presenting the log-likelihood function in the form of (16.17) makes it clear that the only pieces of information that are required to estimate the parameter $\mu$ are the frequencies of the choices; no information is required on the actual sequences of choices. Accordingly, we may say that the choice frequencies are *sufficient statistics* for the estimation of $\mu$.

The log-likelihood function (16.17) may be plotted with only knowledge of the choice frequencies. The following Excel sheet is used to compute the log-likelihood over a range of values of $\mu$. The columns p_PL and p_EL contain the probabilities computed using the mata optimise routine described in Section 16.3.2. The column logl contains the formula for the log-likelihood, which is computed from the probabilities, and from the choice frequencies which are inputted at the bottom of the screen. Here, we are assuming that there are 1000 rounds in total, in which pursuers choose L 500 times, and evaders choose L 667 times. The plot of the resulting log-likelihood is shown in Figure 16.3.



| | A | B | C | D |
|---|---|---|---|---|
| 1 | mu | p_PL | p_EL | logl |
| 2 | 0.001 | 0.6667 | 0.6667 | `=(B$25*LN(B2)+LN(1-B2)*(B$24-B$25)+LN(C2)*B$26+LN(1-C2)*(B$24-B$26))` |
| 3 | 0.1 | 0.641 | 0.686 | `=(B$25*LN(B3)+LN(1-B3)*(B$24-B$25)+LN(C3)*B$26+LN(1-C3)*(B$24-B$26))` |
| 4 | 0.2 | 0.611 | 0.697 | `=(B$25*LN(B4)+LN(1-B4)*(B$24-B$25)+LN(C4)*B$26+LN(1-C4)*(B$24-B$26))` |
| 5 | 0.3 | 0.582 | 0.7 | `=(B$25*LN(B5)+LN(1-B5)*(B$24-B$25)+LN(C5)*B$26+LN(1-C5)*(B$24-B$26))` |
| 6 | 0.4 | 0.556 | 0.697 | `=(B$25*LN(B6)+LN(1-B6)*(B$24-B$25)+LN(C6)*B$26+LN(1-C6)*(B$24-B$26))` |
| 7 | 0.5 | 0.534 | 0.689 | `=(B$25*LN(B7)+LN(1-B7)*(B$24-B$25)+LN(C7)*B$26+LN(1-C7)*(B$24-B$26))` |
| 8 | 0.6 | 0.516 | 0.68 | `=(B$25*LN(B8)+LN(1-B8)*(B$24-B$25)+LN(C8)*B$26+LN(1-C8)*(B$24-B$26))` |
| 9 | 0.7 | 0.502 | 0.669 | `=(B$25*LN(B9)+LN(1-B9)*(B$24-B$25)+LN(C9)*B$26+LN(1-C9)*(B$24-B$26))` |
| 10 | 0.8 | 0.492 | 0.658 | `=(B$25*LN(B10)+LN(1-B10)*(B$24-B$25)+LN(C10)*B$26+LN(1-C10)*(B$24-B$26))` |
| 11 | 0.9 | 0.484 | 0.647 | `=(B$25*LN(B11)+LN(1-B11)*(B$24-B$25)+LN(C11)*B$26+LN(1-C11)*(B$24-B$26))` |
| 12 | 1 | 0.478 | 0.638 | `=(B$25*LN(B12)+LN(1-B12)*(B$24-B$25)+LN(C12)*B$26+LN(1-C12)*(B$24-B$26))` |
| 13 | 2 | 0.466 | 0.575 | `=(B$25*LN(B13)+LN(1-B13)*(B$24-B$25)+LN(C13)*B$26+LN(1-C13)*(B$24-B$26))` |
| 14 | 3 | 0.471 | 0.549 | `=(B$25*LN(B14)+LN(1-B14)*(B$24-B$25)+LN(C14)*B$26+LN(1-C14)*(B$24-B$26))` |
| 15 | 4 | 0.475 | 0.536 | `=(B$25*LN(B15)+LN(1-B15)*(B$24-B$25)+LN(C15)*B$26+LN(1-C15)*(B$24-B$26))` |
| 16 | 5 | 0.479 | 0.528 | `=(B$25*LN(B16)+LN(1-B16)*(B$24-B$25)+LN(C16)*B$26+LN(1-C16)*(B$24-B$26))` |
| 17 | 6 | 0.482 | 0.523 | `=(B$25*LN(B17)+LN(1-B17)*(B$24-B$25)+LN(C17)*B$26+LN(1-C17)*(B$24-B$26))` |
| 18 | 7 | 0.484 | 0.519 | `=(B$25*LN(B18)+LN(1-B18)*(B$24-B$25)+LN(C18)*B$26+LN(1-C18)*(B$24-B$26))` |
| 19 | 8 | 0.486 | 0.517 | `=(B$25*LN(B19)+LN(1-B19)*(B$24-B$25)+LN(C19)*B$26+LN(1-C19)*(B$24-B$26))` |
| 20 | 9 | 0.487 | 0.515 | `=(B$25*LN(B20)+LN(1-B20)*(B$24-B$25)+LN(C20)*B$26+LN(1-C20)*(B$24-B$26))` |
| 21 | 10 | 0.489 | 0.513 | `=(B$25*LN(B21)+LN(1-B21)*(B$24-B$25)+LN(C21)*B$26+LN(1-C21)*(B$24-B$26))` |
| 22 | | | | |
| 23 | | | | |
| 24 | n | 1000 | | |
| 25 | n_PL | 500 | | |
| 26 | n_EL | 667 | | |
| 27 | | | | |

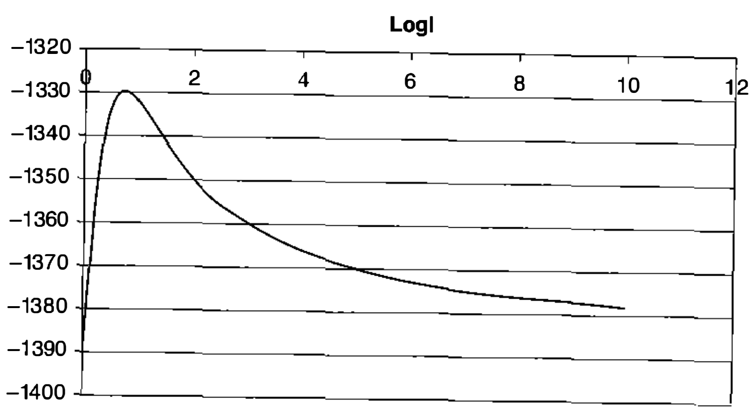



Figure 16.3: Log-likelihood function for the QRE model



We see that the log-likelihood function has only one maximum, and it is verified below that this maximum is at $\mu = 0.72$ and the maximised log-likelihood is $-1329.43$. We further note that the log-likelihood is not globally concave, and this might cause problems if the search were to stray too far from the MLE.

There is another matter relating to the shape of the log-likelihood function that users of this model should be aware of. The model requires that the data follows particular patterns; in this case, the required pattern is that the evader's frequency of choosing L is at least as large as the pursuer's. Let us consider what happens if this requirement is not met. If we reverse the frequencies assumed above, so that n_PL is 667 and n_EL is 500, the log-likelihood function has the very non-standard form shown in Figure 16.4. In this situation, technically, the MLE of $\mu$ is zero, since this is where LogL reaches a maximum. However, a more sensible conclusion is that the QRE model, in this simple form, is incapable of explaining this pattern in the data.

Of course, we can also estimate $\mu$ using the ML routine in STATA, in conjunction with the mata optimise routine described in Section 16.3.2.

The following STATA code first "simulates" 1000 plays of the game. Actually this may not qualify as a "simulation" because the data is set deterministically: the proposer's choices are set to 500L and 500R; the evader's choices are 667L and 333R. A program named "`qre`" is then defined whose purpose is to evaluate the log-likelihood function for a given value of the parameter $\mu$. Note that in order to find the two probabilities, the mata optimise routine presented above is called from within the qre program, using the current value of $\mu$. The results from the simulation are shown below the program.

Note that two different starting-vectors appear in different places in the program: "`start`" is the $(2 \times 1)$ starting vector used by the mata optimise routine when computing the two probabilities; "`sstart`" is the $(1 \times 1)$ vector of starting value(s) for the ML routine.

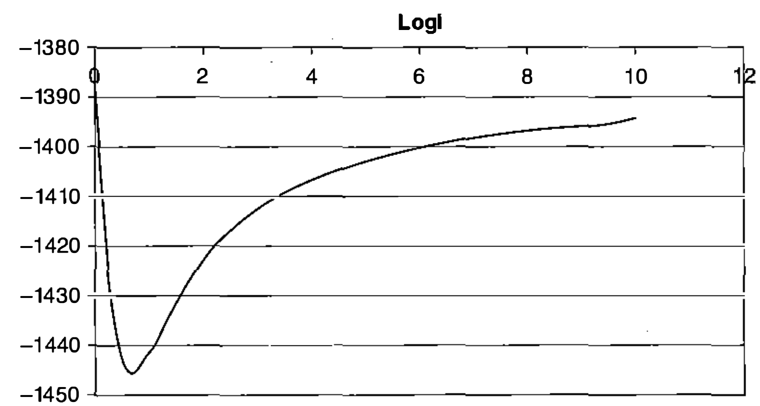



Figure 16.4: Log-likelihood function for the QRE model with an inconvenient sample

```
* HERE: SUPPLY ("vector_min") MATA PROGRAM PRESENTED ABOVE

* GENERATE DATA SET

set obs 1000
gen int y_P = _n<501
gen int y_E = _n<668

* SET STARTING VALUES FOR COMPUTATION OF THE TWO PROBABILITIES

mat start=(0.67,0.67)

prog drop _all

* LOG-LIKELIHOOD EVALUATION PROGRAM (qre) STARTS HERE:

program define qre
args lnf mu
tempvar pp
tempname p_PL p_EL

scalar mu=`mu'

* COPY STARTING PROBABILITIES AND mu INTO MATA

mata: start=st_matrix("start")
mata: mu=st_numscalar("mu")

* PERFORM OPTIMIZATION; STORE SOLUTION IN 2x1 VECTOR p:

mata: p = optimize(S)

* EXTRACT ELEMENTS OF p

mata: st_numscalar("p_PL",p[1])
mata:  st_numscalar("p_EL", p[2])

* GENERATE JOINT PROBABILITY OF PURSUER'S AND EVADER'S DECISIONS:

quietly gen double `pp'=((p_PL*y_P)+(1-p_PL)*(1-y_P))*((p_EL*y_E)+(1-p_EL)*(1-y_E))

* GENERATE LOG-LIKELIHOOD

quietly replace `lnf'=ln(`pp')
end

* SET STARTING VALUE FOR ML ROUTINE (NOTE: ONLY ONE PARAMETER, mu)

mat sstart=.6

* RUN ML

ml model lf qre ()
ml init sstart, copy
ml maximize
```

Results from this simulation are shown below:

```
. ml maximize

initial:      log likelihood = -1330.3224
rescale:      log likelihood = -1330.3224
Iteration 0:  log likelihood = -1330.3224
Iteration 1:  log likelihood = -1329.4369
Iteration 2:  log likelihood = -1329.4302
Iteration 3:  log likelihood = -1329.4302

                                              Number of obs   =       1000
                                              Wald chi2(0)    =          .
Log likelihood = -1329.4302                   Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7197914   .0974524     7.39   0.000     .5287882    .9107947
------------------------------------------------------------------------------
```



We see that the MLE of $\mu$ is 0.72 and the maximised log-likelihood is $-1329.43$. These numbers are consistent with the analysis performed on the same data using the Excel sheet above.

## 16.5 The Risk-averse QRE Model

In Section 16.2.1, we derived the mixed strategy Nash equilibrium for the pursue-evade game, by computing expected values to players of different strategies, and then imposing indifference between strategies. Note that in doing this, we were assuming risk-neutrality. In Section 16.3, we incorporated randomness into the decision process, thereby obtaining the QRE model. Goeree et al. (2003) have proposed an interesting extension to QRE which allows risk aversion instead of risk neutrality. Here, we estimate the "risk-averse QRE model", again in the context of the pursue-evade game.

Following Goeree et al. (2003) we assume a constant relative risk aversion (CRRA) utility function:

$$\begin{aligned} U(x) &= \frac{x^{1-r}}{1-r} && x \geq 0 \\ &= -\frac{(-x)^{1-r}}{1-r} && x < 0 \end{aligned} \tag{16.18}$$

As explained in Chapter 12, the parameter $r$ is the coefficient of relative risk aversion. Where in Section 16.2 we computed expected values, here we compute expected utilities. For Pursuer, the expected utilities from playing L and R are:

$$\begin{aligned} EU_P(L) &= p_{EL} \times \frac{1}{1-r} + (1-p_{EL}) \times 0 = \frac{p_{EL}}{1-r} \\ EU_P(R) &= p_{EL} \times 0 + (1-p_{EL}) \times \frac{2^{1-r}}{1-r} = \frac{2^{1-r}(1-p_{EL})}{1-r} \end{aligned} \tag{16.19}$$

For Evader, the expected utilities from playing L and R are:

$$\begin{aligned} EU_E(L) &= p_{PL} \times \frac{-1}{1-r} + (1-p_{PL}) \times 0 = \frac{p_{PL}}{1-r} \\ EU_E(R) &= p_{PL} \times 0 + (1-p_{PL}) \times \frac{-(2^{1-r})}{1-r} = -\frac{2^{1-r}(1-p_{PL})}{1-r} \end{aligned} \tag{16.20}$$

Using reasoning similar to that used in Section 16.3.2, we arrive at the conditions for equilibrium in the risk-averse QRE model:

$$p_{PL} = \frac{\exp\left(\frac{p_{EL}}{(1-r)\mu}\right)}{\exp\left(\frac{p_{EL}}{(1-r)\mu}\right) + \exp\left(\frac{2^{1-r}(1-p_{EL})}{(1-r)\mu}\right)}$$

$$p_{EL} = \frac{\exp\left(-\frac{p_{PL}}{(1-r)\mu}\right)}{\exp\left(-\frac{p_{PL}}{(1-r)\mu}\right) + \exp\left(-\frac{2^{1-r}(1-p_{PL})}{(1-r)\mu}\right)}$$

The procedure for computing the probabilities is the same as that used in Section 16.3.2, although now it requires knowledge of two parameters, $\mu$ and $r$.

The complete code, containing the mata optimise program for finding the probabilities, and also the ml program for estimating the two parameters ($\mu$ and $r$) is shown below.

```
* START MATA FROM WITHIN STATA

mata:

// SET STARTING VALUES FOR THE TWO PROBABILITIES

start=(0.5,0.5)

// CREATE PROGRAM ("vector_min") FOR EVALUATING 2x1 VECTOR (ss)
// WHOSE ELEMENTS ARE TO BE MINIMISED

void vector_min(todo, p, ss, S, H)
             {
external X, mu, r
                        PP   = p[1]
                        PE   = p[2]

EU_PL=PE/(mu*(1-r))
EU_PR=(2^(1-r))*(1-PE)/(mu*(1-r))

EU_EL=(-PP)/(mu*(1-r))
EU_ER=-(2^(1-r))*(1-PP)/(mu*(1-r))

s1=(PP-exp(EU_PL)/(exp(EU_PL)+exp(EU_PR)))^2
s2=(PE-exp(EU_EL)/(exp(EU_EL)+exp(EU_ER)))^2

                        ss= s1\s2

             }

// BEGIN DEFINITION OF OPTIMISATION PROBLEM,
// RETURNING S, A PROBLEM-DESCRIPTION HANDLE CONTAINING DEFAULT VALUES

 S = optimize_init()

// MODIFY DEFAULTS


optimize_init_evaluator(S, &vector_min())
 optimize_init_evaluatortype(S, "v0")
optimize_init_params(S, start)
 optimize_init_which(S,  "min" )
 optimize_init_tracelevel(S,"none")
 optimize_init_conv_ptol(S, 1e-16)
 optimize_init_conv_vtol(S, 1e-16)

// RETURN TO STATA

end

clear

* GENERATE DATA SET

set obs 1000
gen int y_P = _n<501
gen int y_E = _n<668

* SET STARTING VALUES FOR COMPUTATION OF THE TWO PROBABILITIES

mat start=(0.67,0.67)

prog drop _all
```



```
* LOG-LIKELIHOOD EVALUATION PROGRAM (qre_risk) STARTS HERE:

program define qre_risk
args lnf mu r
tempvar pp
tempname p1 p2 mmu rr

scalar mmu=`mu'
scalar rr=`r'

* COPY STARTING PROBABILITIES AND TWO PARAMETERS (mu and r) INTO MATA

mata: start=st_matrix("start")
mata: mu=st_numscalar("mmu")
mata: r=st_numscalar("rr")

* PERFORM OPTIMIZATION; STORE SOLUTION IN 2x1 VECTOR p:

mata: p = optimize(S)

* EXTRACT ELEMENTS OF p

mata: st_numscalar("p1",p[1])
mata:  st_numscalar("p2", p[2])

* GENERATE JOINT PROBABILITY OF PURSUER'S AND EVADER'S DECISIONS:

quietly gen double `pp'=((p1*y_P)+(1-p1)*(1-y_P))*((p2*y_E)+(1-p2)*(1-y_E))

* GENERATE LOG-LIKELIHOOD

quietly replace `lnf'=ln(`pp')
end

* SET STARTING VALUES FOR 2 PARAMETERS mu AND r

mat sstart=(.1,0.3)

* RUN ML

ml model lf qre_risk () ()
ml init sstart, copy
ml maximize
```

Once again the data has been generated with 1,000 plays, with Pursuers choosing L on 500 occasions, and Evaders choosing L on 667 occasions. The results are as follows.

```
. ml maximize

initial:       log likelihood = -1354.7992
rescale:       log likelihood = -1353.2851
rescale eq:    log likelihood =  -1329.691
Iteration 0:   log likelihood =  -1329.691
Iteration 1:   log likelihood = -1329.4408
Iteration 2:   log likelihood = -1329.4301
Iteration 3:   log likelihood =   -1329.43

                                                  Number of obs   =       1000
                                                  Wald chi2(0)    =          .
Log likelihood =   -1329.43                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
       _cons |   .7203918   .1047659     6.88   0.000     .5150544    .9257293
-------------+----------------------------------------------------------------
eq2          |
       _cons |  -.0021648   .1382919    -0.02   0.988    -.2732119    .2688824
------------------------------------------------------------------------------
```

Firstly, the MLE of the risk aversion parameter (the second estimate in the table of results) is very close to zero, suggesting risk neutrality for the assumed data pattern. Also, the MLE of $\mu(0.7204)$ is very similar to the estimate of the same parameter obtained in Section 16.4 when the risk-neutral QRE was estimated (0.7198). Moreover, the risk-neutral QRE estimates $\mu$ more precisely on the evidence of the smaller standard error. It therefore appears that the risk-averse QRE is not adding very much at all in this case. The departure from the Nash equilibrium is being adequately explained by the stochastic element.

However, let us consider a different data pattern, in which both players choose L too infrequently. Let us assume that Pursuers choose L 500 times, and Evaders choose L 600 times. In this case, the results are:

```
                                                Number of obs   =       1000
                                                Wald chi2(0)    =          .
Log likelihood = -1366.1588                     Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
       _cons |   1.054045   .2420504     4.35   0.000     .5796348    1.528455
-------------+----------------------------------------------------------------
eq2          |
       _cons |   .4150363   .1321526     3.14   0.002      .156022    .6740506
------------------------------------------------------------------------------
```

The estimate of $r$ is now $+0.415$ and is strongly significantly different from zero, reflecting strong evidence of risk aversion in this data. The estimate of $\mu$ is 1.05 and this also seems quite high. However, if the risk-neutral QRE is estimated with this data set, a much higher estimate of $\mu$, of 1.39, is obtained. The lower estimate of $\mu$ in the risk-averse QRE model is consistent with the idea that a decent proportion of the apparent "noise" in the data is being explained by the players' risk aversion.

## 16.6 QRE Applied to Contest Data

In this section we reconsider the contest experiment of Chowdhury et al. (2014) whose data was analysed in Chapter 4. In Chapter 4 panel data techniques were used to estimate the determinants of effort in contests, and in particular to determine the sources of "over-bidding". The data is contained in the file **chowdhury**.

Here we use QRE to explain departures from RNNE. This was, in fact, the approach taken by Chowdhury et al. (2014).

This example differs in an important way from the "pursue-evade" example considered in earlier parts of this chapter, because here the strategy space is not restricted to two outcomes. The effort level can take any value between 0 and 80. Assuming effort must be integer-valued, this means that the strategy space consists of 81 possible choices. This size of strategy space presents problems for the computation of the QRE since it means that 81 different probabilities need to be computed at each iteration. Figure 4.4 in Chapter 4 presented a histogram of effort levels. One clear feature of that histogram was the multi-modality, with modes tending to occur



at multiples of 5. This feature gives rise to a convenient and relatively innocuous means of reducing the size of the strategy space: round all observations to the nearest five units. Let the rounded effort level be $y$. The STATA command for generating $y$ from *bid* is:

```
gen y=5*round(bid/5,1)
```

and the resulting distribution is shown in Figure 16.5. A vertical line is included at the effort level of 15, and it will be recalled that this is the equilibrium prediction for effort in this contest.

We will now proceed to demonstrate how the QRE may be computed from this data.

Each player $i$ forms beliefs in the form of the probability distribution of other players' effort levels. Since we are treating effort as a discrete variable with 17 possible values $y^0 = 0$, $y^1 = 5$, $y^2 = 10, \ldots, y^{16} = 80$, this probability distribution is represented by the set of probabilities $p^0, p^1, \ldots, p^{16}$. On the basis of this distribution, player $i$ forms an expectation of the total effort made by the other $n-1$ players as $(n-1)\sum_{j=0}^{16} p^j y^j$. Player $i$ then computes the expected pay-off for himself from each of the 17 possible strategies:

$$EV_i(y; p^0, \ldots, p^{16}) = \frac{80y}{y + (n-1)\sum_{j=0}^{16} p^j y^j} - y \qquad y \in \{y^0, y^1, \ldots, y^{16}\} \tag{16.21}$$

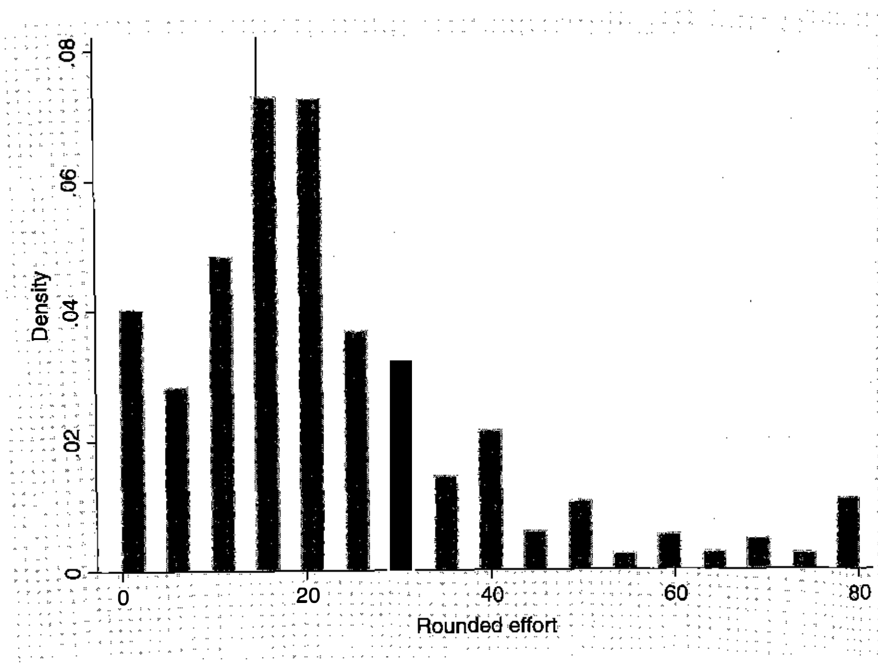



Figure 16.5: Effort data from Chowdhury et al. (2014) rounded to the nearest five units

The probability of player $i$ choosing strategy $j$ may then be expressed as:

$$p_i(y^j; p^0, \ldots, p^{16}; \mu) = \frac{\exp[EV_i(y^j)/\mu]}{\sum_{k=0}^{16} \exp[EV_i(y^k)/\mu]} \quad j = 0, \ldots, 16 \qquad (16.22)$$

where $\mu$ is the noise parameter.

The QRE (for a given value of $\mu$) is defined by the vector of probabilities, $(\tilde{p}^0(\mu), \ldots, \tilde{p}^{16}(\mu))$ say, such that:

$$p_i(y^j; \tilde{p}^0, \ldots, \tilde{p}^{16}; \mu) = \tilde{p}^j(\mu) \quad j = 0, \ldots, 16 \qquad (16.23)$$

For an i.i.d sample of $N$ effort levels, $y_1, \ldots, y_N$, we may then construct the log-likelihood function for estimating the parameter $\mu$ as:

$$LogL(\mu) = \sum_{i=1}^{N} \sum_{j=0}^{16} I(y_i = y^j) \ln[\tilde{p}^j(\mu)] \qquad (16.24)$$

Again the program for computing the QRE conisists of two parts: a mata "optimize" routine for finding the fixed point $(\tilde{p}^0(\mu), \ldots, \tilde{p}^{16}(\mu))$ for a given value of $\mu$; and a STATA "ml" program which finds the value of $\mu$ which maximises the log-likelihood function (16.24).

The code is as follows:

```
*  Maximum likelihood estimation of "mu" in QRE model for contest data.
*  MATA Optimize routine is used to find p0-p16 for given mu.
*  STATA ML routine maximises the log-likelihood over mu

clear mata

mata:

start=(0.08,0.08,0.08,0.08,0.04,0.04,0.04,0.04,0.04,0.04, ///
0.04,0.04,0.04,0.04,0.04,0.04)

void mysolver(todo, p, fff, S, H)
    {
 external mu
p1 = p[1]
p2 = p[2]
p3 = p[3]
p4 = p[4]
p5 = p[5]
p6 = p[6]
p7 = p[7]
p8 = p[8]
p9 = p[9]
p10 = p[10]
p11 = p[11]
p12 = p[12]
p13 = p[13]
p14 = p[14]
p15 = p[15]
p16 = p[16]

p0=1-p1-p2-p3-p4-p5-p6-p7-p8-p9-p10-p11-p12-p13-p14-p15-p16

ee=5*(0*p0+1*p1+2*p2+3*p3+4*p4+5*p5+6*p6+7*p7+8*p8+9*p9+10*p10  ///
 +11*p11+12*p12+13*p13+14*p14+15*p15+16*p16)
```



```
zz0=exp(0*(80/(0+3*ee)-1)/mu)
 zz1=exp(5*(80/(5+3*ee)-1)/mu)
  zz2=exp(10*(80/(10+3*ee)-1)/mu)
   zz3=exp(15*(80/(15+3*ee)-1)/mu)
zz4=exp(20*(80/(20+3*ee)-1)/mu)
 zz5=exp(25*(80/(25+3*ee)-1)/mu)
zz6=exp(30*(80/(30+3*ee)-1)/mu)
 zz7=exp(35*(80/(35+3*ee)-1)/mu)
  zz8=exp(40*(80/(40+3*ee)-1)/mu)
   zz9=exp(45*(80/(45+3*ee)-1)/mu)
 zz10=exp(50*(80/(50+3*ee)-1)/mu)

zz11=exp(55*(80/(55+3*ee)-1)/mu)
   zz12=exp(60*(80/(60+3*ee)-1)/mu)
    zz13=exp(65*(80/(65+3*ee)-1)/mu)
 zz14=exp(70*(80/(70+3*ee)-1)/mu)
  zz15=exp(75*(80/(75+3*ee)-1)/mu)
 zz16=exp(80*(80/(80+3*ee)-1)/mu)


zz=zz0+zz1+zz2+zz3+zz4+zz5+zz6+zz7+zz8+zz9+zz10+zz11 ///
+ zz12+zz13+zz14+zz15+zz16

pp0=zz0/zz
pp1=zz1/zz
pp2=zz2/zz
pp3=zz3/zz
pp4=zz4/zz
pp5=zz5/zz
pp6=zz6/zz
pp7=zz7/zz
pp8=zz8/zz
pp9=zz9/zz
pp10=zz10/zz

pp11=zz11/zz
pp12=zz12/zz
pp13=zz13/zz
pp14=zz14/zz
pp15=zz15/zz
pp16=zz16/zz



fff=(p1-pp1)^2 \
(p2-pp2)^2 \
(p3-pp3)^2 \
(p4-pp4)^2 \
(p5-pp5)^2 \
(p6-pp6)^2 \
(p7-pp7)^2 \
(p8-pp8)^2 \
(p9-pp9)^2 \
(p10-pp10)^2 \
(p11-pp11)^2 \
(p12-pp12)^2 \
(p13-pp13)^2 \
(p14-pp14)^2 \
(p15-pp15)^2 \
(p16-pp16)^2


         }

S = optimize_init()

optimize_init_evaluator(S, &mysolver())
 optimize_init_evaluatortype(S, "v0")
```

```
optimize_init_params(S, start)
 optimize_init_which(S,  "min" )
 optimize_init_tracelevel(S,"none")
 optimize_init_conv_ptol(S, 1e-16)
 optimize_init_conv_vtol(S, 1e-16)

end

clear

scalar mu=10.0

mat start=(1.101704)

prog drop _all

    mata: start=st_matrix("start")

mata: mu=st_numscalar("mu")


mata: p = optimize(S)

mata: start=p

mata: st_numscalar("p1",p[1])
mata: st_numscalar("p2",p[2])
mata: st_numscalar("p3",p[3])
mata: st_numscalar("p4",p[4])
mata: st_numscalar("p5",p[5])
mata: st_numscalar("p6",p[6])
mata: st_numscalar("p7",p[7])
mata: st_numscalar("p8",p[8])
mata: st_numscalar("p9",p[9])
mata: st_numscalar("p10",p[10])

mata: st_numscalar("p11",p[11])
mata: st_numscalar("p12",p[12])
mata: st_numscalar("p13",p[13])
mata: st_numscalar("p14",p[14])
mata: st_numscalar("p15",p[15])
mata: st_numscalar("p16",p[16])


scalar p0=1-p1-p2-p3-p4-p5-p6-p7-p8-p9-p10-p11-p12-p13-p14-p15-p16

clear

mat start=(0.08,0.08,0.08,0.08,0.04,0.04,0.04, ///
0.04,0.04,0.04,0.04,0.04,0.04,0.04,0.04,0.04)

prog drop _all

program define qre
args lnf mu
tempvar pp
tempname p0 p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 mmu

scalar mmu=`mu'

mata: start=st_matrix("start")

mata: mu=st_numscalar("mmu")
mata: X=st_numscalar("X")

mata: p = optimize(S)

mata: start=p
```



```
mata: st_numscalar("p1",p[1])
mata:  st_numscalar("p2", p[2])
mata: st_numscalar("p3",p[3])
mata:  st_numscalar("p4", p[4])
mata: st_numscalar("p5",p[5])
mata:  st_numscalar("p6", p[6])
mata:  st_numscalar("p7", p[7])
mata: st_numscalar("p8",p[8])
mata:  st_numscalar("p9", p[9])
mata:  st_numscalar("p10", p[10])
mata: st_numscalar("p11",p[11])
mata:  st_numscalar("p12", p[12])
mata: st_numscalar("p13",p[13])
mata:  st_numscalar("p14", p[14])
mata: st_numscalar("p15",p[15])
mata:  st_numscalar("p16", p[16])

scalar p0=1-p1-p2-p3-p4-p5-p6-p7-p8-p9-p10-p11-p12-p13-p14-p15-p16

forvalues k=1(1)16 {
scalar p`k'=max(p`k',1.0e-12)
}

scalar p0=1-p1-p2-p3-p4-p5-p6-p7-p8-p9-p10-p11-p12-p13-p14-p15-p16

quietly gen double `pp'=y0*p0+y1*p1+y2*p2+y3*p3+y4*p4+y5*p5+y6*p6 ///
 +y7*p7+y8*p8+y9*p9+y10*p10+y11*p11+y12*p12+y13*p13+y14*p14+y15*p15+y16*p16

quietly replace `lnf'=ln(`pp')
end

use chowdhury, clear

keep if t>15

* ROUND BID TO NEAREST 5

gen y=5*round(bid/5,1)

gen y0=y==0
gen y1=y==5
gen y2=y==10
gen y3=y==15
gen y4=y==20
gen y5=y==25
gen y6=y==30
gen y7=y==35
gen y8=y==40
gen y9=y==45
gen y10=y==50
gen y11=y==55
gen y12=y==60
gen y13=y==65
gen y14=y==70
gen y15=y==75
gen y16=y==80

mat sstart=(10)

ml model lf qre ()
ml init sstart, copy
ml maximize
```

The estimate of $\mu$ is 10.42, and the parameter is estimated quite precisely, with a 95% confidence interval $9.56 < \mu < 11.27$. The estimate is not comparable to the corresponding result in Table 4 of Chowdhury et al. (2014) because they estimate $\lambda = 1/\mu$ and also they have rescaled the outcome variable to be measured in US dollars rather than experimental points. However, the maximised log-likelihood of $-5433.8$ reported above is reasonably close to that reported in Table 4 of Chowdhury et al. (2014) which is $-5424.5$.

## 16.7 Summary and Further Reading

The chapter started by deriving a mixed-strategy Nash equilibrium for a particular two-player game. The method for doing this is explained well in Appendix A1.1 of Camerer (2003). We then tested a particular data set for adherence to the mixed-strategy Nash Equilibrium. The data set was from Rosenthal et al. (2003), and the analysis perfomed here is similar to theirs.

It was stressed in Section 16.2.2 that the runs test is a test of not only randomness but also of the constancy over time of the mixed strategy. It is conceivable that a random sequence of choices appears non-random if the underlying mixed strategy is changing over time. Ansari et al. (2012) and Shachat et al. (2015) estimate hidden Markov models in which players are assumed to switch between pure and mixed strategies. The models are applied to data from an experiment in which human subjects repeatedly play against a computer that always plays according to the mixed strategy Nash equilibrium. The estimation results lead to the conclusions that there are significant amounts of both pure and mixed strategy play, and that there are low transition probabilities between mixed and pure strategies.

The chapter progressed to an analysis of the Quantal Response Equilibrium (QRE) model which was developed by McKelvey & Palfrey (1995). QRE follows in the tradition of early models of individual choice behaviour (Luce, 1959). Many applications of QRE now appear in the literature. In this chapter, we have considered two particular examples, with, as usual, a focus on how to obtain estimates of the model's parameters. For an application of QRE to the ultimatum game, see Yi (2005).

## Exercise

Consider the "battle-of-the-sexes" game, defined by the following pay-off matrix:

| | | Player 2 | |
|---|---|---|---|
| | | Left | Right |
| Player 1 | Up | 2, 1 | 0, 0 |
| | Down | 0, 0 | 1, 2 |

Using the approach of Section 16.2.1, derive the mixed strategy equilibrium.

# Chapter 17
# Depth of Reasoning Models

## 17.1 Introduction

In the last chapter, the focus was on Quantal Response Equilibrium (QRE), a model that assumes that players depart from the Nash prediction in a random manner, but that each player forms correct beliefs about the randomness in others' decisions. In this chapter, we progress to a class of model in which it is assumed that some (or all) players act on *incorrect* beliefs concerning the strategies of others.

The essence of the approach is to assume different levels of reasoning. Each subject forms a belief about the levels of reasoning (and therefore the strategies) of other players, and adopts a strategy which is the best response to these assumed strategies. Of course, not all subjects have correct beliefs about the other subjects' strategies. All of the models developed in this chapter are built on this assumption of incorrect beliefs, and it is this assumption that makes them models of bounded rationality.

Econometrically, the models turns out to be natural applications of the finite mixture approach, of the type discussed in detail in Chapter 8. The "subject types" are simply different levels of reasoning. We also need a stochastic element that allows the actual decision to differ from the "best response" for a player of a given type. We do this in the same way as in Chapter 8.

The game we use as an application is the "$\omega$-beauty contest" game, also known as the "guessing game", developed by Nagel (1995). One reason for focusing on this game is that the data arising from it is highly amenable to mixture modelling. The feature that makes mixture modelling easy is that the outcome is continuously distributed, and tends to have the same types of easily discernible multi-modalities as seen in the mixture distributions modelled in Chapter 8.

Two types of model are considered. The first, labelled the "level-k" model, is one in which it is assumed that each player with level of reasoning greater than zero believes all other players to be one level below them. The second, labelled the "cognitive hierarchy" model, assumes, perhaps more reasonably, that players believe that other agents are distributed between lower levels of reasoning.

## 17.2 A Level-k Model for the Beauty Contest Game

Nagel's (1995) "$\omega$-Beauty Contest game" takes the following form. Each player chooses a whole number between 0 and 100. $\omega$ is a fraction whose value is chosen by the experimenter; it is usually 2/3. In the case $\omega = 2/3$, the winner is the player whose number is closest to 2/3 of the average for the entire group.

Let us imagine that the game (with $\omega = 2/3$) has been played in a large lecture theatre, with 500 players. Simulated data is contained in the file **beauty_sim**. The distribution of simulated guesses is shown in Figure 17.1. We see that the distribution is multi-modal, with one clear mode at around 33 and another around 22. Note that there is another mode close to zero. It should be said that the distribution presented in Figure 17.1 is fairly similar to those obtained from real data sets of guesses in the same game. See, for example Bosch-Domènech et al. (2010).

A popular approach to modelling behaviour in this game is the level-k model, a standard version of which is as follows. We first assume that there are a group of individuals in the population who select a number completely at random, from a (discrete) uniform distribution on $\{0, 1, \ldots, 100\}$. These individuals are labelled "level-0 reasoners". Then, there is a group who believe that all other players are level-0 reasoners, inferring that the mean guess will be around 50, and therefore that the best guess is 33 (being the closest integer to 2/3 of 50). These players are labelled "level-1 reasoners". Next, there is a group who believe that all others are level-1, with a mean guess of 33, so that the best guess for this type of individual

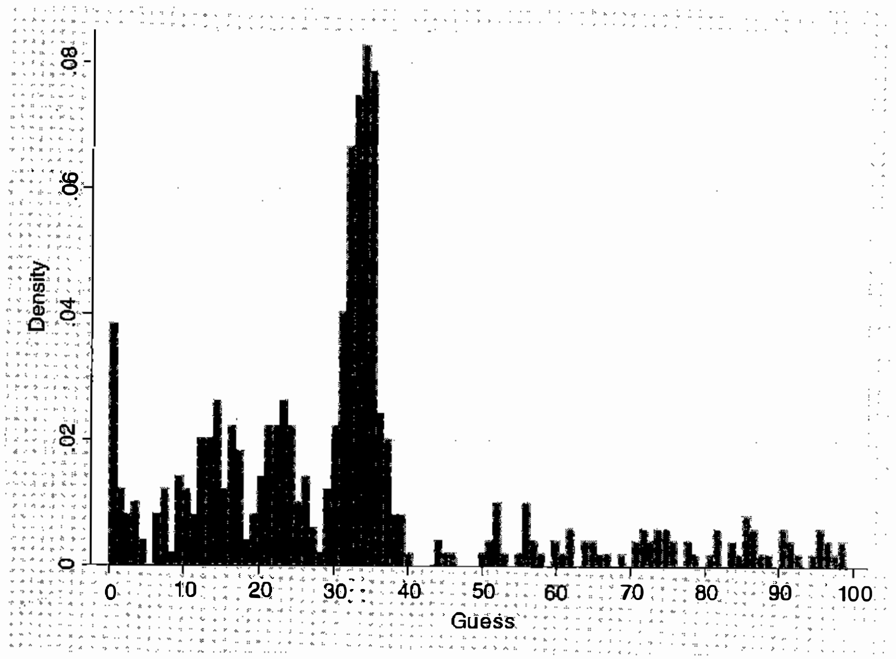



Figure 17.1: The distribution of simulated guesses for the beauty contest game



is 22; these are level-2 reasoners. This sequence continues. Level-3 reasoners will guess 15. Level-4 reasoners will guess 10, and so on.

Note that if every player had immaculate powers of reasoning, they would all supply a guess of 0, and they would all be correct and share the prize. However, needless to say, this is not what happens when the game is played with real subjects.

The estimation problem is to use the data shown in the histogram of Figure 17.1 to estimate the proportion of the population who are at each level of reasoning. We require a parametric model. We assume that there are a finite number $(J + 1)$ of types, and the maximum level of reasoning is level $J$.

In practice, some subjects do go straight to the Nash Equilibrium of zero, so it is sensible to allow for a "naïve-Nash-type" whose best guess is assumed to be zero. They are labelled as "naïve" simply because, although their behaviour corresponds to the Nash prediction, they are very unlikely to win the game. We will assign the level-$J$ label to this type.

Apart from Level-0 reasoners, who are assumed to choose from a (discrete) uniform distribution, we assume that an individual's choice is the best guess for an individual of their type, plus a random normally distributed error with mean zero and standard deviation $\sigma$. That is, we assume that if $y_j^*$ is the best guess for type $j$, then the actual guess $(y)$ will be determined by:

$$\left(y|\text{type } j\right) = y_j^* + \epsilon \qquad \epsilon \sim N\left(0, \sigma^2\right) \quad j = 1, \ldots, J \tag{17.1}$$

These assumptions give us the conditional density functions for each type:

$$f(y|L_0) = 1/100 \qquad\qquad 0 \le y \le 100$$
$$f(y|L_j) = \frac{1}{\sigma}\phi\left(\frac{y - y_j^*}{\sigma}\right) \qquad 0 \le y \le 100 \quad j = 1, \ldots, J \tag{17.2}$$

We also assume that the population is made up of the $J + 1$ types with mixing proportions $p_0, p_1, \cdots, p_J$. Combining the mixing proportions with the conditional densities (17.2) gives us the sample log-likelihood (for a sample of guesses $y_i$, $i = 1, \ldots, n$):

$$\text{LogL} = \sum_{i=1}^{n} \ln\left[\frac{p_0}{100} + \sum_{j=1}^{J} p_j \frac{1}{\sigma}\phi\left(\frac{y_i - y_j^*}{\sigma}\right)\right] \tag{17.3}$$

We set $J = 5$, and the "best guesses" are $y_1^* = 33$, $y_2^* = 22$, $y_3^* = 15$, $y_4^* = 10$, $y_5^* = 0$. Note that the best guess for level 5 is zero, because, as noted above, we assign level $J$ to "naïve-Nash" subjects. The code required to maximise the log-likelihood function is:

```
* LIKELIHOOD EVALUATION PROGRAM STARTS HERE:

program define beauty_mixture
args logl p1 p2 p3 p4 p5 sig
tempvar  f0 f1 f2 f3 f4 f5 1

quietly{
```

```
gen double `f0'=0.01
gen double `f1'=(1/`sig')*normalden((y-33.5)/`sig')
gen double `f2'=(1/`sig')*normalden((y-22.4)/`sig')
gen double `f3'=(1/`sig')*normalden((y-15.0)/`sig')
gen double `f4'=(1/`sig')*normalden((y-10.1)/`sig')
gen double `f5'=(1/`sig')*normalden((y-0)/`sig')

gen double `l'=(1-`p1'-`p2'-`p3'-`p4'-`p5')*`f0'   ///
+`p1'*`f1'+`p2'*`f2'+`p3'*`f3'+`p4'*`f4'+`p5'*`f5'

replace postp0=(1-`p1'-`p2'-`p3'-`p4'-`p5')*`f0'/`l'
replace postp1=`p1'*`f1'/`l'
replace postp2=`p2'*`f2'/`l'
replace postp3=`p3'*`f3'/`l'
replace postp4=`p4'*`f4'/`l'
replace postp5=`p5'*`f5'/`l'

replace `logl'=ln(`l')

 putmata postp0, replace
 putmata postp1, replace
 putmata postp2, replace
 putmata postp3, replace
 putmata postp4, replace
 putmata postp5, replace

}

end

* END OF LIKELIHOOD EVALUATION PROGRAM

* READ DATA

use beauty_sim, clear

hist y, disc xtitle("guess")

* INITIALISE POSTERIOR TYPE PROBABILITIES

gen postp0=. gen postp1=. gen postp2=. gen postp3=. gen postp4=. gen
postp5=.

* SET STARTING VALUES AND RUN ML

mat start=(0.3, 0.4, 0.1, 0.1,0.05, 2) ml model lf beauty_mixture
/p1 /p2 /p3 /p4 /p5 /sig ml init start, copy ml maximize

* DEDUCE PROBABILITY OF LEVEL-0 TYPE

nlcom p0:
1-_b[p1:_cons]-_b[p2:_cons]-_b[p3:_cons]-_b[p4:_cons]-_b[p5:_cons]

* EXTRACT POSTERIOR TYPE PROBABILITIES

drop postp* getmata postp0 getmata postp1 getmata postp2 getmata
postp3 getmata postp4 getmata postp5

* PLOT TYPE PROBABILITIES AGAINST GUESS

sort y

line  postp0 postp1 postp2 postp3 postp4 postp5 y , lpattern(- 1 1 1
1 1) ///
 xline(21.7) xtitle("guess") ytitle("posterior type probability") legend(off)
```



Note that the five mixing proportions $p_1, \ldots, p_5$ are estimated, and then an estimate of $p_0$ is deduced using the delta method. The results are:

```
                                                  Number of obs   =        500
                                                  Wald chi2(0)    =          .
Log likelihood = -1985.0613                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
p1           |
       _cons |   .3982665    .023804    16.73   0.000     .3516116    .4449213
-------------+----------------------------------------------------------------
p2           |
       _cons |   .1128533   .0163975     6.88   0.000     .0807148    .1449919
-------------+----------------------------------------------------------------
p3           |
       _cons |   .0898775   .0159347     5.64   0.000     .0586461     .121109
-------------+----------------------------------------------------------------
p4           |
       _cons |   .0462681   .0135852     3.41   0.001     .0196415    .0728946
-------------+----------------------------------------------------------------
p5           |
       _cons |   .0500939   .0117892     4.25   0.000     .0269876    .0732002
-------------+----------------------------------------------------------------
sig          |
       _cons |   1.929627   .1027345    18.78   0.000     1.728271    2.130982
------------------------------------------------------------------------------

. . nlcom p0:
1-_b[p1:_cons]-_b[p2:_cons]-_b[p3:_cons]-_b[p4:_cons]-_b[p5:_cons]

          p0:  1-_b[p1:_cons]-_b[p2:_cons]-_b[p3:_cons]-_b[p4:_cons]-_b[p5:_cons]

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          p0 |   .3026407    .029052    10.42   0.000     .2456999    .3595815
------------------------------------------------------------------------------
```

We see that the mixing proportions and the standard deviation of the error term are estimated to be:

$$
\begin{aligned}
\hat{p}_0 &= 0.303(0.029)\\
\hat{p}_1 &= 0.398(0.024)\\
\hat{p}_2 &= 0.113(0.016)\\
\hat{p}_3 &= 0.090(0.016)\\
\hat{p}_4 &= 0.046(0.014)\\
\hat{p}_5 &= 0.050(0.012)\\
\hat{\sigma} &= 1.930(0.103)
\end{aligned}
$$

It appears that in this (simulated) data set, around 30% of subjects are estimated to be level-0. Of the remainder, in agreement with similar studies that use real data, the majority are divided between levels 1, 2 and 3. The proportion of "naïve Nash-types" is 0.05.

We may then consider the posterior type probabilities which have been generated in the usual way using Bayes' rule. It is sensible to plot these against the subject's guess. This is done in Figure 17.2. The dashed curve represents the posterior probability of level-0. Note that this is close to 1 for any subject whose guess is greater than about 40. The other posterior probabilities peak in different positions, as expected. The curve peaking at 33 is the level-1 posterior probability; the one peaking at 22 is that for level-2; the one at 15 is for level-3; the one at 10 is for level 4. The curve peaking at zero is for level-5, the "naïve Nash-type". The position of this last curve indicates that subjects whose guess is zero or a *very small positive number* may be categorised as "naïve Nash".

It is perhaps interesting that the dashed line (probability of being level-0) appears to reach peaks in the areas between the various "best guesses". If for example, if a subject's guess is 27 or 28 (i.e. roughly half way between the best guess for levels 1 and 2) they are neither likely to be level-1 or level-2, and are instead categorised as level-0, as indicated by the peak in the dashed curve in the vicinity of this point.

Finally, it is interesting to consider what the winning guess is. The mean of the simulated sample is 32.5, implying that the winning guess is 22. The vertical line drawn in Figure 17.2 represents this winning guess. This happens to be the best guess for a level-2 type. Accordingly, we see that the probability of the winner being level-2 is around 0.86. We also see that the winner has a perhaps surprisingly high probability of around 0.14 of being level-0.

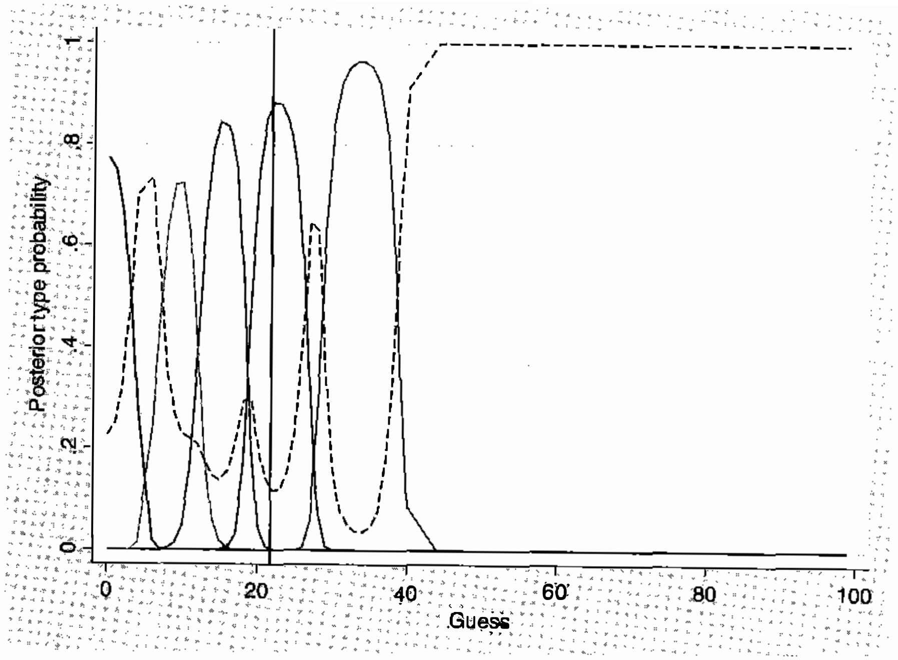



Figure 17.2: Posterior type probabilities in the level-k model. Vertical line drawn at winning guess



## 17.3 The Cognitive Hierarchy Model

A shortcoming of the version of the level-k model constructed and estimated in Section 17.2 is that it assumes that all subjects believe that all other subjects are exactly one level below themselves. An obvious reason to doubt this assumption is that if a player possesses the cognitive ability associated with level 2 or above, they are surely not naïve enough to believe that all other subjects are only one level below.

The Cognitive Hierarchy (CH) model, developed by Camerer et al. (2003, 2004), addresses this issue by assuming that the distribution of the population over reasoning levels is Poisson ($\tau$), and that a subject of type $k$ believes other members of the population to be distributed between types 0, ..., $(k-1)$ according to an upper truncated Poisson ($\tau$) distribution. That is, the (true) distribution of types over the population is given by:

$$p(j) = \left(\frac{e^{-\tau}\tau^j}{j!}\right) \qquad j = 0, 1, 2, ... \tag{17.4}$$

and a subject of type k believes that the probability distribution of other subjects between types is:

$$p_k(j) = \left(\frac{e^{-\tau}\tau^j}{j!}\right) \Bigg/ \left(\sum_{m=0}^{k-1} \frac{e^{-\tau}\tau^m}{m!}\right) \qquad j = 0, ..., k-1 \tag{17.5}$$

On the basis of the probabilities (17.5), the "best guesses" for each type may be computed recursively as:

$$\begin{aligned}
b_1 &= 0.67 \times 0.5 = 0.33 \\
b_2 &= 0.67[p_2(1) \times b_1 + p_2(0) \times 0.5] \\
b_3 &= 0.67[p_3(2) \times b_2 + p_3(1) \times b_1 + p_3(0) \times 0.5] \\
b_4 &= 0.67[p_4(3) \times b_3 + p_4(2) \times b_2 + p_4(1) \times b_1 + p_4(0) \times 0.5] \\
b_5 &= 0
\end{aligned} \tag{17.6}$$

Note that, for practical purposes, we assume reasoning levels up to 4. We also assume that type 5 is the "naïve Nash" type whose "best guess" is zero.

The log-likelihood function is then constructed by combining the observed guesses ($y$) with the type probabilities ($p(j)$) and the best guesses, $b_j$ given in (17.6):

$$\text{LogL} = \sum_{i=1}^{n} \ln\left[\frac{p(0)}{101} + \sum_{j=1}^{J} p(j)\frac{1}{\sigma}\phi\left(\frac{y_i - b_j}{\sigma}\right)\right] \tag{17.7}$$

The following code simulates data from the CH model. Note that there are only two parameters in the data generating process: the computational error parameter ($\sigma$) and the Poisson mean ($\tau$).

```
clear

set more off
set seed 9123456
set obs 500

egen i=fill(1/2)

* set "true" parameter values for simulation

scalar tau=2.0
scalar sigma=2.0

*generate the computational error variable

gen e=sigma*rnormal()

*generate the level-of-reasoning for each individual,
*setting the maximum level to 5

gen level=rpoisson(tau)
replace level=5 if level>5

*generate the first few Poisson Probabilities;
*p5 is one minus the sum of the others.

scalar p0=exp(-tau)
scalar p1=p0*tau/1
scalar p2=p1*tau/2
scalar p3=p2*tau/3
scalar p4=p3*tau/4

scalar p5=1-p0-p1-p2-p3-p4

* generate the "best guesses" for each level of reasoning;
* Note that type 5 is "naive Nash" with best-guess zero.

scalar b0=50
scalar b1=.67*b0
scalar b2=.67*(p1*b1+p0*b0)/(p1+p0)
scalar b3=.67*(p2*b2+p1*b1+p0*b0)/(p2+p1+p0)
scalar b4=.67*(p3*b3+p2*b2+p1*b1+p0*b0)/(p3+p2+p1+p0)
scalar b5=0

* generate the guesses

gen y=round((level==0)*100*uniform()+(level==1)*(b1+e)+(level==2)*(b2+e)+ ///
   (level==3)*(b3+e)+(level==4)*(b4+e)+(level==5)*abs(0+e),1)

hist y, bin(30) xtitle(guess)
```

The data resulting from the above simulation is contained in the file **cog_hier_sim**. The histogram of guesses is shown in Figure 17.3.

An interesting feature of the distribution of guesses is noted from Figure 17.3: aside from the "naïve-Nash" players who are clustered around zero, there are very few guesses of below 20. This is because, given that the Poisson mean ($\tau$) is a small number (according to Camerer et al. (2003), $\tau$ is typically estimated to be around 1.5), the vast majority of individuals are of modest reasoning-level (i.e. 3 or less). For individuals who happen to possess a higher reasoning-level, the best-guess is heavily influenced by the best-guesses of the lower level reasoners. In fact, it can be shown that, for a given value of $\tau$, as reasoning level increases, the best-guess converges towards a lower bound, which in the present case is around 20.



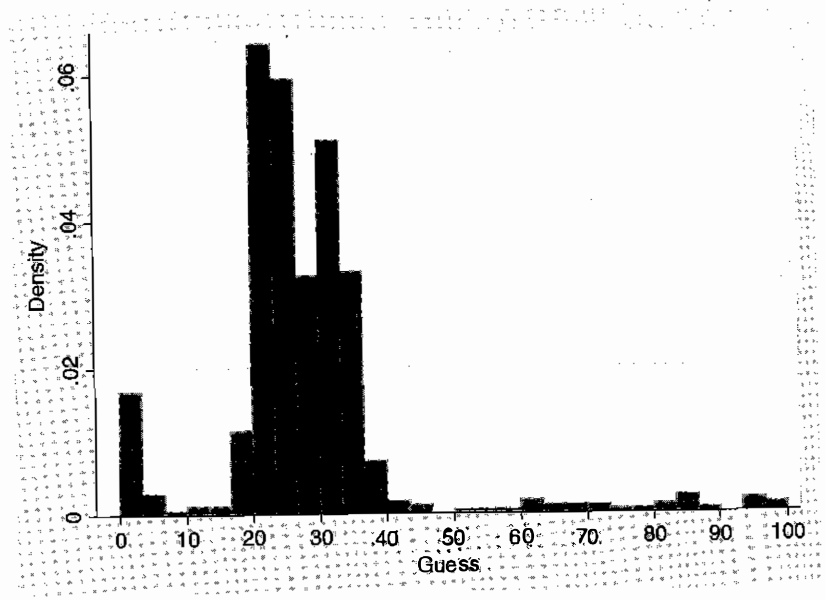



Figure 17.3: A histogram of guesses in the beauty contest game, simulated from the cognitive hierarchy model

This must be considered a shortcoming of the specification of the CH model presented here, because subjects whose guesses are less than 20 can only be classified as level-0 reasoners! This clearly calls for more sophisticated specifications of the CH model.

This last point relates closely to the issue of posterior probabilities, which are once again computed within the estimation program, to be analysed later.

The program for estimating the CH model is as follows. Note that there are only two parameters to be estimated: the computational error parameter ($\sigma$) and the Poisson mean ($\tau$).

```
program drop _all

*Log-likelihood evaluation program (ch) starts here

program define cog_heir
args logl sig tau
tempvar  f0 f1 f2 f3 f4 f5 l
tempname p0 p1 p2 p3 p4 p5 b0 b1 b2 b3 b4 b5

scalar `p0'=exp(-`tau')
scalar `p1'=`p0'*`tau'/1
scalar `p2'=`p1'*`tau'/2
scalar `p3'=`p2'*`tau'/3
scalar `p4'=`p3'*`tau'/4
scalar `p5'=1-`p0'-`p1'-`p2'-`p3'-`p4'

scalar `b0'=50
scalar `b1'=.67*`b0'
scalar `b2'=.67*(`p1'*`b1'+`p0'*`b0')/(`p1'+`p0')
scalar `b3'=.67*(`p2'*`b2'+`p1'*`b1'+`p0'*`b0')/(`p2'+`p1'+`p0')
scalar `b4'=.67*(`p3'*`b3'+`p2'*`b2'+`p1'*`b1'+`p0'*`b0')/(`p3'+`p2'+`p1'+`p0')
```

```
quietly{

gen double `f0'=0.01
gen double `f1'=(1/`sig')*normalden((y-`b1')/`sig')
gen double `f2'=(1/`sig')*normalden((y-`b2')/`sig')
gen double `f3'=(1/`sig')*normalden((y-`b3')/`sig')
gen double `f4'=(1/`sig')*normalden((y-`b4')/`sig')
gen double `f5'=(1/`sig')*normalden((y-0)/`sig')

gen double `l'=`p0'*`f0'+`p1'*`f1'+`p2'*`f2'+`p3'*`f3'+`p4'*`f4'+`p5'*`f5'

replace `logl'=ln(`l')

replace postp0=`p0'*`f0'/`l'
replace postp1=`p1'*`f1'/`l'
replace postp2=`p2'*`f2'/`l'
replace postp3=`p3'*`f3'/`l'
replace postp4=`p4'*`f4'/`l'
replace postp5=`p5'*`f5'/`l'

putmata postp0, replace
 putmata postp1, replace
 putmata postp2, replace
 putmata postp3, replace
 putmata postp4, replace
 putmata postp5, replace
}

end

* create posterior prob variables, set starting values and call ML program (ch)

gen postp0=.
gen postp1=.
gen postp2=.
gen postp3=.
gen postp4=.
gen postp5=.


mat start=( 2,2)
ml model lf cog_heir /sig /tau
ml init start, copy

ml maximize

drop postp*

getmata postp0
getmata postp1
getmata postp2
getmata postp3
getmata postp4
getmata postp5

sort y


line  postp0 postp1 postp2 postp3 postp4 postp5 y , lpattern(- 1 1 1 1 1) ///
legend(off) xlabel(0(10)100) xtitle(guess) ytitle("posterior type probability")
```

The estimation results are:

```
                                                  Number of obs   =        500
                                                  Wald chi2(0)    =          .
Log likelihood = -1746.3042                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
sig          |
       _cons |   1.998404   .1071528    18.65   0.000      1.788388    2.208419
-------------+----------------------------------------------------------------
tau          |
       _cons |   2.028311   .0503448    40.29   0.000      1.929637    2.126985
------------------------------------------------------------------------------
```



We see that both estimates are very close to the true parameter values (both 2.0), confirming the correctness of the simulation and estimation programs.

The cognitive hierarchy (CH) model is not a mixture model in the sense we are used to, because type probabilities are not being estimated. However type probabilities are still an important part of the estimation program, represented by the local scalars 'p0'–'p5'. The reason they are not estimated parameters is because the CH model assumes a particular structure in these type probabilities, namely, that they are Poisson probabilities. Hence the problem of estimating the set of type probabilities is replaced by the problem of estimating a single parameter: the Poisson mean ($\tau$).

Posterior type probabilities can still be computed, and this has been done in the usual way in the code presented above. As with the level-k model above, we plot the posterior probabilities against subjects' guesses. The dashed line is the posterior probability of level-0. Again, the probability of level-0 is close to 1 for all subjects with guesses greater than around 40. For subjects guessing between 20 and 40, the level-0 probability is very low, with one or more of levels 1–4 being suggested with high probability. Subjects guessing close to zero are once again classified as "naïve-Nash" types. For subjects with guesses around 10, there is a problem that was raised earlier: guesses in this range can only be explained by classifying the subjects as level-0 (as seen by the very high position of the dashed line in this range). This is clearly a problem because guesses of around 10 often turn out to be good guesses in the sense of having a high chance of winning the game! As previously mentioned, this simply calls for more sophisticated specifications of the CH model.

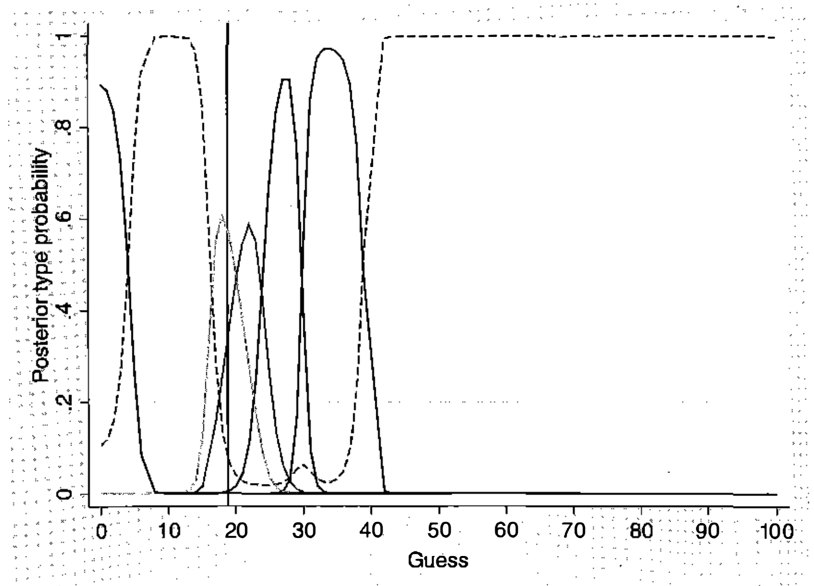



Figure 17.4: Posterior type probabilities from the cognitive hierarchy model. Vertical line drawn at winning guess

## 17.4 Summary and Further Reading

The approach taken in this chapter to constructing finite mixture models, of assuming players' decisions are distributed randomly around their best response, is very similar to the approaches taken by Bosch-Domènech et al. (2010) and Runco (2013). It should be mentioned that different approaches to this one have been followed elsewhere in the literature. For example, Stahl & Wilson (1995) assume that the best guess is always made, and that variation in guesses is explained by variation in beliefs concerning the behaviour of players at lower levels of reasoning.

Depth of reasoning models have also been applied to other types of game. See, for example, Crawford & Iriberri (2007), who applied these models to the "hide and seek" game.

## Exercise

Consider a situation in which the beauty contest game is played repeatedly by the same group of players, with feedback given (in terms of the winning guess) after each round. What would you expect to happen to the distribution of guesses over rounds? What sort of approach would you use to model the process? To answer these questions, it might be useful to consult Stahl (1996).

# Chapter 18
# Learning Models

## 18.1 Introduction

The last two chapters were concerned with the modelling of data from interactive games, and in particular how to allow for departures from equilibrium. The models considered were static, in the sense that there was no scope for strategies to change over time. In this chapter, we allow for the possibility that strategies do change over time. One obvious situation of interest is that in which players are far from equilibrium in the early rounds of a game, but, in the course of the experiment, adapt or evolve, either part of the way or all of the way, towards the equilibrium. The process which causes this evolution towards the equilibrium is a *learning* process.

Learning is defined as an observed change in behaviour owing to experience. Learning processes have already been modelled in earlier chapters. For example, in the auction models of Chapter 4, we included a variable, based on the task number, that allowed convergence towards an equilibrium (which may or may not be the Nash Equilibrium) in the course of the experiment. Similarly, in the public goods model of Section 8.5, task number was found to have a significantly negative effect on contributions to the public fund, and this was interpreted as subjects learning about the game, or, to be more specific, learning how to play Nash. In that model we also found, again using task number, that the tremble parameter, representing lapses of concentration, declined markedly in the course of the experiment. Decay in the tremble parameter was again seen in the risky choice model of Chapter 13.

In those contexts it seemed adequate to model the effect of experience in this way. In contexts of individual decision making, investigating the effect of task number on behaviour is undoubtedly the most natural way to proceed. In interactive contexts, relying on task number is appropriate provided there is reason to believe that the learning is predominantly about the task, rather than about the behaviour of other players or about the outcomes of previous tasks.

However, in some interactive contexts, for example in repeated games against a fixed opponent, learning about the behaviour of other players becomes an important driver of decisions. In each round, subjects may observe both the chosen strategy of the opponent, and the outcome in terms of their and others pay-offs. Hence they learn directly about the behaviour of others, and also about the type of strategies that are most profitable for themselves. The process of learning about other players is complicated by the fact that other players' behaviour changes as they, too, gain experience. A comprehensive model of learning should therefore incorporate the effects of a player's own past pay-offs, and also the effects of past choices by other players.

The first learning model we consider is directional learning (DL). This model is based on the idea that a players adjust their strategies in a *direction* dictated by the *outcome* of the previous play. The second model is reinforcement learning (RL), in which players alter their choice probabilities in reaction to *pay-offs* from previous plays. The third model is belief learning (BL), in which players alter their choice probabilities in response not only to the actual pay-off received, but also to the pay-offs they *would have* received by choosing each possible strategy.

The final learning model we consider is the experience-weighted attraction (EWA) learning model of Camerer & Ho (1999). EWA is a combination of RL and BL that includes the two models as special cases.

In keeping with the central objective of this book, the emphasis will be on how to estimate each of these models using STATA.

## 18.2 Directional Learning (DL)

Directional learning theory, first proposed by Selten & Stoecker (1986), is a good learning theory to start with, because it is relatively simple and highly intuitive. The key idea underlying the theory is that subjects adjust their behaviour in each period in response to the outcome of the previous period. We shall use as an example the experiment conducted by Selten & Stoecker (1986).

The experiment consists of repeated plays of the well-known prisoner's dilemma (PD) game. It is well-known that the PD has a definite game-theoretical solution that predicts non-cooperative behaviour. When the game is repeated a finite number of times (referred to as a *supergame*), the game-theoretic prediction extends to non-cooperation in every round. However, experimental behaviour does not conform to this theoretical prediction. Most experimental subjects are observed cooperating for at least part of the supergame. The typical pattern consists of what may best be described as tacit cooperation until shortly before the end of the supergame, followed by non-cooperation for the remaining rounds.

The focus of Selten & Stoecker's (1986) analysis is the timing of the "end-effect"; that is, the round in which one of the two players decides to deviate. Clearly, the player who deviates first secures a payoff advantage over their opponent. However, the player deviating first may feel that they might have enjoyed a greater total payoff by deviating *later*, since they do not know in which later round their opponent would have deviated, if at all. Similarly, the player who did not deviate first clearly realises that they would have needed to deviate in an *earlier* round than they had intended to do, if they were to have secured the payoff advantage for themselves. If the two players happen to deviate in the same round, it is likely that each would feel regret that they had not deviated one round earlier, thereby securing the advantage.

In order to test for these reactions, it is necessary to observe subjects playing a sequence of supergames, with a different opponent in every supergame. The hypothesised reactions may be tested by measuring the dependence of each subject's intended deviation period in each supergame, on the outcome of that subject's *previous* supergame. It is this dependence that captures the learning process. It is *directional learning* because it is anticipated that the decision variable (intended deviation period) changes in a *direction* dictated by the outcome of the previous supergame.



An important point is that the unit of observation in this analysis is the supergame, not each individual PD game. The dependent variable in the analysis will be an individual subject's intended deviation period in the current supergame.

Selten & Stoecker's (1986) data (extracted directly from Table B1 of their article) is contained in the file **selten-stoecker**. Thirty-five subjects participated in 25 supergames. A supergame consists of a sequence of ten prisoner's dilemma (PD) games with a fixed pair of players. Players are randomly rematched *between* supergames. A section of the data set is shown in Figure 18.1. $i$ indexes subjects, and $t$ indexes round. Missing values in the data set arise from supergames that do not follow the standard pattern (i.e. that do not contain a single continual sequence of cooperative outcomes). The variable `self` represents the intended deviation period of subject. Under normal circumstances, this is a number between 1 and 10 (10 being the number of plays in each supergame). A value of `self`= 11 indicates that subject $i$'s intended deviation period is unknown but known to be later than the opponent's. A value of `self`= 12 indicates that subject $i$ had no intention to deviate, that is, was willing to cooperate throughout the supergame. The variable `other` represents the opponent's intended deviation period, and is defined similarly to `self`.

From the variables `self` and `other` it is a simple matter to generate the three binary variables `before`, `same`, and `after`, indicating respectively whether subject $i$ deviated before, in the same round as, or after, their opponent.

Selten & Stoecker (1986) model this data using a Markov learning model, with the changes in intended deviation period detemined by transition probabilities which depend on experience in the previous period. Here, we adopt a more direct method for capturing the learning process. We simply perform a linear regression with the change in intended deviation period as the dependent variable, and the variables representing the outcome of the previous period as explanatory variables. Since the dependent variable is the *change* in intended deviation period, any individual-specific component of the intended deviation period itself is being eliminated by differencing, and hence the use of OLS regression for this model is valid.

This regression requires the use of the "`difference`" and "`lag`" operators, and for this, it is first necessary to declare the data to be panel data using the xtset command. The results, together with the two commands, are shown below:

```
. xtset i t
       panel variable:  i (strongly balanced)
        time variable:  t, 1 to 25
                delta:  1 unit

. regress d.self l.before l.same l.after, nocon

      Source |       SS       df       MS              Number of obs =     528
-------------+------------------------------           F(  3,   525) =   29.36
       Model |  93.535929     3   31.178643           Prob > F      =  0.0000
    Residual | 557.464071   525  1.06183633           R-squared     =  0.1437
-------------+------------------------------           Adj R-squared =  0.1388
       Total |        651   528  1.23295455           Root MSE      =  1.0305
```

```
------------------------------------------------------------------------------
     D.self |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
     before |
        L1. |   .3645833   .0743666     4.90   0.000     .2184906    .5106761
            |
       same |
        L1. |  -.1626506   .0799788    -2.03   0.042    -.3197683   -.0055329
            |
      after |
        L1. |  -.6117647   .0790322    -7.74   0.000     -.767023   -.4565064
------------------------------------------------------------------------------
```

The "`nocon`" option results in the regression being performed with no constant term. This gives the three coefficients a clearer interpretation. All three are significantly different from zero. The largest coefficient in magnitude is on the lag of "after"; this indicates that if the player's intended deviation period is *later* than their opponent's, they adjust their own deviation period by making it on average 0.61 periods *earlier* in the following supergame. The coefficient of "before" is of the opposite sign, and smaller in magnitude; it indicates that if the player deviates *before* their opponent, they make their deviation period on average 0.36 periods *later* in the following supergame. Finally, if the player's deviation period is the same as their opponent's, they make their deviation period on average 0.16 periods *earlier* in the following supergame, but, as expected, this effect is the smallest of the three in magnitude.

All of these results are in agreement with directional learning theory: there is strong evidence that players adjust their behaviour in response to the outcome of the previous supergame, and these adjustments are all in the expected directions.

| | i | t | self | other | before | same | after |
|---|---|---|---|---|---|---|---|
| 376 | 16 | 1 | . | . | . | . | . |
| 377 | 16 | 2 | . | . | . | . | . |
| 378 | 16 | 3 | . | . | . | . | . |
| 379 | 16 | 4 | 9 | 11 | 1 | 0 | 0 |
| 380 | 16 | 5 | 9 | 7 | 0 | 0 | 1 |
| 381 | 16 | 6 | . | . | . | . | . |
| 382 | 16 | 7 | 9 | 7 | 0 | 0 | 1 |
| 383 | 16 | 8 | 8 | 8 | 0 | 1 | 0 |
| 384 | 16 | 9 | 8 | 11 | 1 | 0 | 0 |
| 385 | 16 | 10 | 8 | 6 | 0 | 0 | 1 |
| 386 | 16 | 11 | . | . | . | . | . |
| 387 | 16 | 12 | 8 | 6 | 0 | 0 | 1 |
| 388 | 16 | 13 | 8 | 6 | 0 | 0 | 1 |
| 389 | 16 | 14 | 5 | 11 | 1 | 0 | 0 |
| 390 | 16 | 15 | 6 | 11 | 1 | 0 | 0 |
| 391 | 16 | 16 | 7 | 11 | 1 | 0 | 0 |
| 392 | 16 | 17 | 8 | 7 | 0 | 0 | 1 |
| 393 | 16 | 18 | 7 | 6 | 0 | 0 | 1 |
| 394 | 16 | 19 | 7 | 7 | 0 | 1 | 0 |
| 395 | 16 | 20 | 8 | 6 | 0 | 0 | 1 |
| 396 | 16 | 21 | 7 | 7 | 0 | 1 | 0 |
| 397 | 16 | 22 | 6 | 6 | 0 | 1 | 0 |
| 398 | 16 | 23 | 6 | 6 | 0 | 1 | 0 |
| 399 | 16 | 24 | 5 | 11 | 1 | 0 | 0 |
| 400 | 16 | 25 | 6 | 6 | 0 | 1 | 0 |

Figure 18.1: Some rows of data from Selten & Stoecker (1986)
Notes: `self` and `other` contain intended deviation period, normally between 1 and 10. Values of 11 indicate that intended deviation period is unknown but known to be later than opponent's; values of 12 indicate no intention to deviate.



## 18.3 Data Used for Estimation of RL, BL and EWA

For an application of the three learning models RL, BL and EWA, we will return to the game as used in Chapter 16, namely the pursue-evade game. Recall that the pay-off matrix was:

| | | Evader | |
|---|---|---|---|
| | | Left | Right |
| Pursuer | Left | 1, −1 | 0, 0 |
| | Right | 0, 0 | 2, −2 |

As in Chapter 16, we are representing each player's decision with a binary variable.

Estimation of all three learning models will be demonstrated using the simulated data set **pursue_evade_sim**. This data set contains simulated data sets on 100 subject pairs, each observed for 50 rounds. The simulation routine used to create this data set will be described later in the chapter.

In all of the analysis that follows throughout the remainder of the chapter, Player 1 is "pursuer" and Player 2 is "evader". Also, the binary variables representing the two players' choices ($y1$ and $y2$) take the value 0 if Left is chosen, and 1 if Right is chosen.

## 18.4 Notation Used in RL, BL and EWA

In this section we introduce the notation that is common to the three learning models, RL, BL and EWA.

To specify the game setting, we use the notation of Camerer & Ho (1999). The two Players are indexed by $i$ $(i = 1, 2)$. The game is repeated over several rounds indexed by $t$ $(t = 1, ..., T)$. In the game that we are using for illustration, there are two possible strategies for each player: $s_i^0$ and $s_i^1$, with the zero superscript indicating "left" and one indicating "right". Let $s_i(t)$ be the strategy chosen by player $i$ in round $t$, and let $s_{-i}(t)$ be the strategy chosen by the other player. Player $i$'s payoff in round $t$ is given by the scalar-valued function $\pi_i\left(s_i(t), s_{-i}(t)\right)$.

The central feature of all three learning models is a set of variables known as "attractions" which are updated following each round. $A_i^j(t)$, $j = 0, 1$ represents player $i$'s attraction to strategy $j$, following round $t$. The three models differ in the way in which the attractions are updated each round.

Players are likely to have relevant experience before the start of the game, and this experience is represented by the prior values, $A_i^j(0)$, known as "initial attractions", which are parameters to be estimated. Identification requires normalising one of the initial attractions to zero for each player.

The choice probabilities in any period are determined by the attractions in the previous period. Following Camerer & Ho (1999), we use the logistic transformation to obtain these probabilities:

$$P_i^j(t) = \frac{\exp\left(\lambda A_i^j(t-1)\right)}{\exp\left(\lambda A_i^1(t-1)\right) + \exp\left(\lambda A_i^0(t-1)\right)} \qquad i = 1,2; \quad j = 0,1. \quad t = 1,\ldots,T \tag{18.1}$$

Note that the probability formula (18.1) satisfies the basic requirement that (for example) the probability of choosing strategy 0 is monotonically increasing in the attraction of strategy 0, and monotonically decreasing in the attraction of strategy 1. The parameter $\lambda$ represents sensitivity to attractions: if $\lambda = 0$, attractions are irrelevant; if $\lambda$ is large, attractions are important.

As mentioned above, the three models RL, BL and EWA differ in the way in which the attractions are updated each round. Next, we consider each of these three models in turn, and demonstrate estimation thereof.

## 18.5 Reinforcement Learning (RL)

Reinforcement learning (Erev & Roth, 1998) is a learning theory based on the idea that players adjust their strategies in response to *payoffs* received in previous periods.

The updating rule for each attraction variable is:

$$A_i^j(t) = \phi A_i^j(t-1) + I\left(s_i(t) = s_i^j\right)\pi_i\left(s_i^j, s_{-i}(t)\right). \qquad i = 1,2; \quad j = 0,1. \quad t = 1,\ldots,T \tag{18.2}$$

where $I(.)$ is, as usual, the indicator function, taking the value one if the statement in parentheses is true, and zero otherwise. The really important feature of (18.2) is that, due to the presence of the indicator function in the final term, a player's attraction to a strategy can only increase if that strategy is chosen, and the attraction increases by the amount of the pay-off received from the chosen strategy.

The parameter $\phi$ appearing in the first term on the RHS of (18.2) is known as the "recency" parameter, and indicates the speed at which past payoffs are forgotten: $\phi = 0$ would indicate that only the most recent payoff is remembered; $\phi = 1$ would indicate that all past payoffs have equal weight in the current decision.

As mentioned in the last section, players are likely to have relevant experience before the start of the game, and this experience is represented by the prior values,



$A_i^j(0)$, known as "initial attractions", which are parameters to be estimated. For identification, the parameters $A_1^1(0)$ and $A_2^1(0)$ are normalised to zero. The other two initial attractions, $A_1^0(0)$ and $A_2^0(0)$ are free parameters.

Choice probabilities are obtained using (18.1) above. Remember that (18.1) contains a "sensitivity parameter" $\lambda$. In total, there are four parameters to be estimated in RL: $\phi$, $A_1^0(0)$, $A_2^0(0)$, and $\lambda$. In the following sub-section, we present the STATA program used to estimate these four parameters, and also show the results obtained from applying the program to the simulated data.

### 18.5.1 Program and results for RL

The (annotated) code used to estimate the RL model is as follows:

```
* LIKELIHOOD EVALUATION PROGRAM STARTS HERE

program define reinforcement

* SPECIFY ARGUMENTS: NAMES OF MAXIMAND AND 4 PARAMETERS

args logl phi lam A10_start A20_start

quietly{

* INITIALISE ATTRACTION VARIABLES FOR CURRENT LIKELIHOOD EVALUATION:

replace A10=.
replace A11=.
replace A20=.
replace A21=.

* UPDATE ATTRACTIONS BY ADDING PAY-OFFS FROM CHOSEN STRATEGIES
* Aij IS PLAYER i's ATTRACTION TO STRATEGY j,
* UPDATED BY ACTUAL PAYOFF IN CURRENT PERIOD,
* FIRST GENERATE VALUES OF ATTRACTION VARIABLES IN PERIOD 1
* (USING INITIAL ATTRACTIONS)
* THEN GENERATE VALUES OF ATTRACTION VARIABLES IN SUBSEQUENT PERIODS

by i: replace A10=`phi'*`A10_start'+wx10 if _n==1
by i: replace A11=`phi'*0+wx11 if _n==1
by i: replace A20=`phi'*`A20_start'+wx20 if _n==1
by i: replace A21=`phi'*0+wx21 if _n==1

by i: replace A11=`phi'*A11[_n-1]+wx11     if A11==.
by i: replace A10=`phi'*A10[_n-1]+wx10     if A10==.
by i: replace A21=`phi'*A21[_n-1]+wx21     if A21==.
by i: replace A20=`phi'*A20[_n-1]+wx20     if A20==.

* GENERATE PROBABILITY OF PLAYER i CHOOSING STRATEGY j (pij)
* USING _PREVIOUS_ PERIOD'S ATTRACTIONS

replace p11=.
replace p21=.

by i: replace p11=exp(`lam'*0)/(exp(`lam'*0)+exp(`lam'*`A10_start')) if _n==1
by i: replace p21=exp(`lam'*0)/(exp(`lam'*0)+exp(`lam'*`A20_start')) if _n==1

by i: replace p11=exp(`lam'*A11[_n-1])/(exp(`lam'*A11[_n-1]) ///
+exp(`lam'*A10[_n-1])) if p11==.
by i: replace p21=exp(`lam'*A21[_n-1])/(exp(`lam'*A21[_n-1]) ///
+exp(`lam'*A20[_n-1])) if p21==.
```

```
* GENERATE LOG-LIKELIHOOD CONTRIBUTION AS THE PRODUCT OF THE PROBABILITIES
* OF THE CHOICES OF THE TWO PLAYERS

quietly replace `logl'=ln((p11*y1+(1-p11)*(1-y1))*(p21*y2+(1-p21)*(1-y2)))
}
end

* LIKELIHOOD EVALUATION PROGRAM ENDS HERE

* READ DATA

use "pursue_evade_sim.dta", clear

* GENERATE AMOUNT EACH PLAYER _WOULD_ RECEIVE BY PLAYING EACH STRATEGY,
* _GIVEN_ THE STRATEGY CHOSEN BY THE OTHER PLAYER
* x_ij IS PAYOFF PLAYER i (i=1,2) WOULD RECEIVE BY PLAYING STRATEGY j
* j=0,1; 0=LEFT; 1=RIGHT).

gen int x11= 2*(y2==1)+0*(y2==0)
gen int x10= 0*(y2==1)+1*(y2==0)

gen int x21= (-2)*(y1==1)+0*(y1==0)
gen int x20= 0*(y1==1)+(-1)*(y1==0)

* GENERATE AMOUNT EACH PLAYER RECIEVES BY PLAYING THE STRATEGY THEY CHOOSE;
* ZERO FOR THE UNCHOSEN STRATEGY
* wx_ij IS AMOUNT RECEIEVED BY i CHOOSING j.  wx_ij = 0 IF j NOT CHOSEN.

gen int wx11= (y1)*x11
gen int wx10= (1-y1)*x10

gen int wx21= (y2)*x21
gen int wx20= (1-y2)*x20

* INITIALISE ATTRACTION VARIABLES, AND CHOICE PROBABILITY VARIABLES

gen double A10=.
gen double A11=.
gen double A20=.
gen double A21=.

gen double p11=.
gen double p21=.

* SET STARTING VALUES:

mat start=( 0.95,0.20,0.0,0.0)

*RUN ML

ml model lf reinforcement  /phi  /lam /A10_start /A20_start
ml init start, copy

ml max, trace search(norescale)
```

The results from running the above code are:

```
                                        Number of obs   =       5000
                                        Wald chi2(0)    =          .
Log likelihood = -6863.0929             Prob > chi2     =          .
```

| | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| phi | | | | | | |
| _cons | .7676348 | .0845556 | 9.08 | 0.000 | .601909 | .9333607 |
| lam | | | | | | |
| _cons | .1095563 | .0156976 | 6.98 | 0.000 | .0787895 | .1403231 |



| | | | | | | |
|---|---|---|---|---|---|---|
| A10_start | | | | | | |
| _cons | 1.389702 | 1.457448 | 0.95 | 0.340 | -1.466843 | 4.246247 |
| A20_start | | | | | | |
| _cons | -8.310659 | 2.286308 | -3.63 | 0.000 | -12.79174 | -3.829578 |

The recency parameter, $\phi$, is estimated to be quite a long way from 1, indicating that past pay-offs are forgotten fairly quickly. The estimate of the sensitivity parameter, $\lambda$, is positive as expected, and strongly significant, indicating that players are indeed influenced by attractions. The estimates of the two initial attractions, $A_1^0(0)$ and $A_1^2(0)$, are respectively positive and negative. This suggests that Pursuer starts with a preference for choosing "Left", while the Evader starts with a preference for "Right". Only the latter is significantly different from zero.

## 18.6 Belief learning (BL)

Belief Learning (BL), also sometimes known as "weighted fictitious play", is a learning theory based on the idea that players adjust their strategies in response to payoffs that *would* have been received under each choice. Which payoff was actually realised is not relevant.

As with RL, the central feature of BL is the concept of "attractions". However, the way in which attractions are updated is different.

It is useful to start with a very simple version of BL, and then to generalise it in steps. We start with a model in which it is assumed that players choose a best response to behaviour (of their opponent) observed in the previous period.

$$A_i^j(t) = A_i^j(t-1) + \pi_i\left(s_i^j, s_{-i}(t)\right) \quad j = 0, 1 \quad t = 1, \ldots, T \tag{18.3}$$

According to (18.3), following each round ($t$), player $i$'s attraction to strategy $j$ simply increases by the pay-off that was, *or would have been*, received, given the choice $s_{-i}(t)$ made by the other player.

The model (18.3) is sometimes called the "Cournot learning model". It has only a small number of free parameters: the initial attractions $A_1^0(0)$ and $A_2^0(0)$ (the other two, $A_1^1(0)$ and $A_2^1(0)$ are, as in RL, normalised to zero); and the sensitivity parameter $\lambda$ appearing in the probability formula (18.1).

It is of course likely that players take account of more than just the behaviour of the other player in the previous round. We next consider a model in which it is assumed that players base their expectations of the others player's behaviour on the other player's behaviour in *all* previous rounds, with equal importance attached to each previous round. The model is known as the "fictitious play model". Assuming the initial attractions to be zero, the model can be presented as follows. Attractions following the first three rounds are given by:

$$A_i^j(1) = \pi_i\left(s_i^j, s_{-i}(1)\right) \quad j = 0, 1$$

$$
\begin{aligned}
A_i^j(2) &= \frac{\pi_i\left(s_i^j, s_{-i}(1)\right) + \pi_i\left(s_i^j, s_{-i}(2)\right)}{2} \qquad j = 0, 1 \\
A_i^j(3) &= \frac{\pi_i\left(s_i^j, s_{-i}(1)\right) + \pi_i\left(s_i^j, s_{-i}(2)\right) + \pi_i\left(s_i^j, s_{-i}(3)\right)}{3} \qquad j = 0, 1 \\
&\vdots
\end{aligned}
\tag{18.4}
$$

According to (18.4), the attraction of strategy $j$ is simply the average pay-off in all rounds up to the current round that would have resulted from choosing strategy $j$, given the observed behaviour of the other player in each round.

An alternative way of presenting model (18.4) is by introducing an "experience" variable $N(t)$, as follows.

$$
\begin{aligned}
N(0) &= 0 \\
N(t) &= N(t-1) + 1 \quad t = 1, \ldots T \\
A_i^j(0) &= 0 \quad j = 0, 1 \\
A_i^j(t) &= \frac{N(t-1)A_i^j(t-1) + \pi_i\left(s_i^j, s_{-i}(t)\right)}{N(t)} \qquad j = 0, 1 \quad t = 1, \ldots, T
\end{aligned}
\tag{18.5}
$$

In (18.5), the experience variable $N(t)$ is simply the round number $t$. However, the experience variable becomes more important in the next generalisation of the model: the "weighted fictitious play model". Here, it is assumed that in forming expectations of the other player's behaviour, the experience of recent rounds carries more weight than that of rounds further back in the past. For this purpose, as in RL, we introduce a "recency parameter", $\phi (0 \leq \phi \leq 1)$. Observations $t$ periods back will be weighted $\phi^{(t-1)}$. Introducing $\phi$ to (18.4), we obtain:

$$
\begin{aligned}
A_i^j(1) &= \pi_i\left(s_i^j, s_{-i}(1)\right) \quad j = 0, 1 \\
A_i^j(2) &= \frac{\phi\pi_i\left(s_i^j, s_{-i}(1)\right) + \pi_i\left(s_i^j, s_{-i}(2)\right)}{\phi + 1} \qquad j = 0, 1 \\
A_i^j(3) &= \frac{\phi^2\pi_i\left(s_i^j, s_{-i}(1)\right) + \phi\pi_i\left(s_i^j, s_{-i}(2)\right) + \pi_i\left(s_i^j, s_{-i}(3)\right)}{\phi^2 + \phi + 1} \qquad j = 0, 1 \\
&\vdots
\end{aligned}
\tag{18.6}
$$

Equation (18.6) makes it clear that attractions are weighted averages of current and previous (hypothetical) pay-offs. We can also introduce $\phi$ to (18.5), as follows:

$$
\begin{aligned}
N(0) &= 0 \\
N(t) &= \phi N(t-1) + 1 \quad t = 1, \ldots T \\
A_i^j(0) &= 0 \quad j = 0, 1 \\
A_i^j(t) &= \frac{\phi N(t-1)A_i^j(t-1) + \pi_i\left(s_i^j, s_{-i}(t)\right)}{N(t)} \qquad j = 0, 1 \quad t = 1, \ldots, T
\end{aligned}
\tag{18.7}
$$



It can easily be verified that (18.7) is equivalent to (18.6). (18.7) is useful because it shows how attractions are updated following each round.

The variable $N(t)$, whose updating rule is given in the second line of (18.7), is a measure of the amount of past experience accumulated at round $t$, measured in "observation equivalents".

For ease of exposition, we have been assuming that the initial attractions $A_i^j(0)$ $j = 1, 2$ and initial experience $N(0)$ are all zero. Actually, it is possible for these to be free parameters. Hence the version of Belief Learning (BL) that we estimate below will be the weighted fictitious play model (18.7), with $A_1^0(0)$, $A_2^0(0)$ and $N(0)$ all as free parameters (and with the other two initial attractions, $A_1^1(0)$ and $A_2^1(0)$, normalised to zero).

As in RL, the recency parameter $\phi$ is the key parameter. Note that if $\phi = 1$ weighted fictitious play becomes standard fictitious play (18.5), because there is no down-weighting of past observations. Note also that if $\phi = 0$, the model becomes the Cournot learning model (18.3).

The complete list of free parameters in our BL model is: $\phi$, $A_1^0(0)$, $A_2^0(0)$, $N(0)$ and $\lambda$. In the next sub-section, we present the STATA code for estimating these parameters, and the estimation results.

### 18.6.1 Program and results for BL

The (annotated) STATA code for estimating the BL model (18.7) is as follows:

```
* LIKELIHOOD EVALUATION PROGRAM STARTS HERE

program define belief
args logl phi lam A10_start A20_start N_start


quietly{

replace A10=.
replace A11=.
replace A20=.
replace A21=.

replace N=.

by i: replace N='phi'*'N_start'+1 if _n==1
by i: replace N='phi'*N[_n-1]+1 if N==.

by i: replace A10=('phi'*'N_start'*'A10_start'+x10)/N if _n==1
by i: replace A11=('phi'*'N_start'*0+x11)/N if _n==1
by i: replace A20=('phi'*'N_start'*'A20_start'+x20)/N if _n==1
by i: replace A21=('phi'*'N_start'*0+x21)/N if _n==1

* Aij is the attraction, updated by payoff (either actual or hypothetical) in t,
* to be used to determine choice probs in t+1

by i: replace A11=('phi'*N[_n-1]*A11[_n-1]+x11)/N    if A11==.
by i: replace A10=('phi'*N[_n-1]*A10[_n-1]+x10)/N    if A10==.
by i: replace A21=('phi'*N[_n-1]*A21[_n-1]+x21)/N    if A21==.
by i: replace A20=('phi'*N[_n-1]*A20[_n-1]+x20)/N    if A20==.

*pij are the probabilities player i choosing strategy j

replace p11=.
```

```
replace p21=.

by i: replace p11=exp('lam'*0)/(exp('lam'*0)+exp('lam'*'A10_start')) if _n==1
by i: replace p21=exp('lam'*0)/(exp('lam'*0)+exp('lam'*'A20_start')) if _n==1

by i: replace p11=exp('lam'*A11[_n-1])/(exp('lam'*A11[_n-1]) ///
+exp('lam'*A10[_n-1])) if p11==.
by i: replace p21=exp('lam'*A21[_n-1])/(exp('lam'*A21[_n-1]) ///
+exp('lam'*A20[_n-1])) if p21==.

replace 'logl'=ln((p11*y1+(1-p11)*(1-y1))*(p21*y2+(1-p21)*(1-y2)))
}

end

* LIKELIHOOD EVALUATION PROGRAM ENDS HERE

* READ DATA

use "pursue_evade_sim.dta", clear

* GENERATE AMOUNT EACH PLAYER _WOULD_ RECEIVE BY PLAYING EACH STRATEGY,
* _GIVEN_ THE STRATEGY CHOSEN BY THE OTHER PLAYER
* x_ij IS PAYOFF PLAYER i (i=1,2) RECEIVES BY PLAYING STRATEGY j
* (j=0,1; 0=LEFT; 1=RIGHT).

gen int x11= 2*(y2==1)+0*(y2==0)
gen int x10= 0*(y2==1)+1*(y2==0)
gen int x21= (-2)*(y1==1)+0*(y1==0)
gen int x20= 0*(y1==1)+(-1)*(y1==0)

* INITIALISE OTHER VARIABLES

gen double A10=.
gen double A11=.
gen double A20=.
gen double A21=.

gen double N=.

gen double wx11=.
gen double wx10=.
gen double wx21=.
gen double wx20=.

gen double p11=.
gen double p21=.

* STARTING VALUES:

mat start=( 0.95,0.20,0.0,0.0,1.0)

*RUNNING ML

ml model lf belief /phi /lambda /A10_start /A20_start /N_start
ml init start, copy
ml max, trace search(norescale)
```

The results from running the above code are:

```
                                                Number of obs   =       5000
                                                Wald chi2(0)    =          .
Log likelihood = -6808.3011                     Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
phi          |
       _cons |   .9531451   .0188062    50.68   0.000     .9162856    .9900046
-------------+----------------------------------------------------------------
lambda       |
       _cons |    .424634   .0355367    11.95   0.000     .3549833    .4942846
-------------+----------------------------------------------------------------
A10_start    |
       _cons |   1.684257   .4686631     3.59   0.000     .7656939    2.602819
-------------+----------------------------------------------------------------
A20_start    |
       _cons |  -4.255275   .7702032    -5.52   0.000    -5.764845   -2.745704
-------------+----------------------------------------------------------------
N_start      |
       _cons |   .4723948   .1498293     3.15   0.002     .1787348    .7660549
------------------------------------------------------------------------------
```



First of all, the estimate of the parameter $\phi$ is clearly greater than zero, amounting to a strong rejection of the Cournot learning model (18.3). The estimate of $\phi$ is in fact fairly close to 1, implying that there is little discounting of previous experience. However, since the 95% confidence interval for $\phi$ does not include 1, we are unable to accept the "standard fictitious play model" (18.5). The estimate of $\lambda$ is again positive and significant, and much larger in magnitude than the estimate of the same parameter in the RL model, but this is simply a consequence of the attraction variables having scales that differ between the two models. The estimates of the two initial attractions, $A_1^0(0)$ and $A_1^2(0)$, are respectively positive and negative, and this time they are both significant. This model us telling us that Pursuer starts with a clear preference for "Left", while the Evader starts with a clear preference for "Right". Finally, the estimate of $N(0)$ is positive and significant although small in magnitude: it tells us that players start the experiment with experience of around 0.5 "observation equivalents".

## 18.7 The Experience Weighted Attraction (EWA) Model

### *18.7.1 Introduction to EWA*

As mentioned in the introduction to the chapter, the experience weighted attraction (EWA) model combines reinforcement learning (RL) and belief learning (BL).

Under EWA, the updating rules for the two variables (experience and attraction) are:

$$N(t) = \rho N(t-1) + 1, \qquad t \geq 1 \tag{18.8}$$

$$A_i^j(t) = \frac{\phi N(t-1) A_i^j(t-1) + \left[\delta + (1-\delta) I(s_i(t) = s_i^j)\right] \pi_i(s_i^j, s_{-i}(t))}{N(t)} \tag{18.9}$$

The parameter $\rho$ in (18.8) represents the depreciation rate of past experience, and is expected to be slightly less than one. Equation (18.9), showing how attractions are updated, is more complicated. The second term of the numerator is either the pay-off received from the chosen strategy, or the pay-off that *would* have been received from the unchosen strategy weighted by $\delta$. $\delta$ is therefore the parameter which tells us whether we have RL or BL: $\delta = 0$ under RL; $\delta = 1$ under BL. And of course, if we find that $\delta$ is a fraction, its value tells us which of the two models is closer to the truth.

As with RL and BL, the choice probabilities in any period are determined by the attractions in the previous period, via the logistic transformation (18.1).

There are a total of seven parameters in EWA:

$$\rho, \delta, \phi, \lambda, A_1^1(0), A_2^1(0), N(0)$$

Restrictions on these parameters give to the other models of interest. As demonstrated by Camerer & Ho (1999), the required restrictions are:

| | |
|---|---|
| RL: | $\delta = 0$; $N(0) = 1$; $\rho = 0$ |
| BL: | $\delta = 1$; $\rho = \phi$ |

A minor point is that the reason why the restriction $N(0) = 1$ is used to give RL, rather than the perhaps more intuitive $N(0) = 0$, is that the latter restriction would cause the first term in the numerator of (18.9) to be zero in the first period, and as a consequence the initial attraction parameters would not be identified. Another point is that under BL, the depreciation parameter $\rho$ and the recency parameter $\phi$ are the same parameter, hence the restriction $\rho = \phi$. Under RL, the depreciation parameter $\rho$ is not present.

### 18.7.2 Simulation of data set using EWA

The following (annotated) STATA code simulates data from the EWA model applied to the pursue-evade game (100 subject pairs each playing 50 rounds). This is the simulated data set used in previous sections of the chapter, and also later in this section.

The parameter values used in the simulation are as follows:

| Parameter | True value |
|---|---|
| $\rho$ | 0.97 |
| $\delta$ | 0.60 |
| $\phi$ | 0.94 |
| $\lambda$ | 0.80 |
| $A_1^U(0)$ | 1.0 |
| $A_2^U(0)$ | −2.0 |
| $N(0)$ | 1.0 |

An important feature of the simulation is the loop over repetitions of the game (i.e. the part of the program starting with the "`forvalues`" command). Before the loop begins, choices and resulting attractions are simulated for period 1 only. Then within each iteration of the loop, the following sequence of tasks is performed in order to simulate the remaining periods:



1. Compute choice probabilities p11 and p21 (i.e. the two players' R-probabilities; using the attractions from the previous period).
2. Generate choices from the probabilities, using random uniform variables.
3. Compute the pay-offs that would result from different choices.
4. Compute the weighted pay-offs, i.e. the pay-offs weighted according to whether they were actually realised.
5. Compute the attractions, for determining the choice probabilities in the following repetition of the loop.

```
/* SIMULATION OF EWA MODEL
n=100 subject pairs, T=50 rounds.
PLAYER 1 = PURSUER; PLAYER 2= EVADER
0 = LEFT; 1 = RIGHT
*/

clear
drop _all
set obs 5000
set seed 56734512
set more off

* SET TRUE PARAMETER VALUES:

scalar rho=0.97
scalar delta=0.60
scalar phi=0.94
scalar lam=0.80
scalar A10_start=1.0
scalar A20_start=-2.0
scalar N_start=1.0

* GENERATE PAIR NUMBER (i), PERIOD NUMBER (t), AND DECLARE PANEL:

egen int i=seq(), f(1)  b(50)
egen int t=seq(), f(1) t(50)
tsset i t

* GENERATE TWO RANDOM UNIFORMS FOR LATER USE:

gen double u1=runiform()
gen double u2=runiform()

* GENERATE PERIOD 1 PROBABILITIES:

by i: generate  double p11=exp(lam*0)/(exp(lam*0)+exp(lam*A10_start)) if _n==1
by i: generate  double p21=exp(lam*0)/(exp(lam*0)+exp(lam*A20_start)) if _n==1

* GENERATE PERIOD-1 CHOICES OF PLAYERS 1 AND 2 (USING RANDOM UNIFORMS):

by i: gen int y1=u1<p11 if _n==1
by i: gen int y2=u2<p21 if _n==1

* GENERATE PERIOD-1 PAY-OFFS:

by i: generate double x11= 2*(y2==1)+0*(y2==0)   if _n==1
by i: generate double x10= 0*(y2==1)+1*(y2==0)   if _n==1
by i: generate double x21= (-2)*(y1==1)+0*(y1==0)   if _n==1
by i: generate double x20= 0*(y1==1)+(-1)*(y1==0)   if _n==1

* GENERATE PERIOD-1 WEIGHTED PAY-OFFS:

by i: generate  double wx11= (delta+(1-delta)*(y1==1))*x11 if _n==1
by i: generate  double wx10= (delta+(1-delta)*(y1==0))*x10 if _n==1
by i: generate  double wx21= (delta+(1-delta)*(y2==1))*x21 if _n==1
by i: generate  double wx20= (delta+(1-delta)*(y2==0))*x20 if _n==1

* GENERATE EXPERIENCE VARIABLE, N(t), STARTING WITH PERIOD 1:

by i: generate  double N=rho*N_start+1 if _n==1
```

```
by i: replace   N=rho*N[_n-1]+1 if N==.

* GENERATE PERIOD-1 ATTRACTIONS:

by i: generate A11=(phi*N_start*0+wx11)/N    if _n==1
by i: generate A10=(phi*N_start*A10_start+wx10)/N    if _n==1
by i: generate A21=(phi*N_start*0+wx21)/N    if _n==1
by i: generate A20=(phi*N_start*A20_start+wx20)/N    if _n==1

quietly{

* LOOP OVER PERIODS STARTS HERE

 forvalues t = 2(1)50 {

* GENERATE p11 AND p21 (PROBABILITIES OF PLAYERS 1 and 2 CHOOSING STRATEGY 1):

by i: replace p11=exp(lam*A11[_n-1])/(exp(lam*A11[_n-1])+exp(lam*A10[_n-1])) ///
  if (_n==`t')
by i: replace p21=exp(lam*A21[_n-1])/(exp(lam*A21[_n-1])+exp(lam*A20[_n-1])) ///
  if (_n==`t')

* GENERATE y1 AND y2 (CHOICES OF PLAYERS 1 AND 2) USING RANDOM UNIFORMS:

by i: replace y1=0 if (_n==`t')
by i: replace y1= (u1<p11)  if (_n==`t')

by i: replace y2=0 if (_n==`t')
by i: replace y2= (u2<p21)  if (_n==`t')

* GENERATE xij (PAY-OFF PLAYER i WOULD HAVE RECEIVED WITH STRATEGY j):

by i: replace x11= 2*(y2==1)+0*(y2==0)   if (_n==`t')
by i: replace x10= 0*(y2==1)+1*(y2==0)   if (_n==`t')
by i: replace x21= (-2)*(y1==1)+0*(y1==0)   if (_n==`t')
by i: replace x20= 0*(y1==1)+(-1)*(y1==0)   if (_n==`t')

* GENERATE wxij (PAY-OFFS WEIGHTED BY DELTA PARAMETER):

by i: replace wx11= (delta+(1-delta)*(y1==1))*x11  if (_n==`t')
by i: replace wx10= (delta+(1-delta)*(y1==0))*x10  if (_n==`t')
by i: replace wx21= (delta+(1-delta)*(y2==1))*x21  if (_n==`t')
by i: replace wx20= (delta+(1-delta)*(y2==0))*x20  if (_n==`t')

* GENERATE Aij (ATTRACTION, UPDATED BY PAY-OFFS IN t, TO BE USED
* TO DETERMINE CHOICE PROBABILITIES in t+1)

by i: replace A11=(phi*N[_n-1]*A11[_n-1]+wx11)/N    if (_n==`t')
by i: replace A10=(phi*N[_n-1]*A10[_n-1]+wx10)/N    if (_n==`t')
by i: replace A21=(phi*N[_n-1]*A21[_n-1]+wx21)/N    if (_n==`t')
by i: replace A20=(phi*N[_n-1]*A20[_n-1]+wx20)/N    if (_n==`t')

}
* END OF LOOP
}

* DISCARD SUPERFLUOUS VARIABLES:

keep i t y1 y2
```

### 18.7.3 Estimation of EWA Model

The estimation program is similar to those used in earlier sections for RL and BL. However it is more complicated because there are more parameters to estimate. In the following code, we first estimate the full EWA model. We then estimate restricted versions that correspond to the RL and BL models estimated earlier. Finally, we conduct LR tests of RL and BL as restricted versions of EWA.



```
* LIKELIHOOD EVALUATION PROGRAM STARTS HERE

program drop _all

program define ewa
args lnf rho delta phi lambda A10_start A20_start N_start

tempvar
tempname

quietly{

replace A10=.
replace A11=.
replace A20=.
replace A21=.

* GENERATE EXPERIENCE VARIABLE, N(t), STARTING WITH PERIOD 1:

replace N=.
by i: replace N=`rho'*`N_start'+1 if _n==1
by i: replace N=`rho'*N[_n-1]+1 if N==.

* GENERATE wxij (PAY-OFFS WEIGHTED BY DELTA PARAMETER):

replace wx11= (`delta'+(1-`delta')*(y1))*x11
replace wx10= (`delta'+(1-`delta')*(1-y1))*x10
replace wx21= (`delta'+(1-`delta')*(y2))*x21
replace wx20= (`delta'+(1-`delta')*(1-y2))*x20

* GENERATE PERIOD-1 ATTRACTIONS:

by i: replace A10=(`phi'*`N_start'*`A10_start'+wx10)/N if _n==1
by i: replace A11=(`phi'*`N_start'*0+wx11)/N if _n==1
by i: replace A20=(`phi'*`N_start'*`A20_start'+wx20)/N if _n==1
by i: replace A21=(`phi'*`N_start'*0+wx21)/N if _n==1

* GENERATE ATTRACTIONS FOR t>1:

by i: replace A11=(`phi'*N[_n-1]*A11[_n-1]+wx11)/N    if A11==.
by i: replace A10=(`phi'*N[_n-1]*A10[_n-1]+wx10)/N    if A10==.
by i: replace A21=(`phi'*N[_n-1]*A21[_n-1]+wx21)/N    if A21==.
by i: replace A20=(`phi'*N[_n-1]*A20[_n-1]+wx20)/N    if A20==.

* GENERATE p11 AND p21 (PROBABILITIES OF PLAYERS 1 and 2 CHOOSING STRATEGY 1):

replace p11=.
replace p21=.

by i: replace p11=exp(`lambda'*0)/(exp(`lambda'*0) ///
+exp(`lambda'*`A10_start')) if _n==1
by i: replace p21=exp(`lambda'*0)/(exp(`lambda'*0) ///
+exp(`lambda'*`A20_start')) if _n==1

by i: replace p11=exp(`lambda'*A11[_n-1])/(exp(`lambda'*A11[_n-1]) ///
+exp(`lambda'*A10[_n-1])) if p11==.
by i: replace p21=exp(`lambda'*A21[_n-1])/(exp(`lambda'*A21[_n-1]) ///
+exp(`lambda'*A20[_n-1])) if p21==.

* GENERATE LOG-LIKELIHOOD

replace `lnf'=ln((p11*y1+(1-p11)*(1-y1))*(p21*y2+(1-p21)*(1-y2)))
}
end

* LIKELIHOOD EVALUATION PROGRAM ENDS HERE

* READ DATA

use "pursue_evade_sim.dta", clear

* GENERATE PAY-OFFS (xij) THAT PLAYER i WOULD RECIEVE FROM CHOOSING STRATEGY j.

gen int x11= 2*(y2==1)+0*(y2==0)
```

```
gen int x10= 0*(y2==1)+1*(y2==0)

gen int x21= (-2)*(y1==1)+0*(y1==0)
gen int x20= 0*(y1==1)+(-1)*(y1==0)

* INITIALISE VARIOUS VARIABLES USED WITHIN PROGRAM:

gen double A10=.
gen double A11=.
gen double A20=.
gen double A21=.

gen double N=.

gen double wx11=.
gen double wx10=.
gen double wx21=.
gen double wx20=.

gen double p11=.
gen double p21=.

* SET STARTING VALUES:

mat start=( 0.993, 0.78,0.998,0.7531,0.657,-1.863,0.833)

*RUNNING ML: FULL EWA MODEL

ml model lf ewa /rho /delta /phi  /lambda /A10_start /A20_start /N_start
ml init start, copy
ml max, trace search(norescale)
est store ewa

* DEFINE CONSTRAINTS REQUIRED FOR RL AND BL:

constraint 1 [delta]_b[_cons]=0.0
constraint 2 [rho]_b[_cons]=0.0
constraint 3 [delta]_b[_cons]=1
constraint 4 [rho]_b[_cons]=[phi]_b[_cons]
constraint 5 [N_start]_b[_cons]=1

*  ESTIMATE RL AS RESTRICTED VERSION OF EWA:

ml model lf ewa /rho /delta /phi  /lambda /A10_start /A20_start /N_start, ///
 constraints(1 2 5)
ml init start, copy
ml max, trace search(norescale)
est store rl

*  ESTIMATE BL AS RESTRICTED VERSION OF EWA:

ml model lf ewa /rho /delta /phi  /lambda /A10_start /A20_start /N_start, ///
 constraints(3 4)
ml init start, copy
ml max, trace search(norescale)
est store bl

* LR TESTS FOR RL AND BL AS RESTRICTIONS OF EWA:

lrtest ewa rl
lrtest ewa bl
```

### 18.7.4 Results from EWA Model

The output corresponding to the full EWA model are as follows. Output from the restricted models (RL and BL) are not shown here, since these results are identical to those from earlier estimations.



```
                                                  Number of obs   =       5000
                                                  Wald chi2(0)    =          .
Log likelihood = -6800.9162                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rho          |
       _cons |   .9834178   .0191123    51.45   0.000     .9459585    1.020877
-------------+----------------------------------------------------------------
delta        |
       _cons |   .5408416   .1124522     4.81   0.000     .3204393    .7612439
-------------+----------------------------------------------------------------
phi          |
       _cons |   .9427034   .0214316    43.99   0.000     .9006982    .9847085
-------------+----------------------------------------------------------------
lambda       |
       _cons |   .8208911   .1502486     5.46   0.000     .5264092    1.115373
-------------+----------------------------------------------------------------
A10_start    |
       _cons |   .9261403   .2732887     3.39   0.001     .3905043    1.461776
-------------+----------------------------------------------------------------
A20_start    |
       _cons |  -2.108425   .5609236    -3.76   0.000    -3.207815   -1.009035
-------------+----------------------------------------------------------------
N_start      |
       _cons |   .8843406   .3065987     2.88   0.004     .2834182    1.485263
------------------------------------------------------------------------------
```

We report the MLEs together with their 95% confidence intervals in Table 18.1. We report these alongside the "true" parameter values used in the simulation. We see that all seven parameters are estimated such that the true parameters are contained within the relevant confidence interval. This result is reassuring because it is consistent with correctness of both the simulation program and the estimation routine.

The most important parameter in EWA is $\delta$, which essentially represents the weight placed on foregone payoffs. The point estimate is 0.54 which is close to the true value of 0.60. The interpretation is that both actual pay-offs (as in RL) and foregone pay-offs (as in BL) are important, but the latter slightly more so because the estimate is greater than 0.5. The 95% confidence interval is from 0.320 to 0.761 and this indicates reasonable precision given the modest dimensions of the sample.

Finally, we compare the performance of EWA against RL and BL. Since EWA is a combination of RL and BL, the latter two are nested within the former, so it is natural to expect EWA to perform better. However, we an obvious question is

| Parameter | True value | 95% CI lower | MLE | 95% CI upper |
|---|---|---|---|---|
| $\rho$ | 0.97 | 0.945 | 0.983 | 1.021 |
| $\delta$ | 0.60 | 0.320 | 0.541 | 0.761 |
| $\phi$ | 0.94 | 0.901 | 0.943 | 0.985 |
| $\lambda$ | 0.80 | 0.526 | 0.821 | 1.115 |
| $A_1^0(0)$ | 1.0 | 0.391 | 0.926 | 1.147 |
| $A_2^0(0)$ | −2.0 | −3.208 | −2.108 | −1.009 |
| $N(0)$ | 1.0 | 0.283 | 0.884 | 1.485 |

Table 18.1: Point and interval estimates from simulation of EWA model, alongside true parameter values

whether the difference in performance is statistically significant. This question can be addressed using LR tests.

In Table 18.2 we collect the results from all three models. The maximised log-likelihoods for the three estimated models, and the results from the two LR tests conducted at the end of the estimation program, are presented below:

| Model | LogL | LR | df | p-value |
|---|---|---|---|---|
| EWA | −6800.92 | | | |
| RL | −6863.09 | 124.34 | 3 | 0.0000 |
| BL | −808.30 | 14.76 | 2 | 0.0006 |

Both RL and BL are rejected strongly by the likelihood ratio test, indicating that EWA is the preferred model of the three. This is hardly surprising, since the data was simulated from an EWA model with $\delta$ parameter chosen so that the "truth" is roughly halfway between RL and BL. It is clear that RL is rejected much more strongly than BL.

It has already been noted from Table 18.1 that the EWA estimates are close to their true values. It is natural to ask what are the consequences on parameter estimation from estimating the "wrong" model. The answer to this can be seen in Table 18.2. Possibly the most striking result is that the "initial attractions" appear to be greatly exaggerated as a consequence of estimating either RL or BL. Also striking is the severe downward bias in both models in the estimation of the parameter $\lambda$ representing sensitivity to attractions. Finally, we see that the recency parameter ($\phi$) is severly under-estimated under RL, but not under BL.

On the basis of these observations, it seems important to identify the best-fitting model before attempting to interpret parameter values.

| Parameter | RL | BL | EWA |
|---|---|---|---|
| $\rho$ | 0 | 0.953(0.019) | 0.983(0.019) |
| $\delta$ | 0 | 1 | 0.541(0.112) |
| $\phi$ | 0.768(0.085) | 0.953(0.019) | 0.943(0.021) |
| $\lambda$ | 0.110(0.016) | 0.425(0.036) | 0.821(0.150) |
| $A_1^0(0)$ | 1.390(1.457) | 1.684(0.469) | 0.926(0.273) |
| $A_2^0(0)$ | −8.311(2.286) | −4.255(0.770) | −2.108(0.561) |
| $N(0)$ | 1 | 0.472(0.150) | 0.884(0.307) |
| $n$ | 100 | 100 | 100 |
| $T$ | 50 | 50 | 50 |
| $LogL$ | −6863.09 | −6808.30 | −6800.92 |

Table 18.2: Maximum Likelihood Estimates from Reinforcement Learning (RL), Belief Learning (BL) and Experience Weighted Attraction (EWA) models applied to simulated data

Notes: Asymptotic standard errors in parentheses. Entries without standard errors are restricted parameters. In BL, estimates of $\rho$ and $\phi$ are constrained to be equal.



## 18.8 Summary and Further Reading

In various places in earlier chapters we have investigated the effect of experience in the form of task number, or round number. This was typically used as a means of assessing whether behaviour is evolving towards the equilibrium. In this chapter we have delved deeper by addressing the questions of why and how the movement towards equilibrium occurs.

The first learning model considered was the Directional Learning (DL) model. The data set was from Selten & Stoecker (1986). The econometric model used by Selten & Stoecker (1986) to analyse this data was a Markov learning model, with the changes in intended deviation period determined by transition probabilities which depend on experience in the previous period. Here, we have simply applied a linear regression model to the same data, and the results appear consistent with the predictions of DL.

The other learning models considered in this chapter were Reinforcement Learning (RL), Belief Learning (BL) and Experience Weighted Attraction (EWA). Readers interested in RL are referred to Erev & Roth (1998), while those interested in BL are referred to Cheung & Friedman (1997). A clear and straightforward summary of all of these models (and other learning models) is provided by (Wilkinson & Klaes, 2012, ch 9). A more advanced summary is provided by Camerer (2003) and readers particularly interested in EWA are referred also to Camerer & Ho (1999). Feltovich (2000) compares the predictive performance of a number of learning models.

An important point is that EWA is a combination of RL and BL that includes both as special cases. Because EWA is a heavily parameterised model, and consequently the estimation program for EWA is quite complicated, we have estimated the models on simulated data only. This has been useful from the point of view of verifying that the estimation program is correct.

The EWA model assumed in this chapter must be classified as a representative agent model because it assumes that all agents are characterised by an identical learning process. As always, we need to consider what the impact of heterogeneity, this time in the learning process, would be. Wilcox (2006) conducts a Monte Carlo study of the effect of parameter heterogeneity in EWA. He finds that when certain parameters vary across subjects, but such heterogeneity is neglected in estimation, it results in a severe bias on the estimation of the parameter $\delta$, in such a way as to favour RL over BL.

Clearly, it would be desirable to estimate EWA assuming between-subject heterogeneity in certain parameters, although this would obviously raise the level of parameterisation even further. A simpler approach to introducing heterogeneity would be to adopt a finite mixture approach. It could be assumed that there are two "types" of learners: RL and BL. An outcome from estimation of such a model would be estimates of the proportions of each of the two types. It is important to understand how such a mixture model would differ from standard EWA: the latter assumes that all agents learn according to exactly the same weighted average of RL and BL.

## Exercises

1. Estimate the EWA model using the real data contained in the file pursue-evade used in Chapter 16.
2. The EWA model combines reinforcement learning and belief learning in the sense that all agents are assumed to be driven by a combination of the two theories. Consider instead the assumption that each agent learns *either* according to reinforcement learning *or* according to belief learning. Such an assumption leads, of course, to a finite mixture model with two types. Consider how you would approach the estimation of such a model.
3. Using the simulate command introduced in Chapter 9, conduct a Monte Carlo investigation of the performance of the estimation techniques applied to the learning models in this chapter. By introducing heterogeneity in appropriate parameters, attempt to reproduce the results obtained by Wilcox (2006).

# Chapter 19
# Summary and Conclusion

The purpose of this concluding chapter is simply to draw together the dominant themes of the book, and also to identify some possible areas of future research in Experimetrics that are motivated by these themes.

## 19.1 Experimental Design Issues

Although the book is mainly about data analysis, some advice was also provided that was intended to be useful at the stage of designing the experiment. Probably the most basic question that needs to be addressed in the process of designing an experiment is: how many subjects? The obvious answer to this is: as many as possible given the budget constraint. A more useful answer is: as many as is required for the adopted framework for statistical inference to be valid. In Chapter 2, a primer in power analysis was provided. This is a useful framework for determining how many subjects are required in each treatment to achieve a chosen level of statistical adequacy. The framework also includes a means of choosing the optimal sample sizes subject to cost constraints, which is of undoubted importance in the planning of economic experiments.

However, the sample-size question becomes more complicated when we then ask: how many tasks should we expect each subject to engage in. Each subject could be set just one task; indeed, theories in their purest form are often presented in the context of a single decision, and it is statistically desirable for a sample to contain only one observation per subject since the independence assumption is more likely to hold. However, there are very strong reasons for using repeated tasks. Most obviously it generates more data for analysis. Also, it provides an opportunity for subjects to learn, and it is likely (although by no means obvious) that the researcher is more interested in the behaviour of experienced than inexperienced subjects. Sometimes it is the learning process that is the focus, in which case repetition is essential.

One issue that may be relevant to choosing the number of tasks is that of the subject's expected earnings: the higher the earnings, the more tasks we can reasonably expect subjects to invest effort in. This then brings us to a trade-off between subject number and task number. If the budget were to increase, would it be better to recruit more subjects, or to increase the number of tasks and offer more generous pay-offs? Of course, this relates to the question of the importance of task-related incentives. Bardsley et al. (2009, chapter 6) provide a balanced discussion of the role of incentives and arrive at the conclusion that the benefits from offering task-related incentives depend very much on the context, and in particular on the objectives of the study.

One obvious alternative to task-related incentives is the offer of a fixed participation fee. This is the style of payment that is required if the tasks are hypothetical in nature. An obvious problem with hypothetical tasks, and also potentially with paid tasks that are characterised by a "flat-maximum" (see Harrison, 1989), is that subjects may not invest due effort in the tasks, in the knowledge that the fee will be paid anyway. This is a problem that has been identified in the marketing area (see, for example, Menictas et al., 2011) in which subjects who respond randomly in online surveys are referred to as "straightliners" (Maronick, 2009) or "satisficers" (Krosnick, 1991). In this book, the concept of the "tremble" has surfaced many times, but it was always taken to be at the level of a single decision. A "straightliner" may be perceived as a subject who "trembles" permanently. A reasonable conjecture is that straightliners are sometimes present in economic experiments. Clearly, the presence of straightliners is likely to bring about a bias in estimation. It is therefore useful to be able to identify them, so that their choices can be deleted before estimation, or at least so that they can be treated as a separate "type" in a mixture. Another reason for identifying straightliners is perhaps so that they can be removed from the experimental database, bringing about a saving of future expense. The method for identifying them would be a Bayesian procedure similar to the one we used to identify "zero-types" in Chapter 11.

Of course, if we imagine keeping a set of conscientious subjects in a lab for an unacceptably long duration, it is natural to expect them to become too bored or tired to continue investing the desired effort. Hence, we expect all subjects to become straightliners eventually, and this leads back to the question of how many tasks each subject should be reasonably expected to engage in, even with the promise of a generous show-up fee.

These issues relate closely to work on decision times which is rapidly gaining the interest of experimental economists. In Chapter 5, we considered the modelling of decision times in a risky choice framework, interpreting decision time as a direct measure of effort expended. An obvious attraction of this measure is that it is easy to measure with great accuracy. In Chapter 5, decision time was seen as the outcome, and we were interested in identifying the features of the task that induce the most effort.

In other contexts, decision times could be used for a different purpose: estimating the pay-offs from higher effort, or in the terminology of Camerer & Hogarth (1999), estimating the "marginal product of labour" for experimental subjects. This could be achieved if decision time were used as an explanatory variable in a model in which actual pay-off is the outcome variable. To our knowledge this has yet to be done.

The recent surge of interest in decision times is perhaps a sign of a growing interest in the decision process generally. Traditionally, economists have been



interested principally in the final decision made. Perhaps a realisation is emerging that valuable insights can also be gained from observing the process leading to the decision. Of course, decision time is only one of many aspects of the process. Another means of observing the process is eye-tracking (Holmqvist et al., 2011) and this is also growing in popularity in experimental economics. Methods for analysing eye-tracking data will inevitably become an important topic within Experimetrics.

Chapter 14 contains material drawn from the wealth of results from statistical optimal design theory, to inform the design of a class of economic experiment. Similar goals have been pursued in a broader context by List et al. (2011). One specific example of the importance of experimental design was seen in section 15.5 in which we set out to estimate the parameters of a utility function for altruism based on hypothetical choices between allocations. There we commented on problems that arose in terms of certain parameters not being identified as a result of shortcomings in the experimental design. This setting is, in fact, very similar to that of the discrete choice experiment (DCE), in which there is a vast literature on optimal design (see, for example Louviere et al., 2008). It may benefit experimental economics hugely to draw on these pools of knowledge from other disciplines.

## 19.2 Experimetrics and Theory Testing

When testing theory, we often start with that theory's "fundamental prediction". It has been stressed that, in situations in which the fundamental prediction is rejected, there are other levels at which the theory can be tested. These often take the form of tests of a model's comparative static predictions. We have demonstrated that the most convenient framework in which to conduct such tests is that of linear regression, since in this context, influences on the outcome other than the treatment of interest can be controlled for, and the perennial problem of dependence in the data can be addressed using well-established techniques.

We have explained in detail how panel data models and multi-level models can be used in order to account for the inevitable phenomenon of between-subject heterogeneity.

We briefly considered meta analysis as an alternative approach to theory testing. This certainly seems to be a useful approach, in view of the enormous numbers of experiments that have been conducted and reported in recent decades. Pooling of results inevitably leads to benefits in terms of stronger conclusions, and better still, new insights.

## 19.3 Data Features

There are many reasons why it is important to pay attention to features of the data being analysed. A distinction that was made early in the book was between nominal, ordinal, and cardinal data. Which of these is available determines, to an extent, the type if treatment test or estimation approach that is appropriate. This distinction is particularly important in the choice between parametric and non-parametric treatment tests.

An important point is that data is always "discrete" in the sense that all systems of measurement require some level of rounding of individual measurements. But this is not a reason for modelling the outcome as discrete. If the researcher is truly concerned that the rounding system is influencing results, interval regression analysis may be used.

Even if the variable of interest is continuous according to the theoretical model, it might be deliberately observed at intervals, or there may be some element of discreteness. For example, in some types of experiment, data is best described as "lumpy", with accumulations of observations at particular "focal points". This sort of data has not been paid much attention in this book. Methods for dealing with lumpiness is likely to be particularly important in the analysis of data from coordination games.

Censoring is a very important data feature. In several different applications, for example testing for reciprocity an a public goods game, we have seen how estimation procedures that disregard censoring can lead to seriously biased estimates.

In the context of Social Preference models, it has been demonstrated in Chapter 15 that it is possible to embed censoring within the economic model, by attributing zero observations to binding non-negativity constraints in the agent's utility maximisation problem. We consider this to be a promising avenue for future research.

We have also promoted the idea that there is often more than one type of zero in experimental data. There are censored zeros, of the type just described, and also "zero types", that is, subjects whose contribution is destined to be zero whatever the circumstances of the task. As a means of dealing with the two types of zero simultaneously, we have promoted the use of the hurdle framework, and we have paid particular attention to the "panel hurdle model", that is appropriate with repeated data, and allows for between-subject heterogeneity.

## 19.4 Experimetrics of Social Preferences

The modelling of social preferences has been an important theme. The modelling is often in the context of dictator game giving. Data on dictator game giving has been analysed in many different ways throughout the book: using treatment tests, both between-subject and within-subject; using the hurdle modelling framework; using structural models containing preference-for-giving parameters; using mixture models; and using discrete choice models to model choices over hypothetical allocations.

This is a very good example of the general principle that the same research questions can be addressed using many different econometric approaches. A common feature of most of these approaches is that they incorporate between-subject heterogeneity. Conclusions (for example about the proportion of the population who are self-interested) are broadly similar between approaches.



There appears to be growing interest in the effect on social preferences of whether income is earned or "manna from heaven". Particularly striking results were reported by Cherry et al. (2002) who found that dictator game giving is dramatically reduced when the endowment is earned. Our own findings on this (Conte & Moffatt, 2014) are that the population divides roughly equally into those who consider the source of the endowment to be relevant to the decision of the final allocation, and those who do not. This result was obtained from estimation of a finite mixture model. A different sort of result was found by Erkal et al. (2011), and verified in Chapter 11 of this book: selfish individuals tend to self-select into high earning positions. Clearly further research is required on the impact of earned income on behaviour in experiments.

## 19.5 Risk Experimetrics

Behaviour under risk has been a recurring theme in the book. Most of this analysis has been carried out in the context of a risky choice experiment, in which subjects engage in a sequence of pairwise lottery choices, on the understanding that, on completion of the sequence, one of their chosen lotteries will be selected at random and played for real. Our favoured modelling approach is the one described in detail in Chapter 13 which assumes a utility function with the risk-attitude parameter varying continuously across the population, and also estimates a probability weighting function to allow deviations from EU. The continuous heterogeneity in risk attitude is handled by the method of maximum simulated likelihood. A number of recently published papers, namely Harrison & Rutström (2009), Conte et al. (2011), and Bruhin & Epper (2010), go one step further, by also allowing heterogeneity in the parameters of the weighting function. Specifically, they assume that a proportion of the population are EU maximisers, and the remainder behave according to prospect theory, with the weighting parameters varying continuously within this second group. Hence, the two different approaches to dealing with heterogeneity that have been covered in this book, namely the finite mixture approach, and the method of maximum simulated likelihood (in the case of Conte et al., 2011), are combined. Both Conte et al. (2011) and Bruhin & Epper (2010) estimate the proportion of EU maximisers in the population to be around 20%; Harrison & Rutström (2009) obtain a somewhat higher estimate of around 50%. There is clearly scope for further research in this area including the trying of alternative parametric specifications for the utility and weighting functions, trying of alternative stochastic specifications, allowing for the effects of experience, and so on.

Utility functions that are typically assumed (such as CRRA and CARA) have a single parameter representing risk aversion, and this parameter is closely related to the curvature (i.e. the second derivative) of the utility function. However, there is growing interest in more elaborate utility functions. There is particular interest in generalisations that allow tests of "prudence" and "temperance", measures closely associated with (respectively) the third and fourth derivatives of the utility function (see Eeckhoudt & Schlesinger, 2006). It would be a worthy exercise to produce a single utility function in which risk aversion, prudence and temperance are all free parameters, and to design a choice experiment in such a way as to enable estimation of all three.

Another extension that appears to be attracting interest is the allowance for "complexity aversion" (see, for example Sonsino et al., 2002; Moffatt et al., 2015). These studies extend the decision maker's utility function to include a term representing the level of complexity of the object (e.g. lottery) being evaluated. Whether individuals are complexity averse, complexity neutral or indeed complexity loving is of obvious interest, and other interesting questions follow, such as whether complexity aversion changes with experience.

An alternative protocol for the elicitation of risk attitudes which received some attention in Chapter 6 is the multiple price list (MPL) method, in which subjects are presented with an ordered list of lottery pairs, and essentially asked where in the list they would "switch" from one column to the other. The advantages and disadvantages of various types of MPL when used as elicitation devices are discussed by Andersen et al. (2006). Some researchers, when working with MPLs, treat the sequence of implied "choices" as choice data, and proceed using the type of strategies suggested in this book for the modelling of repeated choice data. However, there are serious concerns with such an approach, because the expectation on subjects simply to indicate their "switch-point" places a strong correlation structure on the choice sequence. Although "reversals back and forth" are sometimes observed, they are rare. When such a correlation structure is correctly incorporated, the information conveyed by the choice sequence becomes equivalent to simply an interval of the risk-attitude parameter. Given this, it is sensible to approach estimation along the lines of the interval regression model as discussed in Section 6.6.1. However, when this route is taken, a drawback is that any information on within-subject variability is lost. In models using repeated choice data, it is commonly found that within-subject variation is at least as important as between-subject variation. Given the current popularity of the MPL protocol, it is suggested that work needs to be done on finding ways of using the method that allow within subject variability to be identified. Immediately obvious suggestions for meeting this objective are to present the sequence of lottery pairs (1) in random order; and (2) in isolation of each other.

An interesting development of the MPL elicitation method is due to Tanaka et al. (2010). Here, subjects are faced with two MPLs and their switch-point is elicited for each. Then the two switch-points are simultaneously used to infer an interval for risk-attitude, and an interval for the weighting parameter. Both can be modelled using interval regression techniques. Better still, a form of bivariate interval regression could be used to model the two simultaneously, so that the results would include an estimate of the correlation between risk attitude and weighting parameter. This direction has been followed recently by Conte et al. (2015).

A final point about the MPL technique is that it can be used for the elicitation of preferences other than risk preferences. For example, Coller & Williams (1999) and Andersen et al. (2008) have used MPLs to elicit subjective discount rates. Time preference is yet another area of growing interest, which unfortunately it has not been possible to cover in this book due to space constraints.



The "preference reversal" (PR) phenomenon (Grether & Plott, 1979) was used as an application of within-subject treatment testing in Chapter 3. This is the phenomenon of subjects choosing the safer of two lotteries (the "p-bet") when asked to choose between them, but to contradict this choice by placing a higher valuation on the riskier lottery (the "$-bet") when asked to value them (that is, to provide their certainty equivalent) separately. As demonstrated in Chapter 3, treatment tests are useful for detecting the extent of the PR phenomenon. However, our view is that a structural approach could yield much deeper insights. The proposed structural model is based on the following reasoning. It is widely accepted that the vast majority of subjects are risk averse, and this risk aversion is revealed clearly when the subject is asked to choose between two lotteries. In contrast, when the certainty equivalent of a single lottery is elicited, subjects have a tendency to report a valuation that is close to the expected value of the lottery; in other words, they tend towards risk-neutrality in valuation tasks. Since $-bets typically have higher expected values than p-bets, this tendency provides an explanation for the PR phenomenon.

Using appropriate structural models, risk preference parameters should therefore be estimated separately for the two types of task. The results from such estimation could firstly be used to confirm (or refute) the conjecture that subjects tend towards risk neutrality in valuation tasks. Secondly, the estimates of the risk attitude parameters from the two tasks, in conjunction with the estimated variance parameters, could be used to create an algorithm for predicting the proportion of preference reversals for any given pair of lotteries, whether appearing in the experiment or not. Hence the structural estimates could be used in out-of-sample prediction of preference reversals. This is undoubtedly a worthy pursuit.

## 19.6 The Experimetrics of Games

There were noticeable changes in the nature of the econometric modelling when we moved from individual decision making to the analysis of choices in interactive games. Firstly, in the context of Quantal Response Equilibrium (QRE), the task of estimation was made complicated by the fact that the probabilities of a player's actions are a function of the probabilities of the other player's actions, in addition to the model parameters. This essentially means that these probabilites cannot be expressed separately. Consequently, within each likelihood evaluation, both players' choice probabilities need to be computed simultaneously using numerical optimisation techniques.

Secondly, the models of learning that we considered in Chapter 18 were, essentially, dynamic panel data models. In individual decision-making, learning is captured simply by making certain parameters depend on task number. This is because learning is only about the game itself. In interactive games played repeatedly, players learn not only about the structure of the game, but about the behaviour of others. Hence the outcomes from each stage of the game need to be incorporated in the learning process. This gives rise to what are essentially dynamic econometric models of learning.

The most complex of these learning models was Experience Weighted Attraction (EWA). The use of simulated data was useful in the demonstration of this model, not least because it gave us a way of confirming that it is possible to estimate the large number of parameters correctly and with relative precision.

The issue of heterogeneity arose yet again in the context of learning in games. Wilcox's (2006) Monte Carlo study establishes that neglect of heterogeneity leads us to favour one class of learning model over another. There is clearly a suggestion implicit in this finding that estimation of learning models should somehow incorporate between-subject heterogeneity. This could be achieved by applying the MSL approach to selected parameters in the EWA model, although this would require estimation of an even higher number of parameters.

An easier step in this direction would be to use the finite mixture approach to combine RL and BL in order to estimate the proportion of agents who are of each type. It is quite possible that this mixture model would perform better than standard EWA, which is essentially a representative agent model built on the assumption that all agents learn according to exactly the same weighted average of RL and BL.

For depth of reasoning models, heterogeneity is again the key, because agents are assumed to separate between different levels of reasoning. The finite mixture framework has been promoted as the most suitable approach. It must be said that the mixture models assumed were fairly basic, including a small number of low level types and a naïve-Nash type. It is possible to go further. Stahl & Wilson (1995), in the context of 3 × 3 symmetric games, includes in addition a "wordly" type (who assumes the existence of the low-level types and the naïve-Nash types) and a rational expectations (RE) type (who assumes the existence of all the other types and the *co-existence* of other RE agents). Econometric modelling of these extended mixture models becomes relatively complex, especially with regard to the correct modelling of the behaviour of the RE type. This seems to be a promising area for future research.

## 19.7 Heterogeneity

Heterogeneity has been mentioned in nearly all of the sections above, and is arguably the most important theme that has been covered in this book. It appears to be relevant in nearly all areas of experimental economics. On many occasions we have seen that a failure to deal with it leads to incorrect conclusions.

There are broadly two types of heterogeneity: discrete heterogeneity, which is handled using finite mixture models; and continuous heterogeneity, for which our chosen approach has been the method of maximum simulated likelihood (MSL). In certain settings, there are elements of both types of heterogeneity, and the two approaches are used in combination.

Another important distinction is between observed and unobserved heterogeneity. Most of the situations dealt with in this book have been the latter. However, dealing with observed heterogeneity (i.e. between-subject differences explained by observed subject characteristics) is clearly a sensible approach in situations in which



relevant subject characteristics (e.g. gender) are indeed observed. Harrison et al. (2007) and Tanaka et al. (2010) both follow this approach.

In all of the applications of MSL that have been considered, there has been only one dimension of heterogeneity, that is, only one parameter that is assumed to vary between subjects. However, readers should be aware that the MSL approach is particularly well suited to situations in which there is more than one dimension of heterogeneity, for example: risk aversion and probability weighting (Conte et al., 2011); risk aversion and complexity aversion (Moffatt et al., 2015); risk aversion, loss aversion, time preference, and randonmess (Von Gaudecker et al., 2011). It is anticipated that with continuing advances in computer power, such models with multiple dimensions of heterogeneity are likely to become more commonplace.

# Appendix A
# List of Data Files and Other Files

The following files, referred to at different points within the book, will be made available online (www.palgrave.com/moffatt).

**STATA data sets:**

lottery_choice_sim
holdup
Forsythe
give_take_sim
common_value_sim
risky_choice_sim
decision_times_sim
house_money_sim
ug_sim
ug_sm_sim
holtlaury_sim
interval_data_sim
exact_sim
bardsley
emotions
mixture_sim
fairness_sim
acquire_sim
bardsley
erkal
clark
garp
ES_sim
chowdhury
beauty_sim
cog_hier_sim

pursue_evade
selten-stoecker
pursue_evade_sim

**STATA Do-files:**

complete do-file

**Excel sheets:**

house money calculations
risk aversion calculations
proposer decision

# Appendix B
# List of STATA Commands

**General rules for commands in STATA:**

Lines starting with * are treated as comment lines.

Sometimes, a command contains a comma. The main part of the command comes before the comma; options come after the comma. Sometimes, only the beginning of the command is required, e.g. `su` for `summarize`.

STATA is case-sensitive, e.g. the variable `X` is treated as different from `x`.

When you are stuck, a very useful thing to do is to click on help → search and type in a key-word. You will be shown a page of the STATA manual. Often, the most useful part of these pages is the Examples section at the end.

**Combining data sets (definitely use help --> search if you ever need these):**

`append` adds new observations (i.e. rows) from another data set
`joinby` brings in new variables (i.e. columns) from another data set

**Inspecting/altering data:**

To inspect the data, click the Data Editor (Browse) icon.

`order i t x y` moves these four variables to the first four columns of the data matrix

To key in a dataset, or to make minor changes to your data, click on the Data Editor (Editor) icon.

`drop if t ≤ 15` deletes all observations for which $t < 15$
`keep if gender = = 1` keeps all observations for which gender=1
`clear` deletes all variables from memory

### Storing commands in a file: do-files:

Click on the Do-file editor icon. Type your commands into the window. Save to a file. Click the → icon to run the complete set or the next icon the selected set of commands. The following commands are useful within do-files:

`set more off` used in a do.file to enable the output to complete in the results window
`quietly replace x=.` causes command to be run without output appearing on screen.
`capture` command that precedes another command in order to suppress output. It is useful in do-files because it allows execution to continue in spite of errors, for example...
`capture gen x2=x*x` allows execution of do-file to continue even if the variable `x2` already exists
`///` continuation of command line
`//` comment marker; everything to the right of this is treated as a comment

### Doing a quick calculation:

`display 2*2` performs the calculation $2 \times 2$

### Simple data analysis:

`summarize x` gives mean, s.d., minimum and maximum of x
`summ x1-x4, detail` gives more detailed summary statistics (including percentiles) for the four variables x1, x2, x3 and x4
`summ x if d==0` gives summary statistics of x, over observations for which values of another variable, d, takes the value zero
`ci x` gives 95% confidence interval for the population mean of the variable x
`tabulate x` gives frequency distribution of x
`tab x y` gives cross-tabulation of two (discrete) variables x and y
`tab x y, summ(z)` creates a cross-tabulation of the two (discrete) variables x and y, with summary statistics of a third variable, z, shown in each cell
`correlate x1 - x3` finds correlation matrix of x1, x2 and x3

### Graphs:

When you create a graph, you can click on the start graph editor icon, and then make whatever changes you need. Most of the options in the commands listed below can be obtained using the graph editor.

`hist x, bin(20) freq` creates a histogram of x, with 20 bars, and frequency measured on vertical
`hist x, discrete` creates a histogram from a discrete variable x
`scatter y x, title("y against x")` scatter plot of y against x, with title
`scatter y x, xlabel(0(10)100) ylabel(0(1)10)` sets ranges for the two axes; the number in the middle is the increment for tick-marks



`scatter y x, xlabel(0(10)100) ylabel(0(1)10) || lfit y x` adds fitted regression line to scatter
`scatter y x, jitter(0.1)` scatter with "jitter"
`lowess y x` scatter of y against x, with a smoother superimposed
`lowess y x, bwidth(0.2)` smoother obtained using narrow bandwidth
`line y t` time series plot of y; t is the time trend variable
`line u t, yline(0)` plots the variable u against time, with a horizontal line at zero; useful when plotting residuals
`line x1 x2 x3 t, lpattern(solid dash dot)` plots three different lines with different patterns

### Creating new variables:

`rename var1 x` renames the variable from var1 to x
`variable label x "amount contributed"` assigns a label to the variable x
`gen logx = ln(x)` generates a variable logx, being the natural log of x; log(x) can be used in place of ln(x)
`gen double rootx = sqrt(x)` generates a variable rootx, being the square root of x, making sure the new variable is stored in double precision
`gen y = x * z` multiplication
`gen y = x / z` division
`gen x2 = x ^ 2` raising to the power 2
`gen int dum1 = x= =1` creates a dummy variable, dum1, taking the value one when x=1, and zero otherwise, making sure the new variable is stored as integer
`tab x, gen (x)` simultaneously creates a set of dummy variables, x1,x2,x3, from a categorical variable x
`gen dum10 = x >=10` creates a dummy variable, dum10, being 1 when x is greater than or equal to 10, and zero otherwise

Note: if a variable already exists, you need to use the `replace` command, instead of `generate`, to redefine it. `replace` is also useful for the following:

`replace x= . if y= =0` changes x to missing whenever y is 0

Related commands are:

`recode x 0= .` changes zeros to missing in x
`recode x 1=2 2=1 *= .` replaces 1's with 2's, 2's with 1's, and anything else with missing
`rename x y` renames a variable but see above to change the label
`recast double x` change storage type to double precision

### Power analysis:

`sampsi 10 12, sd(5) onesam oneside p(0.8)` find sample size for one-sample one-sided test of population mean = 10
`sampsi 10 12, sd1(4.0) sd2(5.84) oneside p(0.8)` find sample sizes for treatment test

`sampsi 10 12, sd1(4.0) sd2(5.84) oneside p(0.8) r(0.5)` find sample sizes for treatment test with unequal costs

**Treatment tests:**

`ttest y=5` one sample t-test of null that population mean equals 5
`bitest y=0.5` binomial test for testing the proportion of 1's in a 0/1 variable
`tab y1 y2, col exact` Fisher's exact test for two binary variables
`tab y1 y2, col chi2` chi-squared test for two binary variables
`tabi 62 68 13 7, exact chi2 col` direct calculation of chi-squared test
`ttest y, by(treatment)` independent samples t-test, assuming equal variances
`ttest y, by(treatment) unequal` independent samples t-test, assuming unequal variances
`bootstrap t=r(t),nodots rep(10000): ttest y,by(treatment)` bootstraps the t-test
`sdtest y, by(treatment)` test for equal variances
`sktest y` test for normality
`ranksum y, by(treatment)` Mann-Whitney test
`ksmirnov y, by(treatment)` Kolmogorov-Smirnov test
`ttest after=before` paired samples t-test
`signrank after=before` Wilcoxon test

**Regression**

`regress y x1 x2` OLS regression of y on x1 and x2, with an intercept
`regress y x if t<=15` OLS regression only using observations with $t \leq 15$
`hettest` performs Breusch-Pagan test for heteroscedasticity following a regression
`regress y x1 x2, noconstant` leaves out intercept
`regress y x1 x2, robust` generates heteroscedasticity-robust, or White-corrected, standard errors
`regress y x1 x2, vci(cluster i)` generates cluster-robust standard errors, with clustering assumed at the level of i
`regress y x1 x2, vci(bootstrap, rep(999) cluster(i))` block-bootsrap

**Testing hypotheses about regression parameters**

`regress y x1` OLS regression of y on x1 only
`est store rest` store estimates from last regression as "rest"
`regress y x1 x2 x3` OLS regression of y on x1, x2, and x3
`est store unrest` store estimates from last regression as "unrest"
`test x2 x3` test joint significance of x2 and x3 using an F-test
`lrtest unrest rest` test joint significance of x2 and x3 using a LR test
`test x1+x2 = 1` tests restriction that the parameters associated with x1 and x2 in the last regression sum to 1



`test (x1+x2=1) (_cons=0)` tests the restriction that the intercept is zero jointly with the restriction of the previous line

**Functions of parameters: the delta method**

`regress y x`
`nlcom threshold: -_b[_cons]/_b[x]` computes horizontal intercept, along with standard error
`regress y x1 x2`
`nlcom tot_eff: _b[x1]+_b[x2]` computes total effect, i.e. the sum of the slope coefficients, along with standard error

**Fitted values and residuals**

`predict yhat, xb` creates a variable yhat, being the fitted values from the last regression
`predict uhat, resid` creates a variable uhat, being the residuals from the last regression

Note: you need to change the name of the variable if you use these commands more than once.

**Creating scalars**

`scalar ten = 10` creates a scalar named "ten"
`scalar list ten` displays scalar
`scalar rsq = e(r2)` stores R-squared from most recent regression, as "rsq"
`scalar rss=e(rss)` stores residual sum of squares from most recent regression, as "rss")

**Panel data**

`by i: gen t=_n` generates t variable (task number)
`sort i t` sort data by i and then by t
`by i: gen sumy=sum(y)` generate cumulative sum of y separately for each i
`by i: gen int first=1 if _n==1` generate a binary variable taking the value 1 for the first observation in each block; missing otherwise
`by i: gen int last=1 if _n==_N` generate a binary variable taking the value 1 for the last observation in each block; missing otherwise
`xtset i t` declare data to be panel data
`xtdescribe` requests information on the dimensions of the panel
`xtline x` produce time series plots of variable x separately for each block of data
`xtreg y x1 x2 x3, fe` fixed effects
`xtreg y x1 x2 x3, re` random effects
`predict uhat, u` stores estimated random effects as "uhat"

**Hausman Test of fixed versus random effects**

`xtset i t` declare data to be panel data

`xtreg y x1 x2 x3, fe` fixed effects
`est store fe` store fixed effects estimates as "fe"
`xtreg y x1 x2 x3, re` random effects
`est store re` store random effects estimates as "re"
`hausman fe re` perform Hausman test comparing fe and re estimates

## Multi-level Modelling

`xtmixed y x || i:` two-level (equivalent to random effects)
`xtmixed y x || session: || i:` three-level
`xtmixed y x || session: || i: x` three-level with random slope (slope differs for different i)

## Binary, censored, interval, and ordinal data

`logit y x1 x2 x3` simple logit
`probit y x1 x2 x3` simple probit
`margins, dydx(x1) at(x1=0)` computes marginal effect of x1 at x1=0
`margins, dydx(x1)` computes average marginal effect of x1
`tobit y x1 x2 x3, ll(0) ul(10)` two-limit tobit, with limits 0 and 10
`intreg lower upper y x1 x2 x3, ll(0) ul(10)` interval regression with lower limits in "lower" and upper limits in "upper"
`oprobit y x1 x2 x3` ordered probit

## Simulation

`set seed 12345678` set random number seed
`set obs 1000` set sample size
`gen double u=runiform()` generate random numbers from uniform(0,1)
`gen double z=invnorm(uniform())` generate random numbers from Normal(0,1)
`gen double z=rnormal()` easier way to generate random numbers from Normal(0,1)
`mat c = (1,0.5 \0.5,1)` creates a correlation matrix for next command
`drawnorm z1 z2, n(2000) corr(c)` generates two standard normal variables, with correlation matrix given by c
`gen int n=rpoisson(3)` generate random numbers from Poisson(3)

## Simulating panel data

`set obs 1000` set sample size
`egen int i=seq(), f(1) t(50) b(20)` generate subject identifier, with n=50, T=20
`egen int t=seq(), f(1) t(20)` generate task number
`xtset i t` declare data as panel
`by i: gen double u=0.5*rnormal() if _n==1` generate random effect in first row for each subject
`by i: replace u=u[1] if u==.` copy random effect into other rows for each subject



## The ml routine

`ml model lf likprog () () ()` specify ml problem as maximising function computed by program "likprog" with respect to three parameters
`ml max` run the maximization
`ml max, trace` run the maximization showing more detailed output
`ml model lf likprog a b c` specify same ml problem as above, but with names assigned to the three parameters
`ml model d0 likprog a b c` required for panel data models
`ml check` use this to locate error in code
`ml coefleg` obtain the legends for each estimated parameter

## Loops

`local xlist x1-x25` Define variable list
`foreach v in 'xlist'` start loop over variables in xlist
`forvalues t in 1(1)50` start loop over values of 't' from 1 to 50

## Stata add-ons

The following are user-written programs which are used in the text. They are not available in STATA but are available online. Use the `findit` command to locate them and install them.

`escftest` Epps-Singleton test
`cdfplot` cumulative frequency graph
`mdraws` for generating Halton sequences
`fmm` for estimating finite mixture models

# Appendix C
# Choice Problems Used in Chapters 5 and 13

| t | p1 | p2 | p3 | q1 | q2 | q3 | Complexity | Mean of cti | Mean of dt | $r^*$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.05 | 0 | 0.95 | 0 | 1 | 0 | 1 | 0.239 | 3.49 | 0.074 |
| 2 | 0.09 | 0 | 0.91 | 0 | 1 | 0 | 1 | 0.175 | 3.62 | 0.136 |
| 3 | 0.11 | 0 | 0.89 | 0 | 1 | 0 | 1 | 0.145 | 4.11 | 0.168 |
| 4 | 0.13 | 0 | 0.87 | 0 | 1 | 0 | 1 | 0.117 | 4.33 | 0.201 |
| 5 | 0.15 | 0 | 0.85 | 0 | 1 | 0 | 1 | 0.09 | 4.6 | 0.234 |
| 6 | 0.17 | 0 | 0.83 | 0 | 1 | 0 | 1 | 0.066 | 5.09 | 0.269 |
| 7 | 0.19 | 0 | 0.81 | 0 | 1 | 0 | 1 | 0.048 | 6.24 | 0.304 |
| 8 | 0.22 | 0 | 0.78 | 0 | 1 | 0 | 1 | 0.041 | 5.59 | 0.358 |
| 9 | 0.26 | 0 | 0.74 | 0 | 1 | 0 | 1 | 0.061 | 5.36 | 0.434 |
| 10 | 0.3 | 0 | 0.7 | 0 | 1 | 0 | 1 | 0.097 | 4.17 | 0.514 |
| 11 | 0.35 | 0 | 0.65 | 0 | 1 | 0 | 1 | 0.154 | 3.77 | 0.621 |
| 12 | 0.4 | 0 | 0.6 | 0 | 1 | 0 | 1 | 0.209 | 3.25 | 0.737 |
| 13 | 0.45 | 0 | 0.55 | 0 | 1 | 0 | 1 | 0.264 | 2.97 | 0.862 |
| 14 | 0.5 | 0 | 0.5 | 0 | 1 | 0 | 1 | 0.318 | 2.78 | 1.0 |
| 15 | 0.6 | 0 | 0.4 | 0 | 1 | 0 | 1 | 0.426 | 3.41 | 1.32 |
| 16 | 0.75 | 0 | 0.25 | 0 | 1 | 0 | 1 | 0.59 | 2.77 | 2.0 |
| 17 | 0.9 | 0 | 0.1 | 0 | 1 | 0 | 1 | 0.775 | 3.07 | 3.32 |
| 18 | 0.5 | 0 | 0.5 | 0.48 | 0.52 | 0 | 2 | 0.16 | 3.9 | 0.06 |
| 19 | 0.5 | 0 | 0.5 | 0.44 | 0.56 | 0 | 2 | 0.127 | 4.83 | 0.16 |
| 20 | 0.5 | 0 | 0.5 | 0.42 | 0.58 | 0 | 2 | 0.111 | 5.07 | 0.21 |
| 21 | 0.5 | 0 | 0.5 | 0.4 | 0.6 | 0 | 2 | 0.095 | 4.35 | 0.26 |
| 22 | 0.5 | 0 | 0.5 | 0.38 | 0.62 | 0 | 2 | 0.078 | 5.01 | 0.31 |
| 23 | 0.5 | 0 | 0.5 | 0.36 | 0.64 | 0 | 2 | 0.062 | 5.23 | 0.36 |
| 24 | 0.5 | 0 | 0.5 | 0.34 | 0.66 | 0 | 2 | 0.046 | 5.47 | 0.40 |
| 25 | 0.5 | 0 | 0.5 | 0.32 | 0.68 | 0 | 2 | 0.033 | 6.66 | 0.44 |
| 26 | 0.5 | 0 | 0.5 | 0.3 | 0.7 | 0 | 2 | 0.027 | 6.49 | 0.49 |
| 27 | 0.5 | 0 | 0.5 | 0.28 | 0.72 | 0 | 2 | 0.03 | 6.61 | 0.53 |
| 28 | 0.5 | 0 | 0.5 | 0.26 | 0.74 | 0 | 2 | 0.038 | 6.52 | 0.57 |
| 29 | 0.5 | 0 | 0.5 | 0.24 | 0.76 | 0 | 2 | 0.048 | 6.35 | 0.60 |
| 30 | 0.5 | 0 | 0.5 | 0.22 | 0.78 | 0 | 2 | 0.062 | 5.92 | 0.64 |
| 31 | 0.5 | 0 | 0.5 | 0.2 | 0.8 | 0 | 2 | 0.08 | 5.24 | 0.68 |
| 32 | 0.5 | 0 | 0.5 | 0.16 | 0.84 | 0 | 2 | 0.117 | 4.95 | 0.75 |

Table C.1: The 50 choice problems, with complexity level, mean of closeness to indifference, mean of decision time (in seconds), and threshold risk attitude ($r^*$)

| t | p1 | p2 | p3 | q1 | q2 | q3 | Complexity | Mean of cti | Mean of dt | $r^*$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 33 | 0.5 | 0 | 0.5 | 0.1 | 0.9 | 0 | 2 | 0.179 | 4.41 | 0.85 |
| 34 | 0.5 | 0 | 0.5 | 0.02 | 0.98 | 0 | 2 | 0.28 | 3.96 | 0.97 |
| 35 | 0.39 | 0.2 | 0.41 | 0.2 | 0.6 | 0.2 | 3 | 0.105 | 4.88 | 0.93 |
| 36 | 0.38 | 0.2 | 0.42 | 0.2 | 0.6 | 0.2 | 3 | 0.094 | 5.39 | 0.86 |
| 37 | 0.36 | 0.2 | 0.44 | 0.2 | 0.6 | 0.2 | 3 | 0.072 | 6.23 | 0.74 |
| 38 | 0.34 | 0.2 | 0.46 | 0.2 | 0.6 | 0.2 | 3 | 0.05 | 6.82 | 0.62 |
| 39 | 0.32 | 0.2 | 0.48 | 0.2 | 0.6 | 0.2 | 3 | 0.028 | 7.44 | 0.51 |
| 40 | 0.3 | 0.2 | 0.5 | 0.2 | 0.6 | 0.2 | 3 | 0.015 | 8.05 | 0.42 |
| 41 | 0.28 | 0.2 | 0.52 | 0.2 | 0.6 | 0.2 | 3 | 0.02 | 7.71 | 0.32 |
| 42 | 0.27 | 0.2 | 0.53 | 0.2 | 0.6 | 0.2 | 3 | 0.03 | 7.04 | 0.28 |
| 43 | 0.26 | 0.2 | 0.54 | 0.2 | 0.6 | 0.2 | 3 | 0.041 | 6.42 | 0.23 |
| 44 | 0.25 | 0.2 | 0.55 | 0.2 | 0.6 | 0.2 | 3 | 0.053 | 6.37 | 0.19 |
| 45 | 0.24 | 0.2 | 0.56 | 0.2 | 0.6 | 0.2 | 3 | 0.065 | 6.03 | 0.15 |
| 46 | 0.23 | 0.2 | 0.57 | 0.2 | 0.6 | 0.2 | 3 | 0.077 | 5.44 | 0.11 |
| 47 | 0.25 | 0.75 | 0 | 0.5 | 0.5 | 0 | 2 | 0.209 | 3.44 | 0 |
| 48 | 0 | 0.25 | 0.75 | 0 | 0.5 | 0.5 | 2 | 0.072 | 5.17 | 0 |
| 49 | 0.25 | 0 | 0.75 | 0.25 | 0.25 | 0.5 | 2 | 0.072 | 5.07 | 0 |
| 50 | 0.25 | 0.25 | 0.75 | 0.25 | 0.5 | 0.25 | 3 | 0.428 | 3.81 | 0 |

Table C.1: (Continued)

# References


Abdellaoui, M., Barrios, C. & Wakker, P. P. (2007), 'Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory', *Journal of Econometrics* **138**, 356–378.

Aitchison, J. & Silvey, S. (1957), 'The generalisation of probit analysis to the case of multiple responses', *Biometrika* **44**, 138–140.

Al-Ubaydli, O. & List, J. A. (2013), On the generalizability of experimental results in economics: With a response to Camerer, Technical report, National Bureau of Economic Research.

Allais, M. (1953), 'Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine', *Econometrica* **21**, 503–546.

Alos-Ferrer, C., Granic, D.-G., Shi, F. & Wagner, A. K. (2012), 'Choices and preferences: Evidence from implicit choices and response times', *Journal of Experimental Social Psychology* **48**, 1336–1342.

Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. E. (2006), 'Elicitation using multiple price list formats', *Experimental Economics* **9**, 383–405.

Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. E. (2008), 'Eliciting risk and time preferences', *Econometrica* **76**(3), 583–618.

Andreoni, J. (1988), 'Why free-ride? Strategies and learning in public goods experiments', *Journal of Public Economics* **37**, 291–304.

Andreoni, J. & Miller, J. (2002), 'Giving according to GARP: An experimental test of the consistency of preferences for altruism', *Econometrica* **70**, 737–753.

Andreoni, J. & Vesterlund, L. (2001), 'Which is the fair sex? Gender differences in altruism', *Quarterly Journal of Economics* **116**, 293–312.

Ansari, A., Montoya, R. & Netzer, O. (2012), 'Dynamic learning in behavioral games: A hidden Markov mixture of experts approach', *Quantitative Marketing and Economics* **10**, 475–503.

Arellano, M. & Bond, S. (1991), 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *Review of Economic Studies* **58**, 277–297.

Atkinson, A. (1996), 'The usefulness of optimum experimental designs', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 59–76.

Baik, K. H., Chowdhury, S. M. & Ramalingam, A. (2014), Resources for conflict: Constraint or wealth?, Technical report, School of Economics, University of East Anglia, Norwich, UK.

Ball, S. B., Bazerman, M. H. & Carroll, J. S. (1991), 'An evaluation of learning in the bilateral winner's curse', *Organizational Behavior and Human Decision Processes* **48**(1), 1–22.

Baltagi, B. (2008), *Econometric analysis of panel data*, Vol. 1, John Wiley & Sons, New York.

Bardsley, N. (2000), 'Control without deception: Individual behaviour in free-riding experiments revisited', *Experimental Economics* **3**, 215–240.

Bardsley, N. (2008), 'Dictator game giving: Altruism or artefact?', *Experimental Economics* **11**, 122–133.

Bardsley, N. & Moffatt, P. G. (2007), 'The experimetrics of public goods: Inferring motivations from contributions', *Theory and Decision* **62**, 161–193.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P. G., Starmer, C. & Sugden, R. (2009), *Experimental economics: Rethinking the rules*, Princeton University Press, Princeton, NJ.

Bazerman, M. H. & Samuelson, W. F. (1983), 'I won the auction but don't want the prize', *Journal of Conflict Resolution* **27**(4), 618–634.

Becker, G., DeGroot, M. & Marschak, J. (1964), 'Measuring utility by a single-response sequential method', *Behavioural Science* **9**, 226–232.

Bellemare, C., Kröger, S. & Van Soest, A. (2008), 'Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities', *Econometrica* **76**, 815–839.

Berenson, M. L., Levine, D. & Rindskopf, D. (1988), *Applied statistics: A first course*, Prentice Hall, New York.

Berg, J., Dickhaut, J. & McCabe, K. (1995), 'Trust, reciprocity, and social history', *Games and Economic Behavior* **10**(1), 122–142.

Binswanger, H. P. (1980), 'Attitudes toward risk: Experimental measurement in rural India', *American Journal of Agricultural Economics* **62**(3), 395–407.

Bolten, G. & Ockenfels, A. (2000), 'A theory of equity, reciprocity and competition', *American Economic Review* **90**, 166–193.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R. & Satorra, A. (2010), 'A finite mixture analysis of beauty-contest data using generalized beta distributions', *Experimental Economics* **13**(4), 461–475.

Bosman, R. & van Winden, F. (2002), 'Emotional hazard in a power-to-take experiment', *Economic Journal* **112**, 147–169.

Branas-Garza, P. (2007), 'Promoting helping behaviour with framing in dictator games', *Journal of Economic Psychology* **28**, 477–486.

Bruhin, A, a. F.-D. & Epper, T. (2010), 'Risk and rationality: Uncovering heterogeneity in probability distortion', *Econometrica* **78**, 1375–1412.

Buchanan, J. M., Tollison, R. D. & Tullock, G. (1980), *Toward a theory of the rent-seeking society*, number 4, Texas A & M University Press, College Station, TX.

Burnham, T. (2003), 'Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving', *Journal of Economic Behaviour and Organisation* **50**, 133–144.

Burton, M., Tomlinson, M. & Young, T. (1994), 'Consumers decisions whether or not to purchase meat: A double hurdle analysis of single adult households', *Journal of Agricultural Economics* **45**, 202–212.

Buschena, D. & Zilberman, D. (2000), 'Generalized expected utility, heteroscedastic error, and path dependence in risky choice', *Journal of Risk and Uncertainty* **20**, 67–88.

Busemeyer, J. R. & Townsend, J. T. (1993), 'Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment.', *Psychological Review* **100**(3), 432.

Camerer, C. (2003), *Behavioral game theory: Experiments in strategic interaction*, Princeton University Press, Princeton, NJ.

Camerer, C. & Ho, T.-H. (1999), 'Experience-weighted attraction learning in games', *Econometrica* **87**, 827–874.

Camerer, C. & Hogarth, R. M. (1999), 'The effects of financial incentives in experiments: A review and capital-labor-production framework', *Journal of Risk and Uncertainty* **19**, 7–42.

Camerer, C., Ho, T.-H. & Chong, J.-K. (2003), 'Models if thinking, learning, and teaching in games', *American Economic Review* **93**, 192–195.

Camerer, C., Ho, T.-H. & Chong, J.-K. (2004), 'A cognitive hierarchy model of games', *Quarterly Journal of Economics* **119**, 861–898.

Cameron, A. & Trivedi, P. K. (2010), *Microeconometrics using stata* (revised edition), STATA Press: College Station, TX.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø. & Tungodden, B. (2007), 'The pluralism of fairness ideals: An experimental approach', *The American Economic Review* **97**, 818–827.



Cappellari, L. & Jenkins, S. P. (2006), 'Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation', *ISER Working Paper 2006-16. Colchester: University of Essex.*

Chaudhuri, P. & Mykland, P. A. (1993), 'Non-linear experiments: Optimal design and inference based on likelihood', *Journal of the American Statistical Association* **88**, 538–546.

Chaudhuri, P. & Mykland, P. A. (1995), 'On efficient designing of non-linear experiments', *Statistica Sinica* **5**, 421–440.

Cherry, T. L., Frykblom, P. & Shogren, J. F. (2002), 'Hardnose the dictator', *American Economic Review* **92**, 1218–1221.

Cheung, S. L. (2014), 'New insights into conditional cooperation and punishment from a strategy method experiment', *Experimental Economics* **17**(1), 129–153.

Cheung, Y.-W. & Friedman, D. (1997), 'Individual learning in normal form games: Some laboratory results', *Games and Economic Behavior* **19**(1), 46–76.

Chlaß, N. & Moffatt, P. G. (2012), Giving in dictator games: Experimenter demand effect or preference over the rules of the game?, Technical report, Jena Economic Research Papers.

Chowdhury, S. M., Sheremeta, R. M. & Turocy T. L. (2014), 'Overbidding and overspreading in rent-seeking experiments: Cost structure and prize allocation rules', *Games and Economic Behavior* **87**, 224–238.

Clark, J. (2002), 'House money effects in public good experiments', *Experimental Economics* **5**, 223–231.

Clarke, K. A. (2003), 'Non-parametric model discrimination in international relations', *Journal of Conflict Resolution* **47**, 72–93.

Cleveland, W. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* **74**, 829–836.

Cohen, J. (2013), *Statistical power analysis for the behavioral sciences*, Routledge Academic, London.

Coller, M. & Williams, M. B. (1999), 'Eliciting individual discount rates', *Experimental Economics* **2**(2), 107–127.

Collins, L. M. & Lanza, S. T. (2010), *Latent class and latent transition analysis for the social, behavioral, and health sciences*, Wiley, New York.

Conlisk, J. (1989), 'Three variants on the Allais Example', *American Economic Review* **79**(3), 392–407.

Connolly, T. & Butler, D. (2006), 'Regret in economic and psychological theories of choice', *Journal of Behavioral Decision Making* **19**, 139–154.

Conte, A. & Moffatt, P. (2014), 'The econometric modelling of social preferences', *Theory and Decision* **76**(1), 119–145.

Conte, A., Hey, J. D. & Moffatt, P. (2011), 'Mixture models of choice under risk', *Journal of Econometric* **162**, 79–88.

Conte, A., Moffatt, P. G. & Riddel, M. (2015), 'Heterogeneity in risk attitudes across domains: A bivariate random preference approach', *CBESS Discussion Paper 15-10, University of East Anglia* .

Cox, D. R. (1970), *The analysis of binary data*, Methuen, London.

Cragg, J. (1971), 'Some statistical models for limited dependent variables with application to the demand for durable goods', *Econometrica* **39**, 829–844.

Crawford, V. P. & Iriberri, N. (2007), 'Fatal attraction: Salience, naivete, and sophistication in experimental "hide-and-seek" games', *The American Economic Review* **97**, 1731–1750.

Crawford, V. P., Costa-Gomes, M. A. & Iriberri, N. (2013), 'Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications', *Journal of Economic Literature* **51**(1), 5–62.

Croson, R. & Gneezy, U. (2009), 'Gender differences in preferences', *Journal of Economic Literature* **47**, 448–474.

Davis, D. & Holt, C. A. (1993), *Experimental Economics*, Princeton University Press, Princeton, NJ.

Daykin, A. R. & Moffatt, P. G. (2002), 'Analyzing ordered responses: A review of the ordered probit model', *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences* **1**(3), 157–166.

Deaton, A. & Irish, M. (1984), 'Statistical models for zero expenditures in household budgets', *Journal of Public Economics* **23**, 59–80.

Deb, P. (2012), 'Fmm: Stata module to estimate finite mixture models', *Statistical Software Components*.

Dionne, G., Artis, M. & Guillen, M. (1996), 'Count data models for a credit scoring system', *Journal of Empirical Finance* **3**, 303–325.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. & Wagner, G. G. (2011), 'Individual risk attitudes: Measurement, determinants, and behavioral consequences', *Journal of the European Economic Association* **9**(3), 522–550.

Dong, D. & Kaiser, H. (2008), 'Studying household purchasing and nonpurchasing behaviour for a frequently consumed commodity: Two models', *Applied Economics* **40**(15), 1941–1951.

Drichoutis, A. C., Nayga Jr, R. M., Lazaridis, P. & Park, B. S. (2011), 'A consistent econometric test for bid interdependence in repeated second-price auctions with posted prices', *Atlantic Economic Journal* **39**(4), 329–341.

Eckel, C. & Grossman, P. (1998), 'Are women less selfish than men? Evidence from dictator experiments', *The Economic Journal* **108**, 729–735.

Eckel, C. & Grossman, P. J. (2001), 'Chivalry and solidarity in ultimatum games', *Economic Inquiry* **39**, 171–188.

Eeckhoudt, L. & Schlesinger, H. (2006), 'Putting risk in its proper place', *American Economic Review* **96**, 280–289.

Efron & Tibshirani (1993), *An introduction to the bootstrap. Monographs on Statistics and Applied Probability, No. 57*, Chapman and Hall, New York.

El-Gamal, M. A. & Grether, D. M. (1995), 'Are people Bayesian? Uncovering behavioral strategies', *Journal of the American statistical Association* **90**(432), 1137–1145.

Ellingsen, T. & Johannesson, M. (2004), 'Promises, threats and fairness', *The Economic Journal* **114**, 397–420.

Engel, C. (2011), 'Dictator games: A meta study', *Experimental Economics* **14**(4), 583–610.

Engel, C. & Moffatt, P. G. (2012), 'Estimation of the house money effect using hurdle models', *Preprints of the Max Planck Institute for Research on Collective Goods Bonn* **2012/13**.

Engel, C. & Moffatt, P. G. (2014), 'dhreg, xtdhreg, and bootdhreg: Commands to implement double-hurdle regression', *Stata Journal* **14**(4), 778–797.

Engelmann, D. & Strobel, M. (2004), 'Inequality aversion, efficiency, and maximin preferences in simple distribution experiments', *American Economic Review* **94**, 857–869.

Epps, T. W. & Singleton, K. J. (1986), 'An omnibus test for the two-sample problem using the empirical characteristic function', *Journal of Statistical Computation and Simulation* **26**, 177–203.

Erev, I. & Roth, A. (1998), 'Predicting how people play games: Reinforcement learning in experimental games with mixed strategy equilibria', *American Economic Review* **88**, 848–881.

Erkal, N., Gangadharan, L. & Nikiforakis, N. (2011), 'Relative earnings and giving in a real-effort experiment', *American Economic Review* **101**, 3330–3348.

Fechner, G. (1860), *Elements of Psychophysics, Vol. 1*, Holt, Rinehart and Winston, New York.

Fedorov, V. (1972), *Theory of Optimum Experiments*, Academic Press, New York.

Fehr, E & Schmidt, K. M. (1999), 'A theory of fairness, competition and cooperation', *Quarterly Journal of Economics* **114**, 817–868.

Feltovich, N. (2000), 'Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games', *Econometrica* **68**, 605–641.

Fischbacher, U., Gachter, S. & Fehr, E. (2001), 'Are people conditionally cooperative? Evidence from a public goods experiment', *Economics Letters* **71**, 379–404.

Fisman, R., Kariv, S. & Markovits, D. (2007), 'Individual preferences for giving', *The American Economic Review* **97**(5), 1858–1876.

Ford, I., Tosney, B. & Wu, C. F. J. (1992), 'The use of a canonical form in the construction of locally optimal designs for non-linear problems', *Journal of the Royal Statistical Society. Series B (Methodological)* **54**, 569–583.



Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. (1994), 'Fairness in simple bargaining experiments', *Games and Economic Behavior* **6**, 347–369.

Fréchette, G. R. (2001), 'sg158. 1 update to random-effects ordered probit', *Stata May 2001 Technical STB-61*, p. 12.

Fréchette, G. R. (2012), 'Session-effects in the laboratory', *Experimental Economics* **15**(3), 485–498.

Gächter, S. & Fehr, E. (2000), 'Cooperation and punishment in public goods experiments', *American Economic Review* **90**(4), 980–994.

Garcia, B. (2013), 'Implementation of a double-hurdle model', *Stata Journal* **13**(4), 776–794.

Georg, S. J. (2009), 'Nonparametric testing of distributions - the Epps-Singleton two-sample test using the empirical characteristic function', *The Stata Journal* **9**, 454–465.

Goeree, J. K., Holt, C. A. & Palfrey, T. R. (2002), 'Quantal response equilibrium and overbidding in private-value auctions', *Journal of Economic Theory* **104**(1), 247–272.

Goeree, J. K., Holt, C. A. & Palfrey, T. R. (2003), 'Risk averse behavior in generalised matching pennies games', *Games and Economic Behavior* **45**, 97–113.

Gould, W., Pitblado, J. & Poi, B. (2010), *Maximum likelihood estimation with STATA, Fourth edition*, STATA Press, College Station, TX.

Greene, W. (2008), Econometric analysis, *Sixth Edition*, Prentice Hall: New York.

Green, D. P. & Tusicisny, A. (2012), Statistical analysis of results from laboratory studies in experimental economics: A critique of current practice, *in* '2012 North American Economic Science Association Conference'.

Grether, D. & Plott, C. R. (1979), 'Economic theory of choice and the preference reversal phenomenon', *Economic Theory of Choice and the Preference Reversal Phenomenon* **69**, 623–638.

Güth, W., Schmittberger, R. & Schwarze, B. (1982), 'An experimental analysis of ultimatum bargaining', *Journal of Economic Behavior & Organization* **3**(4), 367–388.

Halton, J. (1960), 'On the efficiency of evaluating certain quasi-random sequences of points on evaluating multi-dimensional integrals', *Numerische Mathematik* **2**, 84–90.

Ham, J. C., Kagel, J. H. & Lehrer, S. F. (2005), 'Randomization, endogeneity and laboratory experiments: The role of cash balances in private value auctions', *Journal of Econometrics* **125**(1), 175–205.

Harless, D. & Camerer, C. (1994), 'The predictive utility of generalized expected utility theories', *Econometrica* **62**, 1251–1289.

Harrison, G. (1989), 'Theory and misbehavior of first price auctions', *American Economic Review* **79**, 949–963.

Harrison, G. (2007), 'House money effects in public good experiments: Comment', *Experimental Economics* **10**, 429–437.

Harrison, G. & Rutström, E. E. (2009), 'Expected utility theory and prospect theory: one wedding and a decent funeral', *Experimental Economics* **12**, 133–158.

Harrison, G., Johnson, E., McInnes, M. & Rutström, E. E. (2005), 'Risk aversion and incentive effects: Comment', *American Economic Review* **95**, 900–904.

Harrison, G. W., Lau, M. I. & Rutström, E. E. (2007), 'Estimating risk attitudes in Denmark: A field experiment', *The Scandinavian Journal of Economics* **109**(2), 341–368.

Harrison G.W., Lau, M. and Rutström E.E., (2015), 'Theory, experimental design and econometrics are complementary (and so are lab and field experiments)', chapter 15 in Fréchette G.R. and Schotter A. (eds.), *Handbook of Experimental Economic Methodology*, Oxford University Press, New York.

Harwell, M. R. & Gatti, G. G. (2001), 'Rescaling ordinal data to interval data in educational research', *Review of Educational Research* **71**, 105–131.

Hausman, J. (1978), 'Specification tests in econometrics', *Econometrica* **46**, 1251–1271.

Hey, J. D. (1995), 'Experimental investigations of errors in decision making under risk', *European Economic Review* **39**, 633–640.

Hey, J. D. (2001), 'Does repetition improve consistency', *Experimental Economics* **4**, 5–54.

Hey, J. D. & di Cagno, D. (1990), 'Circles and triangles: An experimental estimation of indifference lines in the marschak-machina triangle', *Journal of Behavioural Decision Making* **3**, 279–306.

Hey, J. D. & Orme, C. (1994), 'Investigating generalisations of expected utility theory using experimental data', *Investigating Generalisations of Expected Utility Theory Using Experimental Data* **62**, 1291–1326.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. & Van de Weijer, J. (2011), *Eye tracking: A comprehensive guide to methods and measures*, Oxford University Press, New York.

Holt, C. (1986), 'Reversals and the independence axiom', *American Economic Review* **76**, 508–515.

Holt, C. & Laury, S. K. (2002), 'Risk aversion and incentive effects', *American Economic Review* **92**, 1644–1655.

Houser, D. (2008), 'Experiments and econometrics', *The new Palgrave dictionary of economics Second edition.* Macmillan, London.

Huber, J. & Zwerina, K. (1996), 'The importance of utility balance in efficient choice designs', *Journal of Marketing Research* **23**, 307–317.

Jakiela, P. (2013), 'Equity vs. efficiency vs. self-interest: On the use of dictator games to measure distributional preferences', *Experimental Economics* **16**(2), 208–221.

Johnson, N. D. & Mislin, A. A. (2011), 'Trust games: A meta-analysis', *Journal of Economic Psychology* **32**(5), 865–889.

Johnson, C. A., Baillon, A., Bleichrodt, H., Li, Z., Van Dolder, D. & Wakker, P. P. (2014), 'Prince: An improved method for measuring incentivized preferences', *Available at SSRN 2504745.*

Jones, A. M. (1989), 'A double hurdle model of cigarette consumption', *Journal of Applied Econometrics* **4**, 23–39.

Kagel, J. H. & Levin, D. (1986), 'The winner's curse and public information in common value auctions', *The American Economic Review* **76**, 894–920.

Kagel, J. H., Levin, D. & Harstad, R. M. (1995), 'Comparative static effects of number of bidders and public information on behavior in second-price common value auctions', *International Journal of Game Theory* **24**(3), 293–319.

Kahneman, D. & Tversky, A. (1979), 'Prospect theory: An analysis of decisions under risk', *Econometrica* **47(2)**, 263–291.

Kanninen, B. (1993), 'Optimal experimental design for double-bounded dichotomous choice contingent valuation', *Land Economic* **69**, 138–146.

Keasey, K. & Moon, P. (1996), 'Gambling with the house money in capital expenditure decisions', *Economics Letters* **50**, 105–110.

Kennedy, P. E. (1998), 'Teaching undergraduate econometrics: A suggestion for fundamental change', *American Economic Review* **88**, 487–492.

Krishna, V. (2010), *Auction theory*, 2nd Edition Academic press, New York.

Laffont, J.-J., Ossard, H. & Vuong, Q. (1995), 'Econometrics of first-price auctions', *Econometrica: Journal of the Econometric Society* **63**, 953–980.

Ledyard, J. O. (1995), *Public goods: A survey of experimental research*, Princeton University Press, Princeton.

Lichtenstein, S. & Slovic, P. (1971), 'Reversals of preference between bids and choices in gambling decisions.', *Journal of experimental psychology* **89**(1), 46.

Likert, R. (1932), *A technique for the measurement of attitudes*, Vol. Archives of Psychology, No. 140, Columbia University Press, New York.

List, J. A. (2007), 'On the interpretation of giving in dictator games', *Journal of Political Economy* **115**, 482–493.

List, J. A., Sadoff, S. & Wagner, M. (2011), 'So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design', *Experimental Economics* **14**(4), 439–457.

Little, I. M. D. (1949), 'A reformulation of the theory of consumer's behaviour', *Oxford Economic Papers* **1**, 90-99.

Loomes, G. & Sugden, R. (1998), 'Testing different stochastic specifications of risky choice', *Economica* **65**, 581–598.

Loomes, G., Moffatt, P. G. & Sugden, R. (2002), 'A microeconometric test of alternative stochastic theories of risky choice', *Journal of Risk and Uncertainty* **24**, 103–130.



Louviere, J. J., Hensher, D. A. & Swait, J. D. (2000), *Stated choice methods: Analysis and application*, Cambridge University Press, Cambridge.

Louviere, J. J., Islam, T., Wasi, N., Street, D. & Burgess, L. (2008), 'Designing discrete choice experiments: Do optimal designs come at a price?', *Journal of Consumer Research* **35**(2), 360–375.

Luce, D. (1959), *Individual choice behavior*, Wesley, New York.

MacKinnon, J. G. (2002), 'Bootstrap inference in econometrics', *Canadian Journal of Economics/Revue canadienne d'économique* **35**(4), 615–645.

Maddala, G. (1983), *Limited dependent and qualitative variables in econometrics*, Cambridge University Press, New York.

Masters, G. N. (1982), 'A Rasch model for partial credit scoring', *Psychometrica* **47**, 149–174.

McDowell, A. (2003), 'From the help desk: Hurdle models', *The Stata Journal* **3**, 1738–184.

McKelvey, R. D. & Palfrey, T. R. (1995), 'Quantal response equilibria for normal form games', *Games and Economic Behavior* **10**(1), 6–38.

McLachlan, G. & Peel, D. (2000), *Finite mixture models*, Wiley, New York.

Menictas, C., Wang, P. & Fine, B. (2011), 'Assessing flat-lining response style bias in online research', *Australasian Journal of Market and Social Research* **19**, 34–44.

Moffatt, P. G. (1991), *Microeconometric models of household purchasing behaviour*, unpublished PhD Thesis, University of Bristol.

Moffatt, P. G. (2005*a*), 'Hurdle models of loan default', *Journal of the Operational Research Society* **56**, 1063–1071.

Moffatt, P. G. (2005*b*), 'Stochastic choice and the allocation of cognitive effort', *Experimental Economics* **8**, 369–388.

Moffatt, P. G. (2007), 'Optimal experimental design in models of decision and choice', Chapter 15 in Boumans M. (ed.), *Measurement in Economics: A Handbook*, Academic Press: London, 357–375.

Moffatt, P. G. & Peters, S. A. (2001), 'Testing for the presence of a tremble in economic experiments', *Experimental Economics* **4**, 221–228.

Moffatt, P. G., Sitzia, S. & Zizzo, D. J. (2015), 'Heterogeneity in preferences towards complexity', *Journal of Risk and Uncertainty*, forthcoming

Müller, W. G. & de Leon, A. C. P. (1996), 'Optimal design of an experiment in economics', *The Economic Journal* **106**, 122–127.

Nagel, R. (1995), 'Unravelling in guessing games: An experimental study', *American Economic Review* **85**, 1313–1326.

Nelson, F. D. (1976), 'On a general computer algorithm for the analysis of models with limited dependent variables', *Annals of Economic and Social Measurement* **5**, 493–509.

Neumann, L. J. & Morgenstern, O. (1947), *Theory of games and economic behavior*, Princeton University Press, Princeton, NJ.

Oehlert, G. W. (1992), 'A note on the delta method', *The American Statistician* **46**(1), 27–29.

Pratt, J. (1964), 'Risk aversion in the small and in the large', *Econometrica* **32**, 122–136.

Prelec, D. (1998), 'The probability weighting function', *Econometrica* **66**, 497–527.

Quiggin, J. (1982), 'A theory of anticipated utility', *Journal of Economic Behaviour and Organisation* **3**, 323–343.

Rabe-Hesketh, S. & Skrondal, A. (2008), *Multilevel and longitudinal modeling using Stata*, STATA press, College Station, TX.

Reiss, P. C. & Wolak, F. A. (2007), 'Structural econometric modeling: Rationales and examples from industrial organization', *Handbook of Econometrics* **6**, 4277–4415.

Roodman, D. (2009), 'How to do xtabond2: An introduction to difference and system GMM in stata'. *Stata Journal*, 9(1), 86–136.

Rosenthal R.W., Shachat J. & Walker M. (2003), 'Hide and seek in Arizona', *International Journal of Game Theory* **32**, 273–293.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M. & Zamir, S. (1991), 'Bargaining and market behaviour in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study', *American Economic Review* **81**, 1068–1095.

Runco, M. (2013), 'Estimating depth of reasoning in a repeated guessing game with no feedback', *Experimental Economics* **16**(3), 402–413.

Saha, A. (1993), 'Expo-power utility: A flexible form for absolute and relative risk aversion', *American Journal of Agricultural Economics* **75**, 905–913.

Seidl, C. (2002), 'Preference reversal', *Journal of Economic Surveys* **16**, 621–655.

Selten, R. (1967), 'Die strategiemethode zur erforschung des eingeschränckt rationalen verhaltens in rahmen eines oligopolexperiments', inbeitraägge zur experimentellen wirtschaftforschung, ed. by H. Sauermann, Tübingen: Jcb mohr.

Selten, R. & Stoecker, R. (1986), 'End behavior in sequences of finite prisoner's dilemma supergames a learning theory approach', *Journal of Economic Behavior & Organization* **7**(1), 47–70.

Shachat, J., Swarthout, J. T. & Wei, L. (2015), 'A hidden Markov model for the detection of pure and mixed strategy play in games', *Econometric Theory* **31**, 729–752.

Sheremeta, R. M. (2010), 'Experimental comparison of multi-stage and one-stage contests', *Games and Economic Behavior* **68**(2), 731–747.

Sheremeta, R. M. (2013), 'Overbidding and heterogeneous behavior in contest experiments', *Journal of Economic Surveys* **27**(3), 491–514.

Siegel, S. & Castellan, N. J. (1988), *Non-parametric statistics for the behavioral sciences, Second edition*, McGraw Hill, New York.

Silvey, S. D. (1980), *Optimal design: An introduction to the theory for parameter estimation*, Chapman and Hall, London.

Skrondal, A. & Rabe-Hesketh, S. (2004), *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*, CRC Press, Oxford.

Smith, M. D. (2003), 'On dependency in double-hurdle models', *Statistical Papers* **44**(4), 581–595.

Smith, V. L. (1982), 'Microeconomic systems as an experimental science', *The American Economic Review* pp. 923–955.

Solnick, S. (2001), 'Gender differences in the ultimatum game', *Economic Inquiry* **39**, 189–200.

Sonsino, D., Benzion, U. & Mador, G. (2002), 'The complexity effects on choice with uncertainty–experimental evidence*', *The Economic Journal* **112**(482), 936–965.

Stahl, D. O. (1996), 'Boundedly rational rule learning in a guessing game', *Games and Economic Behavior*, **16**, 303–330.

Stahl, D. O. & Wilson, P. W. (1995), 'On players models of other players: Theory and experimental evidence', *Games and Economic Behavior* **10**(1), 218–254.

Starmer, C. (2000), 'Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk', *Journal of economic literature* pp. 332–382.

StataCorp (2011), *Stata statistical software: Release 12*, StataCorpLP, College Station, TX.

Stott, H. P. (2006), 'Cumulative prospect theory's functional menagerie', *Journal of Risk and uncertainty* **32**(2), 101–130.

Tanaka, T., Camerer, C. F. & Nguyen, Q. (2010), 'Risk and time preferences: Linking experimental and household survey data from Vietnam', *American Economic Review* **100**, 557–571.

Thaler, R. & Johnson, E. (1990), 'Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice', *Management Science* **36**, 643–660.

Tobin (1958), 'Estimation of relationships for limited dependent variables', *Econometrica* **26**, 24–36.

Train, K. E. (2003), *Discrete choice methods with simulation*, Cambridge University Press, Cambridge.

Tversky, A. & Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and uncertainty* **5**(4), 297–323.

Tversky, A., Slovic, P. & Kahneman, D. (1990), 'The causes of preference reversal', *The American Economic Review* **80**, 204–217.

Von Gaudecker, H.-M., Van Soest, A. & Wengström, E. (2011), 'Heterogeneity in risky choice behavior in a broad population', *The American Economic Review* **101**, 664–694.

Vuong, Q. H. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**, 307–333.

Wakker, P. P. (2010), *Prospect Theory*, Cambridge University Press, Cambridge.





Wales, T. J. & Woodland, A. K. (1983), 'Estimation of consumer demand systems with binding non-negativity constraints', *Journal of Econometrics* **21**, 263–285.

Wilcox, N. T. (1994), 'On a lottery pricing anomaly: Time tells the tale', *Journal of Risk and Uncertainty* **8**, 311–324.

Wilcox, N. T. (2006), 'Theories of learning in games and heterogeneity bias', *Econometrica* **74**, 1271–1292.

Wilcox, N. T. (2008), 'Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison', *Research in Experimental Economics* **12**, 197–292.

Wilkinson, N. & Klaes, M. (2012), *An introduction to behavioral economics*, Palgrave Macmillan, London.

Wooldridge, J. (2012), *Introductory econometrics: A modern approach*, Cengage Learning, Mason, OH.

Yates, F. (1934), 'Contingency tables involving small numbers and the chi 2 test', *Supplement to the Journal of the Royal Statistical Society* **1**, 217–235.

Yi, K.-O. (2005), 'Quantal-response equilibrium models of the ultimatum bargaining game', *Games and Economic Behavior* **51**(2), 324–348.

Zelmer, J. (2003), 'Linear public goods experiments: A meta-analysis', *Experimental Economics* **6**(3), 299–310.

# Index