Mortgage Choice as a Natural Field Experiment on Choice Under Risk*

Philomena M. Bacon

University of East Anglia

Peter G. Moffatt

School of Economics

University of East Anglia

Norwich

NR4 7TJ

United Kingdom

November 15, 2010

MORTGAGE CHOICE AS A NATURAL FIELD EXPERIMENT ON CHOICE UNDER RISK

Abstract

Microdata from the UK Survey of Mortgage Lenders is used to model borrowers' choices between variable and fixed rate mortgages. The data is treated as a large-scale "natural experiment" on risky choice, with the choice of a fixed rate corresponding to the "safe choice" in a more conventional experimental setting. The choice is assumed to depend partly on risk attitude, and partly on expectations of future movements in interest rates. Approximately 280,000 choices, made by borrowers between 1992 and 2001, appear in the sample. The ordered probit model is used for estimation, while taking account of a number of econometric issues including missing counterfactuals, selectivity, and endogeneity. Explanatory variables are divided into three groups: mortgage price variables; interest rate expectations; and borrower characteristics. A number of strong effects are found, including: fixing is more likely when agents expect interest rates to rise; a larger amount borrowed (i.e. higher pay-offs in the choice problem) increases the propensity to fix; the presence of female borrowers increases the propensity to fix; older borrowers are less likely to fix; high-income borrowers are less likely to fix, particularly so if income is "self-certified". The results offer fresh insights into the analysis of risky choice, particularly with regard to the roles of incentives and subjects' income.

1 Introduction

The decision to purchase a home is undoubtedly one of the most important financial decisions made in the course of an economic agent's lifetime. A closely related decision is on how the purchase will be financed. For the majority of homebuyers, the purchase necessitates the taking out of a mortgage, that covers some or all of the purchase price of the property. A variety of types of mortgage are available. The feature of interest in this paper is whether the mortgage interest payments are fixed or variable. Borrowers may choose a "variable rate" mortgage, in which case their monthly interest payments would be determined by the prevailing market rate of interest, which can rise or fall at any time. Alternatively, they may choose a "fixed rate" mortgage, by which they are contracted to pay a fixed monthly amount over a specified (usually shorter) period.

It is clear that the choice between a fixed rate and a variable rate mortgage is a classic risky choice problem in the economic sense, and the decision resembles, in some ways at least, the type of risky choice problem typically created in laboratory settings. The fixed rate is the "safe choice" since it results in a sequence of monthly payments that are known with certainty in advance. The variable rate is the "risky choice", since it results in uncertainty over future monthly payments [1]. The choice that is made clearly depends on the risk attitude (i.e. the "preferences") of the decision maker. However, there is another important determinant of this choice: the decision maker's expectations of future movements of interest rates. It is easy to see that alternative combinations of preferences and expectations can produce the same choice: a risk-averse borrower who expects no change in interest rates might be expected to choose the fixed rate; however, a risk-seeking borrower might also be expected to choose the fixed rate if they happen to expect interest rates to rise sufficiently steeply in the near future. This is essentially the problem raised by Manski (2004), who considered identification in such situations. As prescribed by Manski, we need to make an identifying assumption about agents' expectations of future movements in interest rates. The

---

[1] Of course, there is the possibility that making the "wrong" choice can have more serious consequences than simply higher monthly payments. Imagine, for example, a borrower who takes out a variable rate mortgage, and then experiences an unexpected steep rise in the interest rate, to the extent that they become unable to service the mortgage. Such a borrower is likely to lose their home.

identifying assumption we adopt is that, at any point in time, all agents have the same interest-rate expectations, corresponding to "market expectations" implied by the prevailing term structure of interest rates.

The choice data is obtained from the UK Council of Mortgage Lenders (CML). It contains information on a five percent random sample of mortgage contracts financed by its members each month between 1992 and 2001. After filtering incomplete data, this amounts to approximately 280,000 observations. The dataset contains detailed demographics on the borrower as well as information about the loan.

In the terminology of Harrison and List (2004) this study may be classified as a "natural field experiment", since it is certainly true that subjects are operating in an environment in which the choice task is naturally undertaken, and also that subjects are unaware that they are participating in an experiment. Additional notable features of this study are as follows. First, the subject pool consists entirely of ordinary individuals, who have self-selected to purchase a widely demanded commodity - a mortgage. Second, the commodity is real and is purchased under normal economic conditions, in a highly competitive market. Third, information regarding choices is freely available to all subjects, and the choice itself is at the complete discretion of the subject, who is unhindered by time constraints. Fourth, the task, to choose between a fixed and a variable rate, requires no former experience, although some subjects are not first-time buyers, and so have presumably made a similar decision at least once previously.

One of the major attractions of this natural experiment is its findings in relation to the salience of the incentives. Completion of the choice task results in a binding contract with resulting pay-offs projected months or years into the future. Such incentives are high: the choice made can result in quite substantial gains or losses relative to the rejected alternative. In Section 2, we present examples of the types of outcomes actually arising in the context of the dataset. Given the salience of the outcome and the unlimited decision time, there is little doubt that subjects in this natural experiment can be relied upon to invest whatever cognitive effort is required to be in a position to express their true preferences.

Also, the task is performed once and only once, thereby avoiding the type of between-task contamination (see, for example, Holt (1986)) that may arise when a sequence of tasks is performed, as is routine in laboratory settings.

Another class of natural experiment with salient incentives is that of the Game Show. This research area has enjoyed a flurry of interest in recent years (see, for example, Post *et al.*, 2008). Choices analysed in this setting are clearly for real and have high stakes. However, researchers in this area have faced a raft of criticism: the participants self-select in a rather extreme way, making it hard to justify the use of results in predicting the behaviour of the wider population; only a very short period (usually a few minutes) is available for each deliberation; the decision-maker is operating in an unnatural and unfamiliar setting, and their decision is likely to be influenced by the presence and behaviour of the studio audience. Clearly, all of these problems are categorically avoided in the natural experiment considered in this paper.

Our natural experiment has further advantages over other types of experiment, that we set out to exploit to the full. One concerns the role of subject income. Economists are very interested in the effect of individual income on risk attitude. However, it is widely accepted that, in the context of laboratory experiments, "outside" income (or "permanent" income) has only an indirect effect on behaviour. The usual measure of "income" in these settings is simply the income earned in the experiment, and the utility function is defined over this variable. In contrast, in the context of our natural experiment, the pay-offs take the form of monthly payments that will come directly out of the subject's monthly income. Hence it is natural to expect "outside" income to have a direct effect on behaviour. We are keen to exploit this opportunity to quantify the effect of income on risk attitude.

One final advantage is the huge sample size: the choices of 280,654 decision-makers are used to estimate the various models we construct. This is at least an order of magnitude higher than the sample sizes typically encountered in lab-based research, and, accordingly, has the potential to lead to considerably stronger conclusions regarding some of the effects in which behavioural economists

may be most interested.

Of course, the fact that this analysis is based on a natural field experiment, with features that set it far apart from the "sterility" of the laboratory environment, presents us with a number of interesting and challenging econometric issues which add further potential for this work to contribute to the body of research in this field. The first of these is the absence of data on the counterfactual choice. While the rate struck by each borrower, whether they chose fixed or variable, is observed, the rate that they would have obtained had they made the other choice, is unknown. This requires some method of imputation, of which, as we shall see, there are a number of possibilities. The second econometric issue is that of selection bias: since borrowers have engaged in a process of search for the package that they find most attractive, it is likely that the choice they have made was driven by the unusually attractive features of the chosen deal, for example, having a lower rate than expected. If the rates struck are to be analysed, as they need to be in the process of imputing the counterfactual rates, such selection bias must be addressed. The third issue is endogeneity. In considering borrowers' choice between different rates, we need to examine how these rates have been determined. They are clearly set by the lenders, and, importantly, the lenders are using some of the same criteria in setting rates, namely those relating to interest rate expectations, as borrowers use in choosing between them. Hence the endogeneity of rates offered to borrowers must also be addressed in the estimation process, in order to obtain consistent estimates of the demand parameters.

Work has previously been carried out on the borrower's decision to fix rates. Alm and Follain (1984) and Campbell and Cocco (2003) provide theoretical treatments; the latter, like us, extract a measure of interest rate expectations from the term structure of interest rates. Empirical work on the fixing decision has typically consisted of binary choice models applied to relatively small sample sizes. For example, Dhillon *et al.*(1987) considered the choices of 78 households in Louisiana, USA; Brueckner and Follain (1988, 1989) used a dataset consisting of 475 households from across the USA; Leece (2000) used a sample of 762 UK households; Coulibaly and Li (2009) consider the choices of 2,878 households across the USA. The key determinant of the choice found in most of this litera-

ture is the rate differential, that is, the variable denoted in our section 3 as $p_f - p_v$. Some researchers have gone further by considering more than two categories of the fixing decision. Sa-Aadu and Sirmans (1995) conside five different levels of fixing, and use a multinomial logit model to condiser the choice between these alternatives of 345 midwestern (USA) households. The multinomial logit model is also used by Phillips and Vanderhoff (1994) who consider the choice between the three alternatives: government-backed fix; ordinary fix; and adjustable rate. Their sample size is 6,894.

There are examples of risky choice research being taken into "the field", for example Binswanger's (1980) study of the decisions of Indian farmers in choice problems with "high" pay-offs. However, with the exception of the gameshow studies mentioned earlier, full-scale "natural experiments" on risky choice are hard to find in the literature. To our knowledge, Bacon (2009) is the only empirical researcher to have treated the mortgage choice problem in the context of risky choice, thereby contributing to the literature on the estimation of factors determining risk attitudes. The contribution of the current paper is to bring this research area further into the domain of behavioural economics, in particular by separating the role of risk attitude from that of expectations, in the spirit of Manski (2004).

The paper is set out as follows. Section 2 briefly considers some institutional issues relating to the choice between fixed and variable rates in the UK, and also, for the sake of further motivation, provides specific examples of monetary outcomes that would have arisen from the different choices at selected points in time. In section 3, we develop a model of the choice between fixed and variable rates, and consider how the model should be estimated in order to deal with the various econometric issues raised earlier in this section. In section 4 we examine the data. In section 5 we present and discuss the results. Section 6 concludes.

2   Fixed and Variable Rate Mortgages in the United Kingdom

The uptake of fixed rate mortgages in the UK has been of growing interest to policy-makers in recent years. From the policy maker's viewpoint, fixed rates have the potential to bring about stability in the housing market, as clearly seen in some European countries, notably France and Germany during the pre-credit crunch period. In 2003, the then UK Chancellor of the Exchequer Gordon Brown advocated a widening in the use of fixed-rate mortgages as a precondition for British entry into the Euro. At the time of the 2003 Budget, he commissioned Professor David Miles to investigate why long-term fixed rate mortgages occupied such a small share of total business, and to make recommendations for addressing the issue. The final report appeared as Miles (2004). His recommendations included measures designed to raise levels of borrowers' understanding of interest rate risk, and measures making it less risky for lenders to lend at long-term fixed rates.

Figure 1 depicts the breakdown of new mortgage business between fixed and variable rates in the UK over the period 1992 to 2007. Prior to 2005 there appeared to be a broad preference for variable rates. However, more recently, perhaps partly as a consequence of the Miles report, there appears to have been a steady increase in the take-up of fixed-rate deals. This contrasts with the picture in the USA where according to the month interest rate survey of the USA Federal Housing Finance Board fixed rate mortgages accounted on average for 70% of the newly issued mortgages between the period 1985-2001.

UK borrowers who choose to fix are typically reluctant to fix for long periods. A two-year fix is fairly typical in the UK market. This contrasts starkly with the 10-15 year fixes typically seen in Europe, and the 15-25 year fixes seen in the USA [2]. When the period of the fix has elapsed, the rate automatically reverts to the Standard Variable Rate (SVR). Harsh early-redemption penalties typically apply if a borrower chooses to opt out of a fixed-rate agreement. In recent years upfront fees have also been introduced, some of which are refundable at the end

---

[2]The interest rate survey of the USA Federal Housing Finance Board reveals that long-term fixed rate mortgages accounted for 70% of newly issued mortgages in the USA between 1985 and 2001.
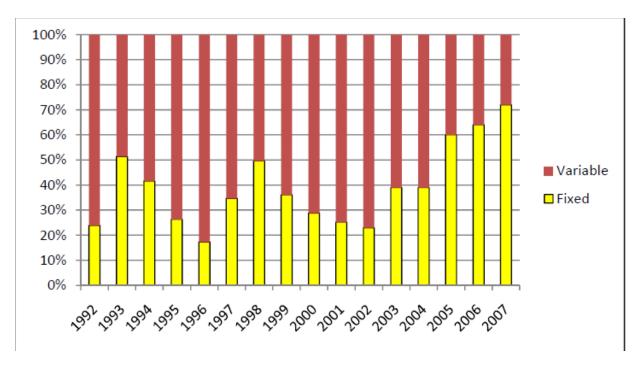
Figure 1: PROPORTIONS CHOOSING FIXED AND VARIABLE: 1992 - 2007.
Source: Survey of Mortgage Lenders. UKDA Essex.

of the fixed rate term, as a deterrent to early redemption.

To complete this section, we bring into focus the size of the pay-offs that characterise the choice problems of interest, by illustrating the monetary gains and losses that arise in particular scenarios. We consider a £100,000 mortgage, and the interest-only monthly payments[3] that would have actually been made by borrowers over a two-year period, under each interest rate regime, had the deals been struck on particular dates. The dates we consider are June 1996, and June 1998. The path of the monthly repayments relating to these deals is shown in Figures 2 and 3. For the fixed-rate calculation we use the mean fixed rate struck in the first month of the two-year period; for the variable-rate calculation, we use the mean variable rate for each month within the two year period. The total amounts payable under each deal (assuming zero discounting) are reported in Table 1.

---

[3]For simplicity we treat capital as being repaid at the end of the mortgage term, and do not add it to the monthly interest costs

Figure 2: PATH OF COMPARABLE MORTGAGE REPAYMENTS: JUNE 1996 – MAY 1998.

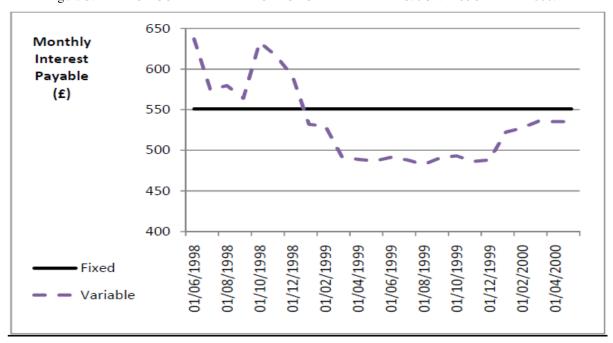Figure 3: PATH OF COMPARABLE MORTGAGE REPAYMENTS: JUNE 1998 – MAY 2000.

Table 1: SUMMARY OF INTEREST PAYABLE UNDER DIFFERENT RATE REGIMES

2 Years From	Variable Rate		Fixed Rate		Outcome
	(Rate)	Total Paid	(Rate)	Total Paid	
Jun 1996	(5.53%)	£13,303.15	(5.92%)	£11,834.98	±£1,485
Jun 1998	(7.64%)	£12,801.21	(6.61%)	£13,222.95	±£421

Starting rate in parentheses

As can be seen in Table 1, for the first period the fixed rate deal was cheaper by £1,485.17, whilst, in the second period the variable rate deal was cheaper by £421.74. The most revealing feature of these two examples is that in both cases, a "myopic" borrower, that is, one who bases their decision only on the price differential prevailing at the start of the two-year period, will make the "wrong" choice. The two graphs make this very clear. In Figure 2, the variable rate starts off below the fixed rate, but is above the fixed rate for most of the 2-year period, resulting in a net loss of £1,485 if the variable rate is chosen. In Figure 3, it is the fixed rate that is lower at the start of the period; thereafter, successive falls in the variable rate result in a net loss of £421, from choosing the fixed rate.

In the models developed in Section 4, we assume that borrowers are not myopic; they are forward-looking, basing their decision on expectations of interest rates that are reflected in the current term structure of interest rates. It is therefore interesting to investigate what these expectations are in the context of our two examples. Referring ahead to Figure 8, which shows the time paths of the 3-month yield and the 10-year yield, we see that the yield spread in June 1996 was +2.52 percentage points, whilst in June 1998 it was -1.71 percentage points. Hence we infer that, in June 1996, the market expected a fairly steep rise in interest rates, while in June 1998, the market expected a fairly steep fall. By incorporating these expectations into the decision-making process, forward-looking borrowers are guided towards the choice of a fixed rate in June 1996, and towards a variable rate in June 1998. Hence, in both cases, the prevailing interest rate expectations are seen to guide borrowers towards the choice that eventually turns out to be "correct".
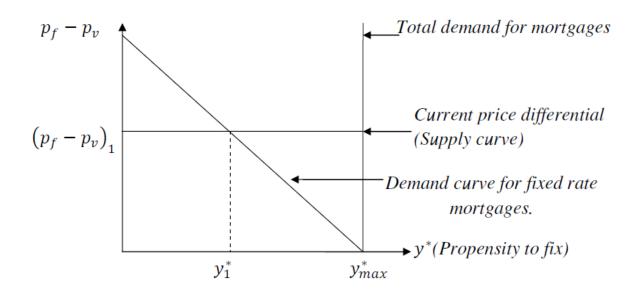
Figure 4: THE MARKET FOR FIXED RATE (AND VARIABLE RATE) MORTGAGES.

3   The Model

Since this paper is concerned with the choice between fixed and variable rate mortgages, and not the overall demand for mortgages, we commence from the assumption that total demand is fixed. If it aids understanding, the market can be perceived as consisting of a single borrower who has already made the decision to take out a mortgage, and is in the process of deciding between a fixed rate and a variable rate. In this context, the demand function discussed below may simply be interpreted as that single individual's propensity to choose a fixed rate over a variable rate. This propensity to fix will be denoted as y^*.

Figure 4 presents a demand and supply diagram for fixed and variable rate mortgages. Total demand (y^*_{max}) is represented by the vertical line. On the vertical axis is measured the fixed-variable price differential (i.e. the difference in the rates): $p_f - p_v$. The focus of the analysis is the proportion of total demand that is fixed-rate. This is represented by the downward sloping curve, which conveys the fact that a rise in the interest rate at which the fixed-rate deal is struck, ceteris

paribus, causes a fall in the propensity of a typical borrower to opt for the fixed rate. Supply of fixed rate mortgages is represented by the horizontal line. This is because we are assuming that the lender chooses both the fixed and variable rates, which gives rise to a horizontal supply curve at $(p_f - p_v)_1$. The resulting quantity demanded of fixed rate deals is then given by y_1^*, and the remainder of deals demanded $y_{max}^* - y_1^*$ are variable-rate.

Clearly, if lenders increase the price of the fixed rate and hold the variable rate constant, the quantity demanded (y_1^*) is expected to fall. The slope of the demand curve is therefore related to the inverse of the negative coefficient on the variable representing $p_f - p_v$, in the demand equation estimated later. However, a crucial point is that the factors that influence lenders' choice of fixed rate are some of the same factors that determine the demand for fixed rate deals. The factors in question relate to interest rate expectations. If the market expects interest rates to rise, borrowers' propensity to choose fixed rates are likely to rise in anticipation of their expectations being fulfilled. However, simultaneously lenders are likely to raise the price of their fixed rate deals. In Figure 4, two things happen in response to the expected increase in interest rates. The entire demand curve shifts to the right; and the horizontal "supply curve" rises. The overall effect on the number of fixed deals is ambiguous, and depends on the relative magnitudes of these two effects. In order to separate out the effects, an Instrumental Variables (IV) estimator is required.

This leads to the important question of how interest rate expectations are measured. Here we appeal to expectations theory (Sargent, 1972), according to which market expectations of future movements of interest rates may be inferred from the slope of the yield curve: the steeper the yield curve, the more steeply the market expects interest rates to rise. We represent this slope using the difference between the yield on a long-dated bond and a short-dated bond. After much experimentation, we settle for using the 10-year yield minus the 3-month yield. We refer to this difference as the *yield spread*, and denote it as s. By including this variable in the demand equation, we are assuming that all borrowers have interest rate expectations that correspond to the current value of the yield spread. This is the identifying assumption, required in the spirit of Manski (2004), that

enables us to separate out the effect of interest rate expectations from that of preferences. By virtue of this assumption, the inclusion of s as a determinant of the fixing propensity allows us to interpret the effects of all other explanatory variables in terms of preferences (i.e. the risk attitude) of the borrower.

Characteristics of the borrower that that determine the demand for fixed rates will be assumed to determine the position of the demand curve in Figure 4. For example, if age of borrower has a negative effect on the propensity to fix, then an increase in age brings about a leftward shift in the entire demand schedule. Age, as well as a number of other characteristics, do indeed turn out to have important effects.

The variable y^* has been introduced as the *propensity to fix*, and this variable will appear on the left hand side of all specifications. However, y^* is unobserved. The observed counterpart is a variable y taking three possible values: $y = 1$ for variable; $y = 2$ for short-fix; $y = 3$ for long fix. In view of the ordinal nature of this dependent variable, as explained more fully at the end of the current section, we shall estimate all models in the framework of the ordered probit model.

All specifications have the following structure:

$$y^* = \alpha_1 s + \alpha_2(p_f - p_v) + \alpha_3 \left(p_v - r\right) + \alpha_4 r + \beta' x + u \tag{1}$$

where:
   y^* is the propensity to fix;
   s is the 10-year less 3-month yield spread;
   $p_f - p_v$ is the fixed-variable price differential;
   $p_v - r$ is the variable-rate premium;
   r is the central bank base interest rate;
   x is a vector containing characteristics of the borrower;
   u is the error term.

The differences between the various specifications are in the manner in which the variables p_f and p_v, and their counterfactuals, are defined. The most obvi-

ous way in which to impute the counterfactual prices is to follow Phillips and Vanderhoff (1994) by simply using the mean of the prices applying to all of the actual data for a given month. That is: for fixed-rate borrowers, the variable price is assumed to be the mean rate struck by all variable-rate borrowers in the same month; for the variable rate borrowers, the fixed price is assumed to be the mean rate struck by all the fixed-rate borrowers in the same month. The means so obtained are used to construct the price differential $\bar{p}_f - \bar{p}_v$, as well as the variable-rate premium $\bar{p}_v - r$. Note that this notation, with a single bar over each price variable, indicates that the means are used only to replace the counterfactual prices; actual price data is still used for the chosen alternative.

A superior method for imputing counterfactual prices is to follow Brueckner and Follain (1989) by using predictions from an OLS regression of prices (for those who made the relevant choice) on a set of explanatory variables. The price equations estimated take the following form:

$$p_f = \gamma_1 r + \gamma_2 s + \gamma_3 s^2 + \gamma_4 s^3 + \delta' z + v \tag{2}$$

$$p_v = \gamma_1 r + \delta' z + v \tag{3}$$

where:

p_f is the fixed price (interest rate of the fixed rate mortgage);
p_v is the variable price (interest rate of the variable rate mortgage);
r is the Bank of England Base rate;
s is the yield spread;
z is a vector of borrower characteristics that determine the rate offered;
v is the error term.

Note that the fixed rate p_f is assumed to depend (non-linearly) on the yield spread whilst the variable rate p_v does not. This is because, while it is logical to assume that lenders set their fixed rates on the basis of expectations of future movements in interest rates, there is no reason for variable rates to be based on such expectations[4].

---

[4]The variable rates of most banks and building societies tend to shadow the central bank rate.

When (2) and (3) are estimated by OLS, and the resulting predictions used to impute counterfactual prices, the price differential and variable-price premium are denoted respectively by $\hat{p}_f - \hat{p}_v$ and $\hat{p}_v - r$. Note that, as previously with the use of "bars", a single "hat" in these expressions indicates that only the counterfactual values have been replaced by OLS predictions; actual data is used where available.

Yet another possible method for imputation, used previously by Brueckner and Follain (1988), is one that allows for selectivity. As previously remarked, a price that is observed is for a deal that has been chosen, and the deal was chosen because of its attractiveness in terms of features including price. Given this, it is likely that the two methods of imputation described above will result in imputed prices that are, on average, biased downwards. To correct for this "selection bias", Heckman's (1979) two-step method is used to estimate (2) and (3). When estimates thus obtained are used to generate predictions, the resulting price differential and variable-price premium will be denoted respectively by $\tilde{p}_f - \tilde{p}_v$ and $\tilde{p}_v - r$.

One further econometric problem that needs to be addressed is the endogeneity of prices. Because lenders, in the process of setting interest rates, take account of the same factors (in particular interest rate expectations) as do borrowers in the process of choosing whether to fix, an assumption of price endogeneity must be incorporated in order to isolate the effect of such factors on the demand for fixes. For this purpose, an IV estimator is used. IV estimation of (1) involves replacing *all* observations of price, *including the observed prices*, with predictions from the first-stage estimation procedure. When all observations are replaced by monthly means, the variables will be denoted by $(\bar{\bar{p}}_f - \bar{\bar{p}}_v)$ and $(\bar{\bar{p}}_v - r)$. When all observations are replaced by OLS predictions, we will have $(\hat{\hat{p}}_f - \hat{\hat{p}}_v)$ and $(\hat{\hat{p}}_v - r)$. Finally, when all observations are replaced by predictions from the Heckman selection model, we will have $(\tilde{\tilde{p}}_f - \tilde{\tilde{p}}_v)$ and $(\tilde{\tilde{p}}_v - r)$.

Equation (1) will be estimated using six different models, which are summarised in Table 2. Model 6 is the most preferred on theoretical grounds, since it simul-

taneously deals with all three of the econometric problems identified: missing counterfactuals; selection bias; and endogeneity.

Table 2: SUMMARY OF MODEL FEATURES AND NOTATION

Model	$p_f - p_v$	$p_v - r$	Counterfactual	Selection bias	Endogeneity
1	$\bar{p}_f - \bar{p}_v$	$\bar{p}_v - r$	✓		
2	$\bar{\bar{p}}_f - \bar{\bar{p}}_v$	$\bar{\bar{p}}_v - r$	✓		✓
3	$\hat{p}_f - \hat{p}_v$	$\hat{p}_v - r$	✓		
4	$\hat{\hat{p}}_f - \hat{\hat{p}}_v$	$\hat{\hat{p}}_v - r$	✓		✓
5	$\tilde{p}_f - \tilde{p}_v$	$\tilde{p}_v - r$	✓	✓	
6	$\tilde{\tilde{p}}_f - \tilde{\tilde{p}}_v$	$\tilde{\tilde{p}}_v - r$	✓	✓	✓

Key to notation	Description
$\bar{\cdot}$	Sample means replace missing values only
$\bar{\bar{\cdot}}$	Sample means replace all values
$\hat{\cdot}$	OLS prediction replaces missing values only
$\hat{\hat{\cdot}}$	OLS prediction all values
$\tilde{\cdot}$	Heckman prediction replaces missing only
$\tilde{\tilde{\cdot}}$	Heckman prediction replaces all values

Let us finally turn to the dependent variable. As mentioned, this has three possible outcomes: variable; fixed for up to one year; and fixed for over one year. If the variable rate is treated as "no fix", then it is clearly seen that the three outcomes are ordered, and the ordered probit model (McKelvey and Zavoina, 1975) is appropriate. The underlying latent continuous variable is "propensity to fix", denoted as y^* throughout this section. Figure 5 shows the distribution if y^*. Note that the mean propensity to fix, $\beta'x$ say, depends on the explanatory variables appearing in the vector x and that actual propensity to fix is distributed normally (with variance normalised to 1) around this mean. The outcome is assumed to be determined according to where the propensity to fix falls in relation to two "cut-points" κ_1 and κ_2. For example, as indicated in Figure 5, a propensity to fix
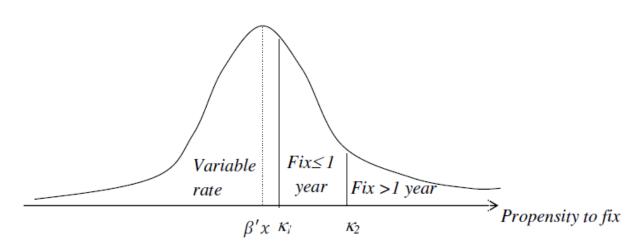
Figure 5: THE ORDERED PROBIT MODEL; DENSITY FUNCTION OF "PROPENSITY TO FIX", AND ITS RELATIONSHIP TO THE OBSERVED FIXING DECISION

between κ_1 and κ_2 would result in a short fix being chosen. The two cut-points are estimated along with the parameters contained in the vector β, by maximum likelihood. Note that the vector β does not include an intercept, since it would not be separately identifiable from the two cut-points.

Another possible approach is the Interval Regression model (Stewart, 1983). Here, the underlying variable is taken to be the desired length of the fix, measured in years. This is illustrated in Figure 6. Desired length of fix is assumed to vary between $-\infty$ to $+\infty$, but negative values are censored and correspond to the choice of a variable rate. If the desired length of fix is positive, we observe the interval (either (0,1) or (1,$+\infty$)) in which the actual length of fix lies. As with the ordered probit model, the underlying variable is assumed to have its mean determined by a set of variables denoted by the vector x, and to be normally distributed around the mean. However, there are a number of differences from ordered probit. Firstly, the variance parameter is identified and is therefore a parameter that requires estimation. Secondly, the intercept in this model is also identified. Thirdly, the cut-points in this model are known values (0 and 1 in this case) and therefore do not require estimation.

A question that inevitably arises is which of these two models is preferred. The answer is that, generally, the interval regression model is preferred on the
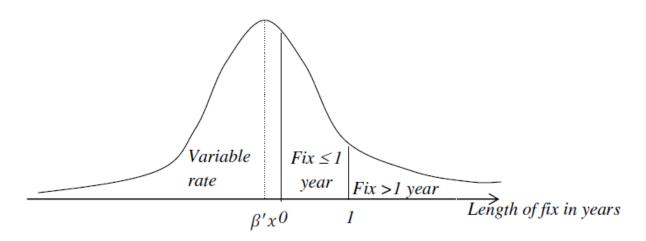
Figure 6: THE INTERVAL REGRESSION MODEL; DENSITY FUNCTION OF LENGTH OF FIX, AND ITS RELATIONSHIP TO THE OBSERVED FIXING DECISION

grounds that knowledge of the cut-points allows greater efficiency with respect to the estimation of the model's other parameters. However, in the special case in which there are only three outcomes, the two models contain exactly the same number of unknown parameters, and, interestingly, they are identical in terms of the value of their maximised log-likelihoods, and therefore their explanatory power[5].

Since the variable of interest here has only three outcomes, we may therefore conclude that ordered probit and interval regressions are equivalent models in explaining our data. Hence, for our demand equation we shall estimate only the ordered probit model, and, accordingly, draw conclusions in terms of the effects of different factors on the borrower's propensity to fix. However, it is still conceptually useful to recognise that an alternative, and effectively equivalent, approach would be to consider the length of the fix as the underlying variable of interest, and to proceed by estimating the interval regression model.

---

[5]Note that, in the trivial case of two outcomes, the ordered probit and interval regression models both become binary probit, and are therefore equivalent, as in the three-outcome case. For any number of outcomes greater than three, the two models are not equivalent, and the interval regression model is preferred.
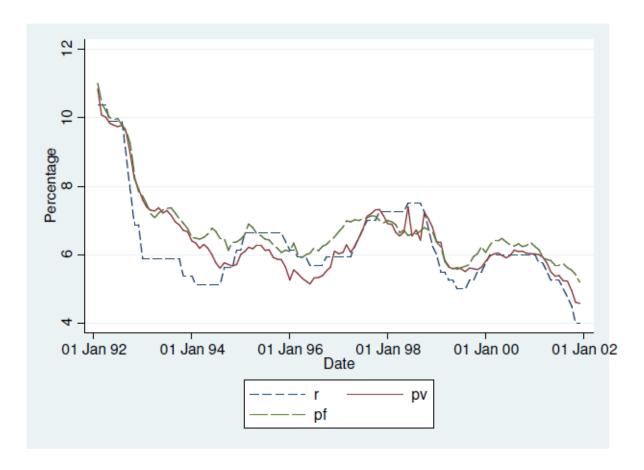
Figure 7: Bank of England Base Rate (r) and Mean Monthly Rates for Fixed (p_f) and Variable (p_v) Mortgages 1992-2001. Source: Fixed and variable rates from the CML dataset; BoE rate from BoE Statistical Interactive Database website: http://www.bankofengland.co.uk

4   Data

We use data from the UK "Survey of Mortgage Lenders" [6], a monthly dataset of approved mortgages by the building societies that are members of the Council of Mortgage Lenders (CML). The CML represents the association of prime lenders in the UK. This means that rates and loan size are competitive and borrowers are comfortably creditworthy (i.e. having low risk of default); in contrast, the non-prime lender sector of the mortgage market specialises in higher risk borrowers and charge a premium rate for mortgages. The survey consists of a random five percent sample of mortgages sold by each lender in the group each month. We

---

[6]This survey was conducted on behalf of the then Dept. of Transport Local Government and the Regions (1978-2001).

use the datasets for 1992 to 2001 giving a total of 438,034 observations; we then further select only those mortgages where full financial details are given leaving a net sample size of 280,654.

Tables A1 and A2 in the Appendix present descriptive statistics for binary explanatory variables, and other explanatory variables, respectively.

Table 3 shows a tabulation of the dependent variable over the sample. We see that the majority of borrowers (62%) chose variable rate mortgages, and that of those who fixed, most chose a "long fix", that is, one lasting longer than one year.

Table 3: Distribution of the Dependent Variable y.

Value of y	Description	Frequency	Percentage
1	Variable rate	172,877	62%
2	Fixed <1 year	11,514	4%
3	Fixed >1 year	96,263	34%
Total		280,654	100%

The explanatory variables are divided into three groups: mortgage price variables; interest rate expectations; and borrower characteristics. Borrower characteristics include income, the number of male and female applicants to the loan, the age of the main borrower, whether income is self-certified, and whether the borrower(s) is (are) first-time-buyers (FTB). Loan characteristics include the size of the advance, the loan-to-income ratio (LIR), the loan-to-value ratio (LVR), and whether the mortgage is of repayment type. Market conditions include the Bank of England (BoE) base rate r, and the term structure variable (s) introduced in Section 3. For the mortgage price there are two variables: $p_f - p_v$ and $p_v - r$. These are derived from the fixed price p_f; the variable price p_v and the BoE base rate r. The time paths of the monthly means of these three variables are presented in Figure 7. As expected, the base rate is usually the lowest of the three. Of the other two, p_f is usually the higher, that is, fixed-rate deals are typically more expensive than variable-rate deals.
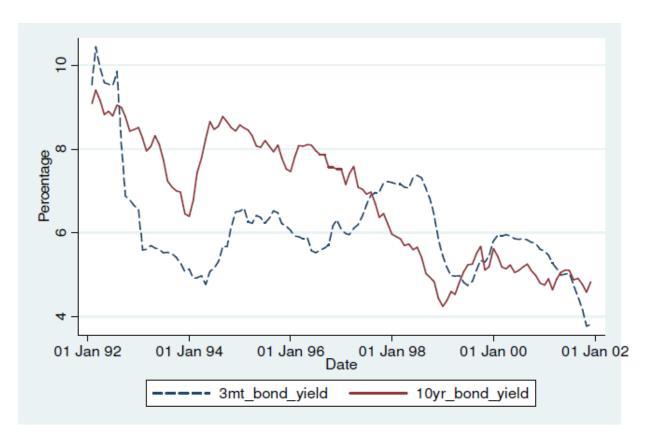
Figure 8: COMPONENTS OF s: THE 3 MONTH BOND YIELD AND 10 YEAR BOND YIELD: 1992 TO 2002.

As discussed in earlier sections, a key determinant of the choice between variable and fixed rates is the expectations of future movements in interest rates. The time paths of the of the two yields that combine to give the yield spread, s, are plotted in Figure 8. The way to use Figure 8 is simply to consider which of the two lines is higher. For example, during the period 1994-1996, the long yield was considerably higher than the short yield, representing the fact that the market was expecting increases in the interest rate. In Figure 7, it is seen that this expectation was correct, with the base rate rising intermittently until mid 1998.

5   Results

The six models outlined in Section 3 have all been estimated using the data described in Section 4. Recall that these six models amount to different approaches to estimating the parameters of equation (1). These estimated parameters are all shown in Table 4. Recall also that some of the models (namely Models 3, 4, 5 and 6) require preliminary estimation of the two price equations (2) and (3), using either OLS or Heckman's 2-step procedure. The estimates from these preliminary models are all presented in Table A3 in the Appendix.

As asserted in Section 4, Model 6 is preferred on theoretical grounds, since it deals with all three of the econometric problems that have been addressed: missing counterfactuals; selection bias; and endogeneity. For this reason, the results from model 6 (i.e. the final column of Table 4) shall be used in interpretation.

Turning first to the price variable on the second sheet for Table 4, we see that the variable $\tilde{p}_f - \tilde{p}_v$ has the expected negative sign, and is strongly significant. This is consistent with the downward-sloping demand curve shown in Figure 4, and simply confirms that a rise in the price of fixes relative to the price of variable deals brings about a fall in the demand for fixes. Next, we see that the coefficient on the variable rate premium ($\tilde{p}_v - r$) is also negative and significant. In interpreting this coefficient, we must imagine that the price differential $\tilde{p}_f - \tilde{p}_v$ is fixed. The interpretation is therefore that if both the variable rate and the fixed rate simultaneously rise by the same amount, there will be a fall in the demand for fixes. Next we see that the coefficient of the base rate r is also negative. This indicates that at times when interest rates, generally, are very high, fixed rates are less popular. This is perhaps because, at times when interest rates are abnormally high, borrowers have particularly strong expectations that they will fall in the future, making fixed deals appear less attractive.

Table 4: ORDERED PROBIT RESULTS FOR MODELS 1 TO 6

Propensity to Fix	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Log of advance	0.449*** (0.0172)	0.422*** (0.0172)	0.470*** (0.0172)	0.494*** (0.0175)	0.460*** (0.0174)	0.506*** (0.0174)
Log of Income	-0.272*** (0.0176)	-0.224*** (0.0177)	-0.289*** (0.0176)	-0.336*** (0.0178)	-0.281*** (0.0178)	-0.338*** (0.0178)
# Males	-0.0313*** (0.00683)	-0.0335*** (0.00687)	-0.0339*** (0.00684)	0.0739*** (0.00730)	-0.0565*** (0.00691)	0.0850*** (0.00723)
# Females	0.0415*** (0.00598)	0.0386*** (0.00601)	0.0373*** (0.00598)	0.0989*** (0.00629)	0.0134* (0.00604)	0.114*** (0.00631)
Age	0.0000893 (0.00144)	0.000183 (0.00145)	-0.000226 (0.00143)	-0.00281 (0.00144)	-0.00130 (0.00145)	-0.00274 (0.00144)
Age^2	-0.0000998*** (0.0000170)	-0.000101*** (0.0000172)	-0.0000966*** (0.0000170)	-0.0000654*** (0.0000171)	-0.0000807*** (0.0000172)	-0.0000663*** (0.0000171)
Self cert'd Income	-0.291*** (0.00492)	-0.283*** (0.00493)	-0.283*** (0.00494)	-0.259*** (0.00494)	-0.306*** (0.00498)	-0.259*** (0.00494)
FTB	-0.00286 (0.00585)	-0.0238*** (0.00587)	0.00242 (0.00586)	0.00981 (0.00589)	0.0204*** (0.00591)	0.00955 (0.00589)
LIR	-0.0733*** (0.00891)	-0.0530*** (0.00895)	-0.0796*** (0.00892)	-0.0862*** (0.00905)	-0.0789*** (0.00901)	-0.0573*** (0.00905)
LVR	0.00873 (0.0141)	0.0154 (0.0142)	-0.00205 (0.0141)	-0.105*** (0.0143)	-0.00580 (0.0143)	-0.0232 (0.0142)
Repayment Mortgage	-0.0548*** (0.00513)	-0.0125* (0.00519)	-0.0633*** (0.00514)	-0.0791*** (0.00517)	-0.0636*** (0.00518)	-0.0852*** (0.00517)
r	-0.0526*** (0.00253)	-0.0909*** (0.00281)	-0.0499*** (0.00257)	-0.193*** (0.00500)	0.00832** (0.00260)	-0.192*** (0.00505)
s	-0.124*** (0.00206)	-0.0287*** (0.00261)	-0.121*** (0.00213)	0.149*** (0.00639)	-0.189*** (0.00214)	0.146*** (0.00648)

Table 4 Continued: ORDERED PROBIT RESULTS FOR MODELS 1 TO 6

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$\bar{p}_v - r$	0.133*** (0.00341)					
$\bar{p}_f - \bar{p}_v$	0.0845*** (0.00311)					
$\bar{\bar{p}}_v - r$		0.0714*** (0.00578)				
$\bar{\bar{p}}_f - \bar{\bar{p}}_v$		-0.659*** (0.0114)				
$\hat{p}_v - r$			0.144*** (0.00375)			
$\hat{p}_f - \hat{p}_v$			0.0782*** (0.00329)			
$\hat{\hat{p}}_v - r$				-0.316*** (0.0171)		
$\hat{\hat{p}}_f - \hat{\hat{p}}_v$				-0.935*** (0.0233)		
$\tilde{p}_v - r$					0.374*** (0.00371)	
$\tilde{p}_f - \tilde{p}_v$					0.362*** (0.00328)	
$\tilde{\tilde{p}}_v - r$						-0.250*** (0.0161)
$\tilde{\tilde{p}}_f - \tilde{\tilde{p}}_v$						-0.941*** (0.0242)
κ_1	1.611*** (0.0618)	1.463*** (0.0628)	1.666*** (0.0621)	0.339*** (0.0754)	1.962*** (0.0624)	0.806*** (0.0649)
κ_2	1.725*** (0.0618)	1.579*** (0.0628)	1.780*** (0.0621)	0.455*** (0.0754)	2.080*** (0.0624)	0.922*** (0.0649)
N	280654	280654	280654	280654	280654	280654
$\chi2^2$	14687.7	20308.6	14784.7	20586.3	25861.1	20478.9
$Degrees\ of\ freedom$	15	15	15	15	15	15
$Log\ likelihood$	-216198.1	-213387.7	-216149.6	-213248.8	-210611.4	-213302.5
$Log\ likelihood\ zero$	-223542.0	-223542.0	-223542.0	-223542.0	-223542.0	-223542.0

Results of ordered probit model of choice between: variable; fixed for up to 1 year; fixed for more than 1 year.

(Standard errors in parentheses). Significance denoted by:* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We see that the coefficient of the yield spread s is positive and significant. This is an expected and a welcome result, since it implies that at times when agents expect interest rates to rise steeply, the demand for fixed-rate deals rises. Recall that the inclusion of s in this demand equation is our chosen method of controlling for interest rate expectations in the demand equation, so that all of the other effects estimated in this equation may be interpreted in terms of effects on risk attitude.

Let us now consider these effects in turn. The log of advance has a positive and significant effect. The size of the advance may be interpreted as a proxy for the level of the pay-offs in the risky choice problem, in which case the positive coefficient leads us to conclude that higher pay-offs bring about an increase in risk aversion. This result is consistent with results obtained in laboratory experiments by, for example by Holt and Laury (2002). More generally, this result adds fuel to the debate over the role of incentives in experiments, discussed, for example, by Camerer and Hogarth (1999).

The log of income has a negative and significant coefficient, suggesting that those with higher income are less risk-averse. We consider this to be a key result because it is filling a gap left by lab-based research. The effect of subjects' income has been keenly pursued in laboratory settings. However, it is not possible to assume, in the lab, that "outside" income is incorporated into the decision-making process; if this is assumed, it becomes impossible to explain the levels of risk-aversion observed, a point made by Rabin (2000). In the context of our natural experiment, however, it is natural to assume that income is embedded in the decision-making process, since the subject is fully aware of the monthly payments that will arise as a result of their choice, and that they will be made directly out of their monthly income. Therefore the strong income effect that we see in Table 4 can be viewed as concrete evidence of the effect of income on risk-aversion.

The negative coefficient of the dummy for "self-certified income" suggests that borrowers who "self-certify" that is, who do not (or perhaps are unable) to provide proof of income, are less risk-averse. Given that self-certifiers tend to be

self-employed, this result is consistent with results from empirical labour economics (see, for example Elkelund *et al.,* 2005 and Masclet *et al.,* 2009) to the effect that risk-seeking individuals tend to select into self-employment.

The number of males and the number of females in the borrowing agreement appear as explanatory variables; both of these variables range from zero to four (see Table A1). It is seen that both have positive coefficients, indicating that the more people in a decision-making unit bring about higher levels of risk aversion. This is broadly consistent with results from work on group (mixed gender) decision-making such as Masclet *et al.,* (2009) and Bateman and Munro (2005). Moreover, the coefficient for the number of females in Table 4 is significantly greater than that of the number of males, indicating that females are more risk averse than males. This is a further development to the existing literature on the role of gender in financial decision making where Jianakoplos and Bernasek (1998) find single women to be more risk averse than couples or single men, while Dwyer *et al.,* (2002) find that such effects are attenuated when specialist knowledge is taken into account. Age of main borrower, and its square, are included. Both have negative coefficients, the latter significantly so. This indicates that older borrowers are less risk averse, or that the accumulation of general life experience (as proxied by age) brings about a reduction in risk aversion. Experience that is more specific to the choice task is captured by the "first-time-buyer" dummy. The positive (although only marginally significant) coefficient on this variable suggests that specific experience, in the form of previous mortgage agreements, also brings about a reduction in risk aversion.

Finally, loan-to-income ratio and loan-to-value ratio each provide a measure of the risk that is being taken by the borrower. The higher these ratios are, the more risk is being taken in terms of the amount borrowed. It is therefore not surprising that both have negative coefficients in model 6. It appears that those who are risk-seeking in their borrowing behaviour are also risk seeking when they come to make the choice between a variable and a fixed rate.

6   Conclusion

The market for mortgages is a highly topical research area at present, not least because it has frequently been identified as the root cause of the global phenomenon that has come to be known as the "credit crunch". In this paper, we have considered one aspect of this market: borrowers' choices over whether and for how long to fix the interest rate. As discussed in Section 2, these decisions, when considered in aggregate, have profound effects, for example by fostering stability in the housing market, and ultimately stability in the economy at large.

In this paper, we have treated the fixing decision as a risky-choice problem in the behavioural economist's sense. We have analysed a large-scale data set containing the fixing decisions of borrowers, in order to estimate the determinants of this decision. The data set has been treated as a natural field experiment, a classification that is justified mainly on the basis that the participating subjects are making their choice in a natural economic environment and are completely unaware that they are participating in an experiment. A significant advantage that this study holds over other natural field experiments on risky choice is that the participating subjects have self-selected only to the extent that they are purchasing a home; they have certainly not self-selected on any basis that relates to their risk-attitude. An advantage of this study over typical laboratory studies on risky choice, is the huge sample size. Many interesting effects that have eluded laboratory researchers have been estimated with very high precision in this study, by virtue of the large sample.

An important theme of the paper has been the need to deal with the various econometric problems that arise as a consequence of the data being collected in the setting of a natural field experiment rather than in that of a laboratory. These problems were: missing counterfactuals; selection bias; and endogeneity. A sequence of models have been estimated that address each of these issues in different combinations. The model that deals with all three problems, and is therefore preferred on theoretical grounds, is Model 6. It was the results of this model that were interpreted in Section 5.

In order to interpret the coefficients of the demand equation in terms of effects on risk attitude, it has been necessary to control for the role of interest rate expectations. The identifying assumption that we have used for this purpose is that all borrowers have interest rate expectations that correspond to the market expectations that can be deduced, via expectations theory, from the prevailing term structure of interest rates. This assumption led to the use of a yield-spread variable s in the demand equation. Reassuringly, this variable had a coefficient which was of the expected sign and is significant.

Most of the results relating to risk attitude appeared to have natural interpretations, and in addition some offered support to results appearing elsewhere in the literature. Moreover, some results have been interpreted as as providing a link between the laboratory and the frontiers of knowledge on risky choice. The most important results in this regard are the postive impact of pay-offs on risk aversion, and the negative effect of income.

7 Appendix

Table A1:
DESCRIPTIVE STATISTICS FOR DUMMY EXPLANATORY VARIABLES

$Year$	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	$Total$
Obs	18,425	23,915	26,633	22,576	30,260	33,419	30,233	29,365	29,737	36,091	280,654
F couple	132	146	145	118	137	144	139	135	171	230	1,497
M couple	110	113	120	89	129	128	131	166	182	290	1,458
H couple	11,495	14,658	16,277	13,476	17,603	18,951	17,095	17,323	17,233	20,362	164,473
Repayment	2,975	5,165	6,705	7,265	11,620	13,136	12,872	13,575	17,809	24,864	115,986
Discount	12,607	14,463	15,295	15,535	17,406	16,175	14,364	17,326	18,791	13,260	155,222
Selfcert'd	6,080	9726	12,108	10480	13,553	15153	13,638	10550	11,267	25897	128452
$\#females$											
0	3,844	4,960	5,478	5,000	7,456	8,808	7,930	6,724	7,014	9,145	66,359
1	14,420	18,775	20,968	17,430	22,632	24,413	22,136	22,469	22,503	26,470	212,216
2	160	177	182	145	169	194	167	170	217	387	1,968
3	1	1	5	1	3	4	0	2	3	70	90
4	0	2	0	0	0	0	0	0	0	19	21
$\#males$											
0	3,039	4,240	4,821	4,051	5,137	5,566	5,126	5,253	5,402	6,305	48,940
1	15,255	19,526	21,656	18,411	24,957	27,674	24,912	23,909	24,097	29,238	229,635
2	130	145	150	112	161	176	186	201	234	452	1,947
3	1	2	5	2	4	3	9	2	3	77	108
4	0	2	1	0	1	0	0	0	1	19	24

Note:

- Figures given are the number of observations where the dummy variable is equal to one.

- "F" "M" and "H" refer to Female, Male and Heterosexual couples respectively.

Table A2.1: DESCRIPTIVE STATISTICS FOR EXPLANATORY VARIABLES

$Year$	1992			1993			1994		
	N	$Mean$	$Stddev$	N	$Mean$	$Stddev$	N	$Mean$	$Stddev$
Advance	18,425	43,617.73	24,652.01	23,915	44,225.23	24,806.30	26,633	46,435.47	27,797.27
Price	18,425	61,208.19	41,001.15	23,915	61,664.61	39,961.04	26,633	63,998.95	43,090.91
Income	18,425	20,800.00	13,157.40	23,915	21,003.17	13,382.15	26,633	21,679.22	14,110.89
Age	18,425	34.22	10.91	23,915	34.67	10.86	26,633	35.12	11.08
r	18,425	9.25	1.18	23,915	5.79	0.19	26,633	5.35	0.31
p_v	13,970	9.53	0.99	11,493	7.12	1.10	15,481	5.88	1.57
$p_f(<1yr)$	1,061	8.87	1.14	2,276	6.66	0.81	1,062	4.59	2.21
$p_f(>1yr)$	3,394	9.58	1.04	10,146	7.32	0.82	10,090	6.72	0.99
LIR	18,425	2.24	0.79	23,915	2.26	0.73	26,633	2.28	0.74
LVR	18,425	0.78	0.22	23,915	0.77	0.22	26,633	0.78	0.22
S Female	2,906	35.26	12.64	4,092	34.52	11.67	4,671	35.42	12.21
S Male	3,733	32.53	10.13	4,846	32.71	9.87	5,352	32.98	10.06
SF FTB	1,807	32.94	12.55	2,700	32.08	11.13	3,100	32.69	11.61
SM FTB	2,488	30.58	9.64	3,314	30.55	8.85	3,895	30.76	8.99
FTB Age	9,469	31.40	10.77	12,775	31.72	10.49	14,578	32.03	10.62

Note:

• Monetary variables: "Advance" "Price" and "Income" are in sterling at current value.

• Variables "SF FTB" (Single Female First Time Buyer), "SM FTB" (Single Male FTB), "S Female" (Single Female) and "S Male (Single Male)" are all dummy variables; the means given for these is the mean of age (in years) for the relevant group.

Table A2.2: DESCRIPTIVE STATISTICS FOR EXPLANATORY VARIABLES

Year	1995			1996			1997		
	N	Mean	Stddev	N	Mean	Stddev	N	Mean	Stddev
Advance	22,576	47,487.00	28,154.57	30,260	51,175.08	31,889.67	33,419	55,307.95	35,695.98
Price	22,576	63,846.62	43,290.60	30,260	70,314.18	49,435.67	33,419	77,056.78	56,159.31
Income	22,576	22,279.62	14,527.35	30,260	24,655.77	17,024.37	33,419	26,222.51	18,424.50
Age	22,576	34.46	10.68	30,260	35.12	10.49	33,419	35.45	10.29
r	22,576	6.57	0.14	30,260	5.87	0.16	33,419	6.64	0.49
p_v	16,607	6.04	1.71	24,137	5.39	1.54	19,643	6.66	1.41
$p_f(<1yr)$	457	5.30	2.12	814	4.68	1.71	667	6.28	1.32
$p_f(>1yr)$	5,512	6.54	1.22	5,309	6.37	1.33	13,109	7.06	0.92
LIR	22,576	2.28	0.73	30,260	2.24	0.77	33,419	2.27	0.83
LVR	22,576	0.80	0.21	30,260	0.79	0.22	33,419	0.78	0.22
S Female	3,933	34.71	11.63	4,998	35.00	11.30	5,420	35.15	11.08
S Male	4,909	32.87	9.80	7,325	33.55	9.43	8,678	33.91	9.15
SF FTB	2,470	31.43	10.19	2,920	31.19	9.66	3,073	31.59	9.73
SM FTB	3,365	30.57	8.65	4,621	30.69	8.15	5,085	30.91	8.06
FTB Age	12,083	30.90	9.48	14,727	30.96	9.04	15,324	31.47	9.17

Note:

• Monetary variables: "Advance" "Price" and "Income" are in sterling at current value.

• Variables "SF FTB" (Single Female First Time Buyer), "SM FTB" (Single Male FTB), "S Female" (Single Female) and "S Male (Single Male)" are all all dummy variables; the means given for these is the mean of age (in years) for the relevant group.

Table A2.3: DESCRIPTIVE STATISTICS FOR EXPLANATORY VARIABLES

Year	1998			1999			2000		
	N	$Mean$	$Stddev$	N	$Mean$	$Stddev$	N	$Mean$	$Stddev$
Advance	30,233	58,387.05	37,648.91	29,365	65,752.61	44,834.98	29,737	71,193.24	50,328.40
Price	30,233	82,603.34	59,827.48	29,365	93,923.49	69,896.46	29,737	102,449.30	78,203.62
Income	30,233	27,404.22	19,229.83	29,365	30,151.89	21,477.78	29,737	31,548.11	22,404.20
Age	30,233	35.44	10.17	29,365	35.89	10.22	29,737	36.40	10.83
r	30,233	7.23	0.34	29,365	5.28	0.25	29,737	5.98	0.06
p_v	12,928	6.80	2.19	17,321	5.69	1.26	18,465	6.01	1.15
$p_f(<1yr)$	372	6.72	1.47	789	5.59	1.19	1,899	5.96	0.67
$p_f(>1yr)$	16,933	6.73	0.77	11,255	5.85	0.89	9,373	6.38	0.85
LIR	30,233	2.28	0.79	29,365	2.33	0.80	29,737	2.38	0.79
LVR	30,233	0.77	0.23	29,365	0.76	0.22	29,737	0.75	0.22
S Female	4,987	35.27	10.87	5,117	35.54	10.61	5,230	36.30	11.37
S Male	7,793	34.06	9.12	6,556	34.10	9.52	6,829	34.86	10.16
SF FTB	2,947	31.94	9.50	2,950	32.12	9.24	2,912	33.14	10.86
SM FTB	4,807	31.48	8.37	4,189	31.70	8.52	4,142	32.48	10.05
FTB Age	14,622	31.80	9.16	13,873	32.29	9.19	13,518	33.12	10.94

Note:

• Monetary variables: "Advance" "Price" and "Income" are in sterling at current value.

• Variables "SF FTB" (Single Female First Time Buyer), "SM FTB" (Single Male FTB), " S Female" (Single Female) and " S Male (Single Male) " are all all dummy variables; the means given for these is the mean of age (in years) for the relevant group.

Table A2.4: DESCRIPTIVE STATISTICS FOR EXPLANATORY VARIABLES

$Year$	2001			Total		
	N	$Mean$	$Stddev$	N	$Mean$	$Stddev$
Advance	36,091	73,311.02	54,273.24	280,654	57,102.21	40,152.75
Price	36,091	116,542.30	88,734.86	280,654	81,806.41	64,193.24
Income	36,091	34,448.12	28,468.19	280,654	26,665.03	20,016.55
Age	36,091	36.04	8.09	280,654	35.38	10.32
r	36,091	5.00	0.62	280,654	6.17	1.16
p_v	22,832	5.31	0.96	172,877	6.28	1.80
$p_f(<1yr)$	2,117	5.47	0.80	11,514	6.05	1.67
$p_f(>1yr)$	11,142	5.76	0.67	96,263	6.66	1.17
LIR	36,091	2.28	0.92	280,654	2.29	0.80
LVR	36,091	0.69	0.26	280,654	0.76	0.23
S Female	6,049	35.72	8.61	47,403	35.33	11.11
S Male	8,823	34.98	7.42	64,844	33.83	9.39
SF FTB	2,771	31.81	8.15	27,650	32.07	10.26
SM FTB	4,110	31.35	7.46	40,016	31.15	8.65
FTB Age	13,197	31.72	8.13	134,166	31.75	9.71

Note:

- Monetary variables: "Advance" "Price" and "Income" are in sterling at current value.

- Variables "SF FTB" (Single Female First Time Buyer), "SM FTB" (Single Male FTB), " S Female" (Single Female) and " S Male (Single Male) " are all all dummy variables; the means given for these is the mean of age (in years) for the relevant group.

Table A3:
OLS AND HECKMAN PRICE EQUATIONS FOR MODELS 3 TO 6

	OLS Fixed	*OLS Variable*	*Heckman Fixed*	*Heckman Variable*
Main				
Log of Advance	-0.106*** (0.00618)	-0.144*** (0.00604)		
Female Couple	0.0122 (0.0739)	0.182** (0.0686)	0.0534 (0.0745)	0.180** (0.0688)
Male Couple	-0.00167 (0.0741)	0.178* (0.0691)	0.0148 (0.0748)	0.144* (0.0693)
Hetero Couple	0.0437 (0.0596)	0.117* (0.0526)	0.0776 (0.0601)	0.0809 (0.0528)
Single Female	-0.0728 (0.0600)	0.129* (0.0531)	-0.0000151 (0.0605)	0.163** (0.0532)
Single Male	-0.0750 (0.0599)	0.000143 (0.0529)	-0.0352 (0.0604)	0.00901 (0.0530)
Loan to Income Ratio	0.00801 (0.00476)	-0.0124* (0.00485)	-0.00107 (0.00460)	-0.0665*** (0.00456)
Loan to Value Ratio	-0.00500 (0.0160)	0.174*** (0.0152)	0.0213 (0.0163)	0.0730*** (0.0151)
Repayment Mortgage	0.00243 (0.00666)	0.0344*** (0.00685)	-0.00789 (0.00674)	0.0245*** (0.00686)
Distmort Mortgage	-0.322*** (0.00639)	-1.091*** (0.00681)	-0.388*** (0.00730)	-1.068*** (0.00734)
r	0.807*** (0.00374)	0.877*** (0.00278)	0.805*** (0.00377)	0.886*** (0.00277)
s	0.251*** (0.00413)		0.244*** (0.00418)	
s^2	-0.0924*** (0.00262)		-0.0887*** (0.00262)	
s^3	0.0270*** (0.000950)		0.0263*** (0.000950)	
Constant	2.927*** (0.0918)	2.852*** (0.0819)	1.536*** (0.0663)	1.387*** (0.0565)

Cont'd

Table A3. Continued: HECKMAN SELECTION EQUATIONS FOR MODELS 5 AND 6

	Fixed	*Variable*
Selection	Fix	Nofix
Female Couple	0.201***	-0.204***
	(0.0549)	(0.0548)
Male Couple	0.126*	-0.132*
	(0.0551)	(0.0551)
Hetero Couple	0.223***	-0.226***
	(0.0432)	(0.0432)
Single Female	0.256***	-0.254***
	(0.0435)	(0.0435)
Single Male	0.157***	-0.158***
	(0.0434)	(0.0434)
Age	-0.00813***	0.00827***
	(0.000279)	(0.000281)
Self Cert'd Mortgage	-0.339***	0.335***
	(0.00500)	(0.00514)
First Time Buyer	0.0324***	-0.0143*
	(0.00584)	(0.00589)
Log of Advance	0.192***	-0.174***
	(0.00491)	(0.00501)
Loan to Income Ratio	0.0533***	-0.0577***
	(0.00367)	(0.00368)
Loan to Value Ratio	0.0730***	-0.0962***
	(0.0140)	(0.0140)
Repayment Mortgage	-0.0633***	0.0612***
	(0.00524)	(0.00525)
Distmort Mortgage	-0.341***	0.342***
	(0.00500)	(0.00501)
r	-0.0689***	0.0701***
	(0.00251)	(0.00251)
s	-0.0881***	0.0896***
	(0.00201)	(0.00200)
Constant	-1.643***	1.465***
	(0.0713)	(0.0720)
		Cont'd

Table A3. Continued:

SUMMARY REPORT FOR OLS AND HECKMAN PRICE EQUATIONS

	OLS Fixed	OLS Variable	Heckman Fixed	Heckman Variable
Rho			0.251***	0.0876***
			(0.0114)	(0.0107)
Sigma			0.0419***	0.318***
			(0.00287)	(0.00178)
N	107777	172877	280654	280654
χ^2			52586.9	122039.0
Degrees of freedom	14	11	13	10
Log likelihood	-155119.1	-299626.3	-332549.0	-477360.7
Log likelihood zero	-177128.2	-346529.1		
R^2	0.335	0.419		

Dependent variable in main equation Interest Rate (fixed or variable as appropriate).
Dependent variable in Heckman Selection equation Fix or No fix (as appropriate).
(Standard errors in parentheses).   Significance denoted by:* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: that the base case for couple-type/ single category in the OLS and Heckman (main and selection stages) models is multiple occupancy households i.e. three or more signatories to the mortgage, the number of observations for these is 3,934.

References

[1] ALM, J. and J. R. FOLLAIN (1984): "Alternative Mortgage Instruments, the Tilt Problem, and Consumer Welfare", *Journal of Financial and Quantitative Analysis*, **19**, 113-116.

[2] BACON, P. M. (2009) "Tenure Choice, Mortgage Choice and Lender Behaviour in the Housing Market of England and Wales", *Unpublished Ph.D Thesis*, UEA, Norwich, UK.

[3] BATEMAN, I. and A. A. MUNRO (2005): "An Experiment on Risky Choice Amongst Households", *Economic Journal*, **115**, C176-C189.

[4] BINSWANGER, H. P. (1980) : "Attitude towards risk: Experimental Measurement in rural India." *American Journal of Agricultural Economics*, **62**, 395-407.

[5] BRUECKNER, J. K. and J. R. FOLLAIN (1989): A.R.Ms and the Demand for Housing", *Regional Science and Urban Economics*, **19**, 163-187.

[6] BRUECKNER, J. K. and J. R. FOLLAIN (1988): "The Rise and Fall of the A.R.M: An Econometric Analysis of Mortgage Choice", *The Review of Economics and Statistics*, **70**, 93-102.

[7] CAMERER, C. and R. HOGARTH (1999): "The effects of financial incentives in experiments: A review and capital-labor-production framework", *Journal of Risk and Uncertainty*, **19** (1), 7-42.

[8] CAMPBELL, J. Y. and J. F. COCCO (2003): "Household Risk Management and Optimal Mortgage Choice", Quarterly *Journal of Economics*, November, 1449-1494.

[9] COULIBALY, B. and G. LI (2009): "Choice of Mortgage Contracts: Evidence from the Survey of Consumer Finances", *Real Estate Economics*, **37**, 659-673.

[10] DEPT. of TRANSPORT, LOCAL, GOVERNMENT, and the REGIONS (2001): "Survey of Mortgage Lenders", *Colchester, Essex: UK Data Archive*.

[11] DHILLON, U. S., J. D. SHILLING, and C. F. SIRMANS (1987): "Choosing between Fixed and Adjustable Rate Mortgages: Note", *Journal of Money Credit and Banking*, **19**, 260-267.

[12] DWYER, P. D., J. H. GILKESON and J. A. LIST (2002): "Gender differences in revealed risk taking: evidence from mutual fund investors", *Economic Letters*, **76**, 151-158.

[13] ELKELUND, JOHANSSON, JARVELIN, and LICHTERMANN (2005): "Self-Employment and Risk Aversion - Evidence from Psychological Test Data", *Journal of Labour Economics*, **12**, 649-659.

[14] HARRISON. G. W. and J. A. LIST (2004) "Field Experiments", *Journal of Economic Literature*, **42**, pp. 1009-1055.

[15] HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, **47**, 153-161.

[16] HOLT, C. A. (1986) "Preference reversals and the independence axiom", *American Economic Review* **76**, 508-514.

[17] HOLT, C. A and S. LAURY (2002): "Risk Aversion and Incentive Effects", *American Economic Review*, **92**, 1644-1655.

[18] JIANAKOPLOS, N. A, and A. BERNASEK (1998): "Are Women More Risk Averse?" *Economic Inquiry*, vol**XXXVI**, 620-630.

[19] LEECE, D. (2000): "Household Choice of Fixed Versus Floating Rate Debt: A Binomial Probit Model with Correction for Classification Error", *Oxford Bulletin of Economics and Statistics*, **62**, 61-82.

[20] MANSKI, C. (2004): "Measuring Expectations," *Econometrica*, **72**, 1329-1376.

[21] MASCLET, D., N. COLOMBIER, L. DENANT-BOEMONT and Y. LOHEAC. (2009): "Group and Individual Risk Preferences: A lottery choice experiment with self-employed and salaried workers", *Journal of Economic Behaviour and Organisation*, **70**, 470-484.

[22] MCKELVEY, R. D. and W. ZAVOINA (1975): "A Statistical Model for the Analysis of Ordinal Level Dependent Variables", *Journal of Mathematical Sociology*, **4**, 103-120.

[23] MILES, D. (2004): "The U.K. Mortgage Market: Taking a Longer-Term View - Final Report and Recommendations", *London: H.M. Treasury*, 1-132.

[24] PHILLIPS, R. A. and J. H. VANDERHOFF (1994): "Alternative Mortgage Instruments, Qualification Constraints and the Demand for Housing: An Empirical Analysis", *AREUEA Journal*, **22**, 453-477.

[25] POST, T. M. VAN-DEN-ASSEM, G. BALTUSSEN, and R. H. THALER (2008): "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show", *American Economic Review*, **98**, 38-71.

[26] RABIN, M. (2000): "Risk Aversion and Expected Utility Theory: A Calibration Theorem" *Econometrica*, **68**, 1281-1292.

[27] SARGENT, T. J. (1972): "Rational Expectations and the Term Structure of Interest Rates", *Journal of Money, Credit and Banking*, **4**,74-97.

[28] SA-AADU, J. and C. F. SIRMANS (1995): "Differentiated Contracts, Heterogeneous Borrowers, and the Mortgage Choice Decision", *Journal of Money Credit and Banking*,**27**, 498-510.

[29] STEWART, M. B. (1983): "On Least Squares Estimation When the Dependent Variable Is Grouped", *Review of Economic Studies*, **50**, 737-753.