Chapter 24

NEW DEVELOPMENTS IN ECONOMETRIC METHODS FOR LABOR MARKET ANALYSIS

ROBERT A. MOFFITT*

Johns Hopkins University

Contents

* The author would like to thank Michael Keane and the participants of a conference at Princeton University in September 1997 for comments, and Julie Hudson for research assistance.

Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card
© 1999 Elsevier Science B.V. All rights reserved.

Abstract

Econometric practice in labor economics has changed over the past 10 years as probit, logit, hazard methods, instrumental variables, and fixed effects models have grown in use and selection bias methods have declined in use. To a large degree these trends reflect an increasing preference for methods which are less restrictive, more robust, and freer in functional form than older methods, although not all trends are consistent with this view. The trends also reflect a tension between structural and reduced-form estimation that has not yet been resolved. A major point of the review is that this trend in labor economic practice has paralleled a trend in econometrics involving the use of flexible forms and semi-parametric and non-parametric methods but has not incorporated the lessons from that field. © 1999 Elsevier Science B.V. All rights reserved.

JEL codes: J00; C1

Labor economists have long regarded their field as the most econometrically sophisticated of the various fields in microeconomics. Many of the major developments in microeconometrics in the 1970s – limited dependent variable and selection bias models, panel data models, hazard analysis, and so on – were often developed with labor economics applications in mind, although certainly not exclusively. Many of the econometric developments of the 1970s were stimulated by the new availability of data from household surveys, both cross-section and panel, which contained information on relatively large numbers of individuals. Many, if not most, of the types of issues that such datasets are best suited for study are found within the scope of labor economics. The development of computational hardware and software in the 1970s also grew rapidly, and this aided the development of many of the econometric methods which required somewhat more computational burdens than ordinarily least squares (OLS). Thus the confluence of econometric developments, data availability, and computational aids all contributed to the rapid advance in methodology in the 1970s in labor economics.

Volumes 1 and 2 of the *Handbook of Labor Economics* (Ashenfelter and Layard, 1986) appeared at a point subsequent to these major econometric developments. While there was no single chapter in that *Handbook* devoted to econometric methods, several of the chapters discussed econometric methods at least briefly as part of a particular topic area, such as those on labor supply (Pencavel, Killingsworth and Heckman), labor demand (Nickell), education (Willis), and hedonics (Rosen). The econometric issues discussed in these chapters were primarily the econometric developments of the 1970s just referred to. As is typically the case, debates in the profession concerning those methods can usually be detected only implicitly in print, for much of the debate over econometric methods is oral rather than written.

This chapter is concerned with econometric methods in labor economics and takes both a retrospective and prospective approach. Retrospectively, the use of econometric methods since 1986 is surveyed and discussed. Over the past decade there have been significant changes in the types of methods used in labor economics which, while mostly known to applied labor economists who have practiced over this period, will be documented here for the first time. In addition, the discussion in this chapter will provide one author's view of why some of those trends have occurred. The chapter will then focus on a subset of the methods used by labor economists, namely, those for qualitative and limited dependent variables (probit, logit, Tobit) and those for selection bias models. The chapter will discuss prospects for the econometric developments in these areas in labor economics over the next 10 years. Perhaps the safest prediction is that methodological views will continue to evolve, for it does not appear to this author that econometric practice is now in equilibrium. Several new econometric developments, such as non-parametric and semi-parametric methods, are discussed and their prospects for increased use are discussed as well.

The results of the survey in the chapter reveal that labor economics is not quite as econometrically sophisticated today as it might be thought. The techniques which have seen the largest increase in usage in the top journals over the 1986–1996 period have been instrumental variables and fixed effects methods, both of which were essentially fully developed considerably prior to 1986. However, there has also been a slight increase in the number of more advanced probit and logit methods used, which may be testimony to lags in the application of econometric methods and/or to their introduction into software packages, for these methods were also developed prior to 1986. The number of applications of frontier econometric methods – simulation methods, non-parametric and semi-parametric methods, and the like – remains exceedingly small. Their use may, of course, be even smaller in other fields of applied economics, but those are not surveyed here.

More generally, the survey reveals that econometric practice in labor economics is shifting toward techniques that are, or at least can be argued to be, less restrictive and more robust than some of those used in the past. Identification of the parameters of econometric models has become a more important focus of attention than it has been historically. This is a trend which would appear to be occurring across many fields of economics, in other social science disciplines, and in fact in the field of statistics itself. It is safe to predict that this trend will continue.

1. What labor economists do

A survey of econometric methods used in labor economics should first determine what labor economists actually do. Table 1 shows the results of a survey of labor economics articles appearing in six highly-ranked general interest journals and two field journals in labor economics in 1985–1987 and then again in 1995–1997 to ascertain trends, using the labor economics classification scheme employed by the *Journal of Economic Literature*.

Interestingly, there has been a decline in the total number of labor economics papers published in those journals surveyed over the period. It is possible that this could be an artifact of the change in the *JEL* classification scheme between the periods but this, if anything, would work in the opposite direction because the scheme in the latter period was more inclusive (e.g., the economics of minorities and discrimination was moved into the

Table 1
Empirical and econometric work in labor economics, 1985–1997[a] (numbers of articles and methods)

	1985–1987	1995–1997
All articles	440	295
With empirical content	278	227
Without empirical content	162	68
Types of datasets used		
Cross-section	113	74
Panel	97	118
Repeated cross-section	40	46
Time series	18	26
Econometric methods used		
OLS	182	132
WLS	5	6
GLS	13	13
NLS	5	3
Probit	33	35
Ordered probit	3	5
Bivariate probit	1	5
Multinomial probit	1	3
Logit	29	23
Multinomial and conditional logit	5	17
Nested logit	1	4
Linear probability model	2	6
Simulation estimation	0	4
Tobit	21	25
Two-step selection bias	19	8
FIML selection bias	5	1
Other selection bias	6	10
IV and 2SLS	36	53
3SLS	4	0
Non-linear 2SLS	2	0
GMM	0	2
Fixed effects	27	56
Random effects	6	9
Hazard	21	25
Non-parametric	1	6
Semi-parametric	0	2

[a] Table surveys all articles appearing in 1985, 1986, and 1987 issues and 1995, 1996, and 1997 issues of the *American Economic Review*, *Econometrica*, *Journal of Human Resources*, *Journal of Labor Economics*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Review of Economics and Statistics* which were listed under the labor economics headings in the classification codes of the *Journal of Economic Literature*.

labor economics category). The large majority of the decline, however, is revealed in Table 1 to have resulted from a decline in the number of theoretical articles published in the field, reflecting a genuine decline in theory in labor economics. This represents a reversal of the trend noted by Stafford (1986) of a marked rise in theory in labor economics from 1965 to the late 1970s and early 1980s.[1] Whether the decline in theory represents a return to the older research style in labor economics, with its institutional and non-theoretical orientation, remains to be seen.

There has also been a slight drop in the total number of articles published with empirical content – defined as having at least one table of non-artificial, non-simulated data – but this could be the result of fluctuations in the numbers of articles published each year, the particular years used in Table 1, and the particular journals chosen.

There have also been shifts over the decade in the types of datasets used in labor economics articles, also displayed in the table. The number using single cross-sections has drastically declined while the number using panel datasets has increased. Although there are few panel datasets available in the later period that were not available in the earlier one – the Survey of Income and Program Participation is an exception – the growing length of panels like the PSID and NLSY, together with the increased appreciation of panel econometric methods, is no doubt responsible for the marked increase in their use. There has also been a slight increase in the number of repeated cross-section datasets used – sometimes called pseudo-panels – primarily resulting from the growth in applications using the Current Population Survey (CPS). There has also been, perhaps surprisingly, an increase in the number of times articles have used time-series datasets, although the absolute number in both periods is far smaller than the numbers for the other dataset types.

The rest of the table shows the number of econometric methods of different types used in labor economics articles in the two periods. Individual articles can contribute more than one entry to the table if they used more than one technique, an issue to which I shall return momentarily.

While least squares is still the workhorse of empirical work in economics – the number of times it is used dominates the others in the table by an order of magnitude – there has nevertheless been a 30% decline in the number of times it has been used. The major sources of this substitution appear later in the table and will be seen to be growth in the use of instrumental variables and fixed effects methods.

Turning to qualitative dependent variable methods, the survey shows that the number of times probit methods of any type have been used has increased slightly, although most of the growth has been in the use of nonstandard variants such as ordered probit, bivariate probit, and multinomial probit. The continued popularity of probit is no doubt partly a result of the large number of labor economics dependent variables which are dichotomous – labor force participation, union status, migration, and educational categorizations, to name just a few – but also the degree to which probit has been incorporated into the major

[1] Manser (1999) detected a decline in theory in labor economics in a survey updating Stafford through 1993, however.

software packages used by applied economists. There has also been a slight increase in the use of logit-related methods, although here the shift away from simple binary logit to more advanced variants such as multinomial, conditional, and nested logit is more pronounced than for probit. While the more advanced logit techniques had been in existence for some time by the mid-1980s, they were still relatively new and their incorporation into econometric practice had not been completed. In addition, once again, these variants were not as widely incorporated into software packages as they are today, which no doubt is an additional contributor to the trend. An alternative hypothesis is that the types of topics which labor economists study have shifted toward types for which these techniques are most appropriate – that is, topics in which multiple discrete outcomes are the object of interest, such as occupational choice – but there is little evidence that there has been any significant shift of this kind.[2]

Table 1 also shows that there has been a slight increase in the use of the linear probability model in labor economics. While the number of times it is used is minuscule compared to what are clearly the more popular techniques of probit and logit, it has grown to a nontrivial number in the later period. This model will be discussed more in the next section and some reasons for this growth will be advanced.

Simulation methods, which are often used for the estimation of large-dimensional discrete choice models, have grown in use considerably over the period and will be discussed in the next section. The use of Tobit analysis has increased slightly but not as dramatically as some of the other methods, and it would be fair to characterize its use as relatively stable. It is a popular technique, used almost as frequently as probit.

Selection bias methods of all types have shown a marked decline over the period. This includes the two-step methods as well as full-information maximum likelihood methods. The reasons for this decline will also be discussed below. Moving in the opposite direction are methods using IV or two-stage least squares (2SLS), which have grown enormously. It will be argued below that these trends are related and are the result of a shift in econometric practice toward methods which require fewer distributional assumptions on unobservables, although it will also be argued that this is to some extent an oversimplification which ignores the less-parametric selection bias methods which have become available in recent years. The growth in IV and 2SLS is still quite remarkable given that those techniques have been widely used in economics for 30 years and had been developed long before that. While it could be argued that recent debates on the use of IV (e.g., Imbens and Angrist, 1994; Bound et al., 1995; Heckman, 1997; Staiger and Stock, 1997) have deepened the profession's understanding of the nature, interpretation, and limitations of IV and 2SLS, very few of the recently-discussed issues had not surfaced before in the econometric literature on these methods. Thus this trend, alone among those in the table, must be largely ascribed to a shift in the preferences of users rather than from the development of new econometric methods, to which the growth of other entries can be ascribed.

The growth in the use of panel datasets mentioned above is necessarily accompanied by

[2] Manser (Table A1, 1999), for example, finds no major shifts in the topics studied in labor economics articles over the past 10 years.

Table 2
Types of fixed effects models used in labor economics, 1985–1997[a] (numbers of times used)

	1985–1987	1995–1997
Individual	10	20
Family	3	11
Cohort	3	2
State	2	9
Geographic, non-state	2	5
Firm	2	3
Industry	1	0
School	0	2
Nationality	0	1
Other	2	1

[a] See Table 1.

a growth in the use of econometric methods for panels. The major growth has been in the use of fixed effects methods, whose use has doubled over the period. Table 2 shows, however, that not all of the growth in the use of this method can be traced to the increased use of panels. While individual fixed effects are indeed the modal category of use, and while models with fixed effects of that type have indeed grown more than those using any other type, the growth of models using family fixed effects and geographic fixed effects (state, city, country, etc.) has been equally dramatic. The resurgence of interest in sibling models, for example (Ashenfelter and Krueger, 1994; Behrman et al., 1994 to cite two examples among many) – models which have a long history in social science research – is part of this trend. The use of state fixed effects has been aided by the growth in the availability of more years of the CPS and of the growth in its sample size which makes estimation of state-specific intercepts more feasible.

The rest of Table 1 shows that the use of hazard methods – also called event-history, transition, or duration methods – has increased slightly, no doubt a combined result of the increase in data availability of panels and of the spread of knowledge and software incorporation of these techniques. Non-parametric and semi-parametric methods have grown significantly over the decade as well but still, in the later period, remain extremely limited in use. A major issue for the future is whether the use of these techniques will grow and become more common and, if so, at what level their use will plateau and what role they will come to play in the toolkit of techniques available to labor economists.

A close reading of Table 1 reveals that many more techniques have grown in usage than have declined. This reflects another trend in econometric practice in labor economics, which has been the growth in the number of multiple techniques used in the typical article. Table 3 shows the distribution of numbers of techniques used in different articles and shows this trend clearly, for the fraction of labor economics articles using only one method

Table 3
Number of different econometric methods used in labor economics articles, 1985–1997[a] (percent distribution)

	1985–1987	1995–1997
1 Method	59	40
2 Methods	26	33
3 Methods	11	16
4 Methods	3	9
5 Methods	1	1
6 Methods	0	0
Total	100	100

[a] See Table 1.

has dropped from almost 60% of all articles in the earlier period to 40% in the latter one. The offset is shown in uniform increased usage of two or more methods.

This last trend reflects a more basic underlying pattern affecting many of the other findings from the survey, of a movement toward the use of less parametric methods, more use of sensitivity testing and multiple methods to test for that sensitivity, and use of robust techniques that are less sensitive to assumptions. Much of applied thinking in labor economics practice today – and in the practice in many other fields inside economics and outside of it – is centered on these sets of issues. A safe prediction to make is that practice is not in steady state and that the trend in this direction will continue, although it would be hazardous to speculate on what exactly where it will be 10 years hence. Nevertheless, this will be the central theme of the rest of the paper, which will discuss trends and developments in qualitative and limited dependent variable models and selection bias models.

2. Developments in qualitative, limited-dependent, and selection bias models

The remainder of the chapter focuses on developments in several of the techniques which have changed in usage and in which thinking has developed considerably over the last 10 years. These are models for qualitative and limited-dependent variables, and selection bias models. The binary choice model will be discussed first and most exhaustively because, despite its simplicity, many of the developments in the other models can be seen most easily and simply when the outcome is dichotomous.

2.1. Binary choice model

2.1.1. Basic considerations

The most popular binary choice framework in labor economics is the probit model, which can be written (suppressing individual-observation subscripts):

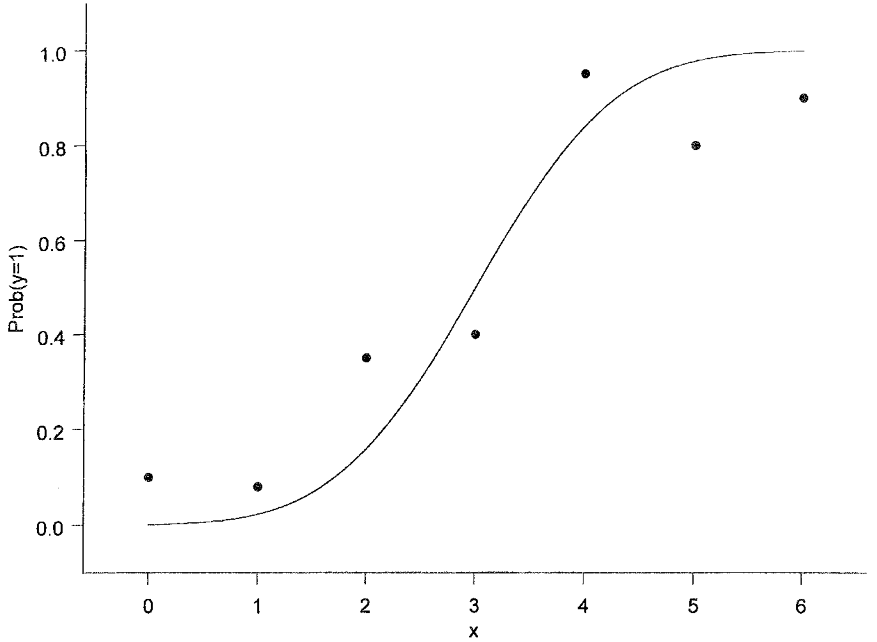

Fig. 1. Probit probability curve with single x with positive coefficient. Dots represent sample means of y.

$$y^* = X\beta + \varepsilon, \tag{1}$$

$$y = 1 \text{ if } y^* \geq 0, \qquad y = 0 \text{ if } y^* < 0, \tag{2}$$

$$\varepsilon \approx N(0, 1), \tag{3}$$

where X is a row vector of variables an observation and β is a column vector of coefficients. The usual normalizations are imposed here – the variance of the error is normalized to 1 and the cutoff point is normalized to zero implying, respectively, that the coefficient estimates are ratios to standard deviations and that the estimated intercept absorbs the true (constant) cutoff. The model implies that

$$\mathrm{Prob}(y = 1 \mid X) = \mathrm{E}(y \mid X) = F(X\beta), \tag{4}$$

where F is the cumulative normal distribution function and is shown as the familiar S-shaped (sigmoid) curve in Fig. 1.

The probit model has a long history – the treatise by Finney (1947) claims that it originated in 1860 – and has been primarily used in the study of bioassay. In the social sciences it has also been extremely popular as a method of estimation of models with dummy dependent variables. While its use in bioassay is based on the theoretical notion that the susceptibility of a group of experimentally-treated subjects to a stimulus has a distribution of responses, the technique is mainly used simply as a curve-fitting method ("...a convenient way...of fitting a function" (Berkson, 1951, p. 327)). Its appeal in economics is based instead on developments in the psychological literature which conceived of human subjects as having a scalar index representing an unobserved propensity, or utility, of an alternative; that differences in the index could be scaled; and that there exists a distribution of those propensities in the population (Thurstone, 1927). This led in turn to the familiar random utility model (Quandt, 1956; McFadden, 1974) as a justification for probit and other binary choice models. More generally, one of its major attractions to economists arises because choices of workers and firms are generally regarded as based on optimizing behavior resulting from utility or profit maximization, and that y^* can be thought of as the difference between the utilities or profits of alternative choices.

The other popular model is the logit model (Berkson, 1944), in which

$$\text{Prob}(y = 1 \mid X) = \frac{e^{X\beta}}{1 + e^{X\beta}}. \tag{5}$$

This model can also be generated by a latent index model such as (1) where the values of each alternative have error terms which are distributed independent extreme value and where y^* again represents the difference in those values and therefore fits just as well into the random utility model as probit (McFadden, 1974). There is an oral tradition in applied microeconomics that probit and logit estimates are almost always close to one another, which is based on the result that the c.d.f.'s of the two normalized distributions are quite close to one another except in the tails (see, e.g., Domenich and McFadden, 1947, p. 58). The oral tradition that the coefficient estimates are not much affected is, however, a somewhat different result and depends on the configuration of the data and the distribution of y and X. Nevertheless, it is based on empirical findings from a large number of applications where both probit and logit have been estimated and hence has a grounding in typical applications.[3]

The linear probability model (LPM) – that is, the model which posits a linear relationship $y = X\gamma + e$ and interprets γ as the effect of a unit change in X on the probability that $y = 1$ – has been fairly uniformly criticized in econometrics textbooks, both older and current. The most cited discussion among the early econometric textbooks is that in Goldberger (1964), who noted that the LPM could produce fitted values of y outside the zero–one range and that e was necessarily heteroskedastic. This criticism appears in the textbook by Theil (1981) and in current textbooks as well. From the current vantage point neither of these disadvantages seems fatal to the LPM because fitted values of y close to the mean of the data are unlikely to lie outside the zero–one range, and thus do not cause a

[3] At one period in the early 1970s it was humorously thought that logit seemed to be preferred by West Coast economists and probit, by East Coast economists. Today the choice between the two is generally thought to be entirely a matter of taste. It was not always so, for in the early development the difference was seriously debated (e.g., Berkson, 1951).

problem if prediction or inference only close to the mean is desired; and because there are a variety of methods now available for the correction of heteroskedasticity.[4]

The reason the LPM has seen a slight return to use is not that anyone believes it should be taken to be the true model – in which case there are many statistical objections to it – but that it can be thought of merely as a linear approximation to the true model, and as a reduced form.

According to the random utility model,

$$y = \mathrm{E}(y \mid X) + u = F(X\beta) + u, \tag{6}$$

where u is a mean-zero, albeit heteroskedastic and bounded, error term. The LPM can be thought of merely as a linear approximation to the non-linear function $F(X\beta)$ and its coefficient estimates are unambiguously and correctly interpretable as such. The coefficients estimated in the LPM are direct estimates of $\partial\mathrm{E}(y \mid X)/\partial X$ and hence do not estimate β but rather some combination of β and F. If one is not particularly interested in estimating the index function coefficient (i.e., β) itself, but only in the net, or reduced-form, effect of a change in X on the probability that $y = 1$, the LPM is arguably an acceptable place to start.[5]

The case for the LPM in this respect must therefore rest partly on whether the question being asked by the analyst requires an estimate of the underlying index coefficient β or just the reduced-form effect of X on $\mathrm{Prob}(y = 1 \mid X)$. The random utility theories which underlie so much of the work in this area make β the object of interest and it is therefore natural to seek estimates of that parameter. A contrary argument can be made that there are not many labor economics hypotheses that depend on the magnitude of β and, if it can be assured that the signs of β and of $\partial\mathrm{E}(y \mid X)/\partial X$ are the same, hypotheses on the sign of the former can be tested from the estimated sign of the latter. If the contrary argument is accepted, the remaining question is whether significant non-linearities are present in the data that are missed by a linear approximation. As with OLS in general, the relevant information in the data used for the coefficient estimation is the set of means of y at each value of x (Fig. 1 includes some scatter points of those means which, according to the model, are generated by $F(X\beta)$ plus a mean of u at each x). The non-linearity of the curve varies over the range of x and the second derivative of the function, which is one measure of non-linearity, reaches an maximum in absolute value at two points on the function; the linear approximation is likely to be the worst in those regions and it is likely to be the best if the data are tightly clustered in either of the two tails or around the 0.50 probability point, where the curve has the lowest curvature. Data configurations which

[4] Fitted values outside the zero–one range do nevertheless cause problems of interpretation of the R-squared and also problems with heteroskedastic fixups such as weighted least squares. Another criticism of the LPM is that it is inconsistent with the random utility model. See Heckman and Snyder (1996) for a discussion of this point.

[5] Amemiya (1981, p. 1487) recommends the LPM in the early stages of an analysis as a convenient way to summarize the data and to gain quick approximate estimates of relationships, but believes that a more formal binary choice model should be used for the final analysis. This is no doubt the way it is used by most applied economists.

have more dispersed sample probabilities over the unit interval are likely to be poorer approximations because more non-linearities are present in that case.[6]

If non-linearities are an issue, a logical alternative procedure is non-parametric (NP) regression (Härdle and Linton, 1994), for NP regression can capture arbitrary non-linearities in a fitted curve. NP regression of y on X in one of its standard forms (e.g., kernel regression or series regression) can be applied without modification even if y is binary, for the method merely fits the means of y in the data at each x, $E(y \mid x)$, to a non-linear curve; that the means of y are fractions is irrelevant. A less drastic solution would be to introduce polynomials, splines, interactions, and other forms of non-linearities in X but within the framework of the basic linear-in-parameters model so that OLS could still be applied. Unfortunately, this approach is not feasible if X is high-dimensional and hence is not a practical alternative.

However, NP regression is still not widely used in labor economics or in other applied areas of economics despite its attractiveness as a method of capturing non-linearities. There is no consensus on why it has not been incorporated in econometric practice more than it has been. NP regression has now been widely known and heavily researched among econometricians (not to mention the wider community of statisticians) for at least 10 years. One simple reason may be that the lags in the incorporation of new econometric techniques into practice is still quite long. A related reason may be that new techniques may have to be incorporated into software packages and that NP regression has not, to date, been incorporated into any of the major packages used by economists.[7] A more problematic reason for the lack of use of NP regression is that estimation can still be somewhat computationally burdensome, especially if there are large numbers of regressors; that large sample sizes are often needed for estimates without major bias, and that rates of convergence might be low; and that the choice of bandwidth has an impact on the estimates but the preferred method of choosing bandwidth has not, to date, been sufficiently standardized through rules of thumb and guides to practice. Unfortunately, the theory of bandwidth selection implies only that it should go to zero asymptotically, as well as satisfying a few other general criteria (Manski, 1991, pp. 43–44), which is not enough of a guide for the practitioner community. A body of practical experience must be built up and widely-agreed upon rules of thumb will have to be adopted instead. It is possible that this will occur in the future and that more use of NP regression will be consequently seen.[8]

[6] See Aldrich and Nelson (1984) for further discussion of using the LPM as a linear approximation to the true curve.

[7] Some packages have non-parametric density estimation, which can be used to estimate NP regressions. Also, LIMDEP permits the estimation of the maximum score model (see below) and some related models.

[8] See the recent article by Blundell and Duncan (1998) for a start in this direction. It should also be noted that some econometricians argue that the semi-parametric models to be discussed momentarily – which permit some forms of non-linearity but still retain some parametric structure – should be considered the viable alternative to NP regression, given the problems with the latter, even for exploring non-linearities and for estimating the reduced-form function $E(y \mid x)$. How to evaluate semi-parametric versus non-parametric models in that light has not been articulated, however. This argument is to be sharply distinguished from the purpose of semi-parametric models discussed in the next section, which is to identify β.

2.1.2. Further issues: heteroskedasticity, identification, and policy experiments

Assuming that one is interested in β itself, and not just in the reduced-form effect of X on y, probit or logit remain the main alternatives currently used by the analyst. Unfortunately, the recent literature on these techniques highlights important restrictions they impose that are often only implicit, and which raise some discomfiting issues that are not easily resolved. One restriction is that the distribution of the unobservables is normal or logistic, and violation of this restriction results in inconsistent estimates of β. Interestingly, however, the literature to date indicates, at least for the binary choice case, that reasonably minor deviations from the normal or logistic – for example, deviations that remain in the class of unimodal distributions – do not much affect the estimates (Manski and Thompson, 1986; Horowitz, 1993b). To obtain a major change in the magnitude of estimated β from this source requires that the error term be distributed very differently, e.g., to be bimodal rather than unimodal (Horowitz, 1993b).

What appears to be a potentially more serious problem is heteroskedasticity, which also (unlike the linear model) results in inconsistent probit and logit coefficients. One approach to the heteroskedasticity problem is to allow for some parametric form of heteroskedasticity, e.g., $\mathrm{Var}(e \mid X) = (X\delta)^2$, and to build this into the likelihood function for probit (or logit). An alternative is to conduct specification tests for such a form of heteroskedasticity or for a more general type; there are a wide variety of such tests available (see Pagan and Vella, 1989; Maddala, 1995, for reviews). Nevertheless, the econometric literature in this area has revealed that a fundamental identification problem lurks if the functional forms are relaxed beyond a certain point. The expected value of y in Eq. (4) relies on two functional form assumptions: (*i*) that F is the normal c.d.f. and that the parameters of the normal distribution do not depend on X, and (ii) that the index is linear in β, as in $X\beta$. These assumptions are generally ignored in linear-model estimation not because they are thought to be true but because OLS coefficients are consistent if they are not (in the case of heteroskedasticity) or are a matter of convenience which could easily be relaxed (in the case of linearity). On the contrary, heteroskedasticity in labor applications could be pervasive, at least to some degree, and non-linearities are often uncovered in studies of the effect of education on earnings, wages on labor supply, and other traditional topics in labor economics.[9]

In the absence of either of these two functional form assumptions, the means of y at each X are given by

$$\mathrm{E}(y \mid X) = F[h(X, \beta); X], \tag{7}$$

where $F(u; X)$ is a proper c.d.f. whose parameters depend on X (i.e., heteroskedasticity is present) and $h(X,\beta)$ is an unknown function which is the main object of interest. It is clear

[9] There is no consensus on whether existing evidence supports serious concern with heteroskedasticity in these models. For an application showing specification tests which reveal no evidence of heteroskedasticity, see Melenberg and Van Soest (1996); for a study reporting specification tests which fail to reject sizable biases from heteroskedasticity, see Horowitz (1993a).

from (7) that one could never hope to disentangle F from h using only the pairs of y-means and X in the data. In Fig. 1, for example, it is impossible to know whether the rough increase in the mean of y as x increases is a result of a true change in the latent index h, or merely a result of a change in the distribution of the error term as x changes. One could fall back on the reduced-form approach, giving up on separating h from F and merely regressing y on X with either OLS or NP regression, but that does not solve the problem; it is still the case that the estimated effect of X on y may merely be picking up heteroskedasticity.

A sizable econometric literature has been built up around this and related issues (see Manski, 1988; Horowitz, 1993b; Powell, 1994, for reviews). One early treatment of the identification problem is addressed by Manski (1988), who termed (7) a "structural" model because it contains an unobserved latent index, as opposed to the reduced-form model $E(y \mid X) = G(X)$ which combines F and h. Estimation of structural models, in general, requires identification restrictions and this model is no exception, and the Manski paper as well as the later reviews discuss a variety of different restrictions that can be imposed on F and/or h to be able to identify β. One can, for example, take a position on the functional form of the index, e.g., $h(X, \beta) = X\beta$ and leave the form of F to be dictated by the data. This is sufficient for identification in what are known as "single index" models, although the form of any heteroskedasticity in those models must be further restricted. The maximum score estimator (Manski, 1975) is also in this class and allows arbitrary heteroskedasticity but at the price of assuming that the median of u is independent of X, and using only the sign of $X\beta$ to determine where y is 1 or 0, thereby using minimum information in the data. The maximum score estimator is one of the few that has been packaged (LIMDEP). Alternatively, one can assume normality (or some other distribution) for F and then let $h(X,\beta)$ be free to be determined by the data which, together with some other restrictions, is also sufficient for identification.[10] None of these alternative approaches is particularly attractive because they convert linearity in $X\beta$ and normality, respectively, from assumptions of convenience to assumptions necessary for identification which are not relaxable (within their class). Assuming linearity of $X\beta$, for example, is tantamount to ascribing all deviations from linearity to F. In addition, the estimation techniques which have been devised for these models have not, as of this writing, been standardized in a way that has made it easy for practitioners to estimate them easily, even as a side test to the robustness of probit or logit. Neither has much empirical experience has been built up upon which standardization could be based. For some techniques, some estimation has been conducted (e.g., for the maximum score estimator) but the empirical experience has not been particularly encouraging. The convergence rates of some of the estimators is also quite slow.

Perhaps the strongest restriction that can be imposed on (7) is simply to assume independence of u and X and therefore to assume that heteroskedasticity is not present. Under

[10] Another branch of this literature which is evolving is one which permits the estimation of a type of function $h(X,\beta)$ which is not fully parametric, as in $X\beta$, but partially parametric such as, for example, representing h as a sum of separable but unknown functions of different variables in X. This permits non-linearities in the effect of X to be estimated without going to the complete non-parametric approach.

that assumption there is still an identification problem because separating F from h requires some restrictions. Matzkin (1992) has provided the most well-known exposition of identification when both F and h are unknown but independence is maintained. Computation is problematic, however, and additional restrictions are required. An alternative is to retreat further and assume either a functional form for F (e.g., normality) or a functional form for h, either of which would simplify matters and make identification easier. Estimation in these cases would still require some non-parametric or related estimation method. One of the few empirical applications of this type of semi-parametric model is that of Newey et al. (1990), who estimated a labor-force-participation equation for women with probit and with two semi-parametric methods which assumed linearity of $X\beta$ and independence of X from the error term, but was non-parametric on the form of F. The authors found that the semi-parametric coefficient estimates were statistically no different than those from the conventional probit model. Still another approach is to assume a functional form for F that is more general and more flexible than the normal or the logistic and thereby move some distance toward the semi-parametric approach. One of the easiest approaches of this type is to assume that the distribution of the error term is a weighted sum of independent normals rather than a single normal (e.g., Geweke and Keane, 1999); this allows the probit model to be a special case but also allows F to take on a wider variety of shapes, for weighted sums of normals can capture many different types of distributions.

The independence assumption is the most attractive one for applications where a policy experiment is the main object of interest, for a genuine policy experiment is by definition one in which a particular X changes over time, or varies cross-sectionally, in a way that is independent of the underlying distributions of unobservables for different values of X. The heteroskedasticity problem, if interpreted as arising from preference heterogeneity correlated with X, for example, is essentially a problem of non-experimental methodology; to avoid it requires changing values of X while holding the types of populations being made subject to each value of X unchanged. This is achieved in a true randomized trial because both experimental and control group error distributions are, on average, the same aside from the effect of the treatment.[11] Indeed, the early literature on the development of probit analysis which took place in bioassay was explicitly experimental in focus, and hence heteroskedasticity was rarely explored.[12]

Where this leaves the practitioner is still somewhat in limbo. If estimation of β itself is desired, then if arbitrary assumptions on distributional form and linearity of the index are to be avoided, difficult and yet-to-be-standardized, or packaged, techniques are still required. There is some evidence that the normality assumption is not damaging as long as the true distribution is unimodal which, if this is maintained, would allow the investigator to simply introduce parametric non-linearities into the index function (still assuming

[11] It is assumed here that (1) is still the true causal model; that is, that there is no treatment effect on any part of the distribution other than the mean. If there is, the model must be modified and different parameters of interest must be introduced.

[12] Finney (1947) urged his readers to read the classic experimental text of R.A. Fisher (1925) before using probit analysis, for example.

independence as well). Moreover, many applied labor economists take the view that it has still not been demonstrated that probit and logit are not, in fact, very robust methods which will generally give approximately correct answers; there is no widespread evidence as yet that contravenes this view. Thus a defensible position at the present time is still to use one of the conventional techniques. On the other hand, if β itself is not of interest, either the LPM or, preferably, non-parametric regression is probably the current or possibly soon-to-be-current standard. These standards could, and probably will, change over the next several years.

2.2. Multinomial choice model

In the multinomial choice model, outcomes consist of multiple discrete categories rather than only two. For example, in a study of occupational choice there will be as many choices as there are occupations. A distinction worth making at the start is that between sets of choices which are mutually exclusive and sets which are not. The occupational choice example is clearly one with mutually exclusive choice, but a different case is the hedonic model in which the individual chooses jobs which have multiple discrete characteristics (pension/no-pension, health insurance/no health insurance, etc.). The latter can be converted into a mutually exclusive set of outcomes by crossing all the individual discrete outcomes with each other and, in so doing, generating a mutually set of combinations of job characteristics, but often this is not desired. This discussion will concentrate on the more common mutually-exclusive case. The non-mutually exclusive case should be thought of more in the class of multiple equation models like the seemingly unrelated regression model.

In the popular multinomial-conditional logit model, individual i must choose from among $j = 1, \ldots, J$ alternatives. Define y_{ij} as a binary variable equal to 1 if the individual chooses j and 0 if not. Then the model proposes that the probability that individual i chooses alternative j is[13]

$$\mathrm{Prob}(y_{ij} = 1 \mid X_i, Z_j) = \frac{\exp(X_i\beta_j + Z_j\delta)}{\sum_{k=1}^{J} \exp(X_i\beta_k + Z_k\delta)}. \tag{8}$$

In Eq. (8), a distinction is made for clarity between variables that vary across individuals, X_i (race, sex, etc.), and variables that vary across alternatives, Z_j (e.g., characteristics of an occupation). In order for the X_i to have a sensible effect on the probability that individual i chooses j, it is necessary that the coefficient on X_i (β_j) vary across alternatives; otherwise, multiplying the top and bottom of (8) by $\exp(X_i\beta)$ would eliminate it from the model and it

[13] Individual subscripts i are added in this section but in no other in the paper. They are shown here because the distinction between X and Z, which is important for identification of the multinomial choice model, is less clear without the individual subscripts.

would not be estimable. Variables which vary across alternatives, like Z_j, on the other hand, can have a constant coefficient (δ).

As in the binary choice model, part of the appeal of the logit model for economists is that it can be derived from a random utility model in which utility maximization is assumed. As shown by McFadden, (8) is the probability that results from a choice problem in which an individual obtains utility from each alternative equal to

$$V_{ij} = X_i\beta_j + Z_j\delta + \varepsilon_{ij} \tag{9}$$

and in which the alternative j with the maximum value of V_{ij} is chosen. The form in (8) requires that the errors ε_{ij} be independently and identically distributed extreme value across alternatives for each individual i. The independence assumption generates the well-known property of independence of irrelevant alternatives (IIA) in the model, namely, that the ratio of the probabilities of choosing any two alternatives is independent of the parameters and the variables (Z_j) for all other alternatives, as can be seen by taking the ratio of two probabilities of the form of (8) for two alternatives j and j'.

The IIA problem and consideration of alternatives that do not require it has dominated the discussion of multinomial choice in applied econometrics in the last 10 years (and, in fact, for some period prior to that). The typical econometric discussion of IIA states that the assumption is violated if some of the alternatives are close substitutes, as would be the case if the individual were choosing between red and blue buses (an example due to McFadden). This is a little misleading because the IIA problem should be thought of more generally as a problem of correlated error terms or, perhaps easier to relate to, as a selection bias problem that arises whenever selecting subsamples of a population leads to inconsistencies in parameter estimates. The latter interpretation makes use of the implication of the IIA assumption that the model can be consistently estimated simply by using for estimation only the individuals in the sample who select one of two of the alternatives, say j and j', and by analyzing their relative choice with binary logit. As should be familiar from the general principles of sample selection bias, estimation on such a subpopulation may yield biased and inconsistent parameter estimates if the subpopulation that chooses only j or j' is systematically different from the rest of the population. The subsample choosing j or j' is a self-selected sample and their relative choices between the two alternatives are likely to be different than the choices that the rest of the population might make. Hence estimates based on the subpopulation will not yield parameters β_j and δ which apply to the total population. The IIA assumption presumes this not to be true; that the rest of the population would make the same relative choices between j and j'.[14]

The underlying issue is how to estimate choices when the ε_{ij} are correlated across j, as one would expect them to be in almost any occupational choice, job choice, or other labor

[14] Indeed, one of the tests for the IIA assumption (Hausman and McFadden, 1984) is based exactly on this formulation, for the test involves comparing the coefficients from the logit estimation on the full sample with the estimates obtained on subsamples like that choosing only two of the alternatives. Under the IIA assumption, the two methods should yield the same coefficient estimates. See Maddala (1995) for a review of this and other specification tests for the IIA assumption.

application where the individuals making the choice have unobserved preferences, or unobserved variables more generally, which are correlated across those alternatives (it would be surprising, in general, if they were not). One approach is to give up on the estimation of the structural model and to seek a reduced form which does not impose any independence on the errors across alternatives. A linearized reduced form might lead to a counterpart to the LPM in the binary choice model, for example. This option is feasible is there are no alternative-specific regressors (Z_j) but may not be if the number of Z_j is large. According to the choice model, an alternative j is chosen if the utility differences between it and other j' are all the same sign, i.e.,

$$V_{ij} - V_{ij'} = X_i(\beta_j - \beta_{j'}) + (Z_j - Z_{j'})\delta + u_{ijj'}, \tag{10}$$

$$u_{ijj'} = \varepsilon_{ij} - \varepsilon_{ij'}, \tag{11}$$

$$\text{Choose } j \text{ iff } V_{ij} - V_{ij'} \geq 0 \qquad \forall j'. \tag{12}$$

Therefore the reduced-form probability of choosing alternative j is

$$\begin{aligned} \mathrm{E}(y_{ij} \mid X_i, Z_1, \dots, Z_J) &= \mathrm{Prob}(y_{ij} = 1 \mid X_i, Z_1, \dots, Z_j) \\ &= \mathrm{Prob}(u_{ij1} > W_{ij1}, \dots, u_{ij,j-1} > W_{ij,j-1}, u_{ij,j+1} > W_{ij,j+1}, \dots, u_{ijJ} > W_{ijJ}) \\ &= g(X_i, Z_1, \dots, Z_J), \end{aligned} \tag{13}$$

where

$$W_{ijj'} = -X_i(\beta_j - \beta_{j'}) - (Z_j - Z_{j'})\delta. \tag{14}$$

Thus the reduced form for the choice of j must contain as arguments not only X_i and Z_j but also the $Z_{j'}$ for all other j'. That is, the probability of choosing j is a function of all characteristics of all alternatives. If the application involves a large number of alternatives, or if there are very many variables Z_j for each alternative, this yields a model with an impractical number of independent variables. Moreover, the large number of coefficients obtained from estimating such long regressions separately for every alternative would be an inefficient method of estimating the model compared to any alternative that recognizes that there is a much smaller number of underlying structural coefficients which are determining all the reduced form coefficients.

Nevertheless, if there are no Z variables in the application or if the number of Z variables or alternatives is small, the reduced-form approach is quite feasible. A LPM which projects y_{ij} onto all X_i and all Z_j, or an NP regression which does the same but better captures the non-linearities involved, are interpretable approaches.[15] There is also an

[15] Independence of the distribution of the errors from X is assumed throughout. As in the binary choice model, for example, heteroskedasticity would adversely affect any of these reduced-form approaches as well as structural approaches.

approach called "universal" logit which assumes (13) to have a logit form in which X_i and all Z_j enter the model.[16] In this case the logistic assumption is arbitrary and does not follow from the underlying error structure of the ε_{ij}, but is simply a way of capturing non-linearities and keeping the dependent variable within the unit interval. In all these approaches, because reduced-forms are estimated, the random-utility interpretation is lost and none of the estimated coefficients can be directly related to those in (9). Also, as in the binary choice model, the value of these approaches depends on whether direct knowledge of the parameters of (9) is of interest rather than reduced form estimates of $\partial \mathrm{E}(y \mid X, Z)/\partial X$ and $\partial \mathrm{E}(y \mid X, Z)/\partial Z$.

Other approaches to the IIA problem retain the object of interest as estimating the parameters of (9) and hence are structural to some degree. The nested multinomial and generalized extreme value (GEV) models (McFadden, 1981), which have increased somewhat in popularity as noted in Section I, are of this type. In these models it is necessary to be able to assign the alternatives to a tree, or sequential, structure in which some of the alternatives are chosen after (in a temporal sense) or independently of (in a more general sense) some of the other alternatives. One application of this approach is to assume that a woman first chooses whether to work and only then whether to work part-time or full-time; another is that an unmarried woman with a child first decides whether to marry and only then, if she does not, whether she will go onto welfare. The nested logit and GEV models permit a degree of correlation between the error terms of the equations for the value of the lower-level alternatives, while maintaining independence from the upper levels. Unfortunately, as the two examples just given illustrate, the behavioral assumptions involved are strong and may be untenable. Most women undoubtedly jointly choose whether to work or not, and whether to work part-time or full-time; the choices are not sequential or separable. Nevertheless, these models have a role to play at least as a specification test for the fully independent model and are often worth estimating (the nested logit model is available in software packages)

An alternative approach that has undergone additional discussion in the last several years is multinomial probit with correlated errors. In this model, (9) is assumed to be the correct specification of utility for each alternative but the ε_{ij} across j are assumed to be distributed multivariate normal with a relatively full covariance structure (i.e., with non-zero correlations between the ε_{ij}). The probability of choosing each alternative is again in the form of (13) but now the aim is to actually evaluate that probability under the assumption that the underlying errors are multivariate normal. The problem in this case is entirely a computational, or numerical, one, for evaluation of high-order multivariate normal probabilities was long considered computationally infeasible even with modern hardware. However, Lerman and Manski (1981) showed that such probabilities could nevertheless be numerically computed by means of Monte Carlo simulation methods in which random draws from a multivariate normal distribution are repeatedly taken to form an estimate of the probability in question. Later work by McFadden (1989) and Pakes and Pollard (1989)

[16] See Amemiya (1985, p. 307) for a discussion of this model.

demonstrated the consistency and other properties of this and related estimators. A variety of alternatives have developed – methods of moment and maximum likelihood simulation methods – and a sizable literature has grown up around them. Several surveys are now available which outline the various approaches to estimation that have been developed (Hajivassiliou, 1993; Keane, 1993; Hajivassiliou and Ruud, 1994; Stern, 1997) and the methods have been extended to panel data (Keane, 1994).

To date these techniques have not been utilized to the extent that their potential would allow.[17] Multinomial logit is still by far the norm in estimation of multinomial choice models. A major reason for this lack of use is probably entirely practical, namely, that simulation methods have not been incorporated into software packages or standardized sufficiently to allow their routine use by applied economists. While writing a program to conduct the necessary numerical computations is in principle not difficult, it is sufficiently time-consuming as to be beyond the time capacities for most applied work. When and if the software firms incorporate these simulation techniques into their products will probably largely determine when and whether these techniques will spread in use.

A second and possibly more serious problem that has received some attention in the econometric literature is the identification of the multinomial probit and other models with correlated errors and, in particular, the identification of the across-alternative correlation coefficients that are at the heart of the contribution of multinomial probit over multinomial logit.

In the linear model, cross-equation correlation coefficients and covariances can usually be estimated from the sample covariance of residuals across equations, but in this case no similar approach would be feasible because the covariance of y_{ij} and $y_{ij'}$ is identifiably zero for all pairs – an individual is observed to choose only one alternative by definition. Therefore it is difficult to see how one could ever estimate a correlation in unobserved tastes (for example) between alternatives. The cross-equation correlations must instead therefore be identified from the conditional mean function g for each choice shown in (13), which relate the choice of each alternative to the X_i and the $Z_{j'}$ for all alternatives j'. The functional form in which the $Z_{j'}$ enter the g functions will differ depending on whether the $\varepsilon_{ij'}$ are or are not independent across all j', and it is this difference that must furnish identification. In the case where there are no Z_j at all, and hence each g function in (13) is simply a non-linear function of X_i, it is clear that no correlation coefficients could be identified in a completely distribution-free specification of the ε_{ij} and that all identification would come only from the non-linearities inherent in the multivariate normal distribution. Identification seems more possible if Z_j exist because the relation between y_{ij} and the $Z_{j'}$ for other alternatives should provide some information on the correlation. One consequence of this problem is that estimates of the multinomial probit model appear to be quite sensitive to the existence, and choice, of alternative-specific variables, as demonstrated by Keane (1992) and Geweke et al. (1994). This problem has not been completely worked-out in the econometric literature.

[17] For two examples of labor economics applications to date, see Berkovec and Stern (1991) and Keane and Moffitt (1998).

The non-parametric and semi-parametric literature has also not yet addressed multinomial models in depth. An extension of the maximum score model mentioned previously to the multinomial case has been proposed but has been little used in practice (Manski, 1975). The identification and estimation problems that arise when the normality or homoskedasticity assumptions are dropped have also not been extended to the multinomial model yet as well (see Horowitz, 1993b, and Powell, 1994, for references). Development of practice in labor economics in this direction must therefore await more progress in the econometric literature.

2.3. Censored regression model (Tobit)

The survey in Section 1 revealed that the Tobit model has retained its popularity in labor economics for the last decade. The model in its simplest form can be stated as

$$y^* = X\beta + \varepsilon, \tag{15}$$

$$y = y^* \text{ if } y^* \geq 0, \qquad y = 0 \text{ if } y^* < 0, \tag{16}$$

with $\varepsilon \sim N(0, \sigma^2)$. One of the most common uses of the Tobit model is in the analysis of the labor supply of married women, of whom a significant fraction do not work at any given point in time.

The general popularity of the model in labor economics, as well as in economics in general, is, as in the binary choice model, traceable to its easy interpretation in terms of individual and firm choice, where y^* represents either the demand or supply of a good, which will equal zero at a theoretically well-defined corner solution, or y^* is simply some other continuous choice which includes the option of not engaging in the consumption or activity at all.[18]

Despite the popularity of the Tobit model, the econometric literature on the model (there usually called the censored regression model) as it has developed over the last decade has revealed its fragility in the face of its assumptions. One source of fragility is the assumption of homoskedasticity which, as in the probit model, is necessary for consistent estimation (Hurd, 1979; Arabmazar and Schmidt, 1982; Brown and Moffitt, 1983). Monte Carlo evidence suggests that the asymptotic bias can be quite large. A second source of difficulty is the distributional assumption of normality for the unobservables, a problem examined explicitly by Arabmazar and Schmidt (1982) and Goldberger (1983). These papers show that the use of different distributions than the normal can yield quite different coefficients.

While both of these problems were present in the probit and logit models, they are potentially more severe in Tobit. In the two binary choice models, the dependent variable has a limited range from zero to one and there is considerable evidence that the S-shaped curves followed by a moderately wide range of different distributions are not far different, at least in the implications for the coefficient vector on a latent index. In the Tobit model,

[18] The Tobit model is due to Tobin (1958). See the September–October 1997 issue of the *Journal of Applied Econometrics* for a series of papers replicating and extending the demand function studied by Tobin.

on the other hand, the availability of continuous data on y leads, paradoxically perhaps (given that more information in the data must be regarded as better than less information), to a greater susceptibility to misspecification. The requirement that that conditional distribution of y, for those with positive y, be truncated normal and the same across individuals is a strong assumption that is commonly violated in many applications.

Although the labor economics literature has not yet absorbed the lessons of these results, the potential failure of the normality assumption in Tobit does have a counterpart in the labor supply literature in one area, which is the long-standing recognition that the distribution of hours of work per week and over longer periods as well, is highly clustered and decidedly non-normal (e.g., Pencavel, 1986); and that the determinants of the decision to work may be different than those for the choice of hours conditional on working. Hours per week are strongly clustered around 35–40, for example, and attempts to fit the conditional hours distribution to a truncated normal results in a poor fit to the fraction of those working zero hours; it is difficult to use the normal distribution to fit both. Fixed costs of work, which have been incorporated into econometric models of labor supply for some time (Hausman, 1980; Cogan, 1981) imply that the marginal labor supply function is not the same as the function describing work choice (see Heckman, 1993, for a discussion of this issue in the context of a review of the labor supply literature). Several articles in the labor supply literature have tested the Tobit model for hours of work and have rejected it, not only for men, whose hours are especially clustered, but also for women (Moffitt, 1984; Mroz, 1987).

In the econometric literature, this issue has been partly reflected in discussions of what is known as the Cragg model (Cragg, 1971), which separates the model for y for those with positive y from the model for whether y is positive. The Cragg model is properly considered to be a multiple-equation selection bias model rather than a censored regression model, although the distinction is not important for anything other than nomenclature. In the labor supply literature, the estimation of conditional hours worked functions reflects this same type of model.

Assuming that the object of interest is still the Tobit model, and not a selection bias model – that is, that the model of interest is a single-equation model – we may ask, once again, what the alternatives are to Tobit and how some of these issues may be addressed. As in the models discussed thus far, provided the problem is with the distributional assumption on the error and not with heteroskedasticity or other failure of independence of X and the error term, one solution is to give up on the estimation of β and seek only estimates of $\partial \mathrm{E}(y \mid X)/\partial X$. In the censored regression model in general, without the normality assumption on ε (but still maintaining independence of ε),

$$\begin{aligned} \mathrm{E}(y \mid X) &= \mathrm{Prob}(y > 0 \mid X)\mathrm{E}(y \mid X, y > 0) = [1 - F(-X\beta)]X\beta + \int_{-X\beta}^{\infty} \varepsilon f(\varepsilon \mid X) d\varepsilon \\ &= g(X), \end{aligned} \tag{17}$$

where F and f are the unknown c.d.f. and p.d.f of ε, respectively. A least-squares projection of y onto X yields a linear approximation to the non-linear curve represented in (17). As before, a NP regression is likely to do a better job in picking up the non-linearities in the curve than least squares.

If β is the object of interest, which it often will be – perhaps more so than in the binary choice case, for here it seems more likely that the continuous sample of the data should allow identification of the index function – a wide variety of econometric methods have been proposed but none has been applied in more than a handful of articles to date, and rarely in labor economics (a recent exception is Chay and Honore, 1998). These include the least absolute deviations estimator (Powell, 1984), the quantile restriction estimator (Powell, 1986a), and the symmetrically trimmed estimator (Powell, 1986b), none of which requires full independence of the error term and hence can accommodate heteroskedasticity. Maintaining independence has led to proposals for a wide variety of additional estimators (Honoré and Powell, 1994 and others). As with the other semi-parametric estimators that have been discussed, there has been insufficient practical experience with these estimators in labor economics for standardized practice to have built up or for very much information to have been gathered on their impact on coefficient estimates in typical applications. Nor have the estimators been incorporated into the major software packages. Given the potential importance of the breakdown of assumptions in the Tobit model, more work in this direction would seem particularly warranted.

As with most semi-parametric estimators in general, most often a parametric assumption on the form of the index function is a maintained assumption. The usual assumption is the standard linear model form $X\beta$, and hence estimates of the features of unknown distributions are implicitly and partly based on deviations from linearity. Little work has been done as well in investigating the interactions of relaxing the linearity of this function with the relaxation of distributional and other assumptions just referred to, either in terms of feasibility, properties, or typical practical performance. As with the binary choice model, there have to be limits to the extent to which linearity can be reduced if independence of the errors is not maintained, because identification of the model can fail completely under a non-parametric specification for the index function combined with arbitrary forms of heteroskedasticity. Unfortunately, the presence of a subset of continuous observations with $y > 0$ does not alter this fundamental problem that also arises in the binary choice model.

2.4. Sample selection bias model

The traditional selection bias model in econometrics began with the work of Heckman (1974) on wages and labor supply and was developed, expanded, and elaborated further in a series of papers in the late 1970s by Heckman (1978, 1979), Lee (1979), and others. The literature has two distinct branches, one of which concerns estimation of equations which are observed for only a subsample, either by definition – as in the case of wage rates, which are by definition observed only for those working – or by fortune of data available. This model could be termed the "partial-population" sample selection model but will here be termed the "sample selection" model for simplicity. The other branch presumes that the total population is available in the data but that there are one or more regressors of interest which take on their values as a result of some type of selection process. The canonical case assumes interest to center on a single dummy variable for some type of treatment and hence these models are often termed "treatment-effect" models. Barnow et al. (1980) drew an analogy between the sample selection model and the treatment-effect model, and proposed estimation techniques for the latter that were based on those developed for the former. But since that time it has become understood that the treatment-effects model admits of a much larger class of estimators, many of which are not applicable to the sample selection model – IV is perhaps the leading case. While there is still some relationship between the two types of models, the literatures have sufficiently diverged that the discussion here will, for space reasons, be restricted entirely to the sample selection model.[19]

The canonical sample selection model can be written as

$$y = X\beta + \varepsilon, \qquad y \text{ observed if } I = 1, \tag{19}$$

$$I^* = Z\delta + v, \tag{20}$$

$$I = 1 \text{ if } I^* \geq 0, \qquad I = 0 \text{ if } I^* < 0, \tag{21}$$

and with the assumption that ε and v are distributed bivariate normal with means zero, variances σ^2 and 1, respectively, and with correlation ρ. The variables I and Z are assumed to be available for the total population. In the sample with observed y, the conditional mean of y is equal to

$$\mathrm{E}(y \mid X, I = 1) = X\beta + \mathrm{E}(\varepsilon \mid X, I = 1) = X\beta + \theta\lambda(Z\delta), \tag{22}$$

where $\theta = \sigma\rho$ and $\lambda(Z\delta) = f(Z\delta)/F(Z\delta)$ is the inverse Mills ratio, and where f and F are the unit normal p.d.f and c.d.f., respectively. Given the result in (22), consistent estimates of β can be obtained either by estimating the two equations in (19)–(21) by maximum likelihood, by a two- step procedure in which probit estimates of (20)–(21) are used to estimate (22) is estimated by least squares (or WLS) using estimates of $Z\delta$ from the first stage, or by a variety of other methods.

Empirical practice in labor economics has seen a decline in the use of these methods, as noted in Section I. This decline in use has a variety of rationales. One is that $\lambda(Z\delta)$ is often

[19] For later developments of the treatment-effects model see Heckman and Robb (1985), Imbens and Angrist (1994), and Manski (1994). The literature is too large to cite many of the developments. The importance of the distinction between the two types of models depends heavily on whether the treatment effect is homogeneous, which is itself related to whether different groups have different equations with separate unobservables. If completely separate equations are specified for the two treatment groups, the model comes closer to the sample selection model. It has also been shown that if the conventional treatment effect coefficient is assumed to be random and a function of the same X variables that appear in the outcome equation, the treatment-effect model separates into the Lee (1979) switching regression model where there are two subpopulations with completely different parameters, which is the same as two separate sample selection models (Björklund and Moffitt, 1987).

highly collinear with X and hence estimates of β tend to be unstable, non-robust, and sensitive to minor changes in the specification of the X and Z vectors. Monte Carlo results of Nelson (1984) show that the standard errors of the elements of β can indeed be very large if the degree of collinearity is high. Other Monte Carlo results show that the inverse Mills ratio is close to linearity over middle ranges of selection probabilities, and exhibits non-linearities that would reduce collinearity with $X\beta$ only in the tails (Leung and Yu, 1996). The argument is usually made for estimation by maximum likelihood as well even though it is more efficient than the two-step method under the model assumptions. A second rationale often mentioned is that the distributional assumption of bivariate normality is unwarranted and may be false and, especially if X and Z coincide, identification of the model is made on the basis of an arbitrary distributional assumption. A third argument often given is that adjustment for sample selection bias does not matter in any case. This rationale is partly in conflict with the first two, for if either collinearity is high or the normality distribution is false, the estimates from the model are not capable of leading to a conclusion one way or the other on the importance of selection bias.[20]

The first two issues are related and have been addressed by the developing semi-parametric literature on sample selection models (Powell, 1994; Vella, 1998). This literature has shown that the bivariate normality assumption can be greatly weakened. A simple relaxation that is partially apparent from (22) already is that all that is really needed for the two-step method is that ν be normally distributed and that ε be linearly related to ν; normality of ε and bivariate normality between the two is not needed. More important, it is clear from (22) that even the normality of ν can be relaxed as long as the joint distribution of ε and ν is independent of X and Z, for in that case the conditional mean of ε depends only on the index function $Z\delta$ which, in turn, depends only on $\text{Prob}(I = 1 \mid Z) = F(Z\delta)$. That is

$$\begin{aligned} \mathrm{E}(y \mid X, I = 1) &= X\beta + \mathrm{E}(\varepsilon \mid X, I = 1) = X\beta + \mathrm{E}(\varepsilon \mid \nu > -Z\delta) = X\beta + h(Z\delta) \\ &= X\beta + h'(p), \end{aligned} \tag{23}$$

where h and h' are unknown functions and $p = \text{Prob}(I = 1 \mid Z\delta)$. Eq. (23) makes no significant distributional assumption on ε and ν (aside from the usual independence assumption from X and Z) and hence can be used as a basis for estimation under relaxed assumptions. Under the same approach as other two-step methods, (23) shows that a first-stage estimate of $Z\delta$ by itself, or of the probability that $I = 1$, if obtainable, can be entered into the equation and used to control for selection bias provided the unknown functions h and h' can be estimated as well.

Approaches along these lines have been elaborated by Gallant and Nychka (1987), Robinson (1988), Choi (1990), Ahn and Powell (1993), and Newey (1988), among

[20] Some of the literature on the robustness of sample selection models is in the treatment-effects literature instead, particularly in the study of the effects of unions and the effects of training programs. Both the nature of the problem and the solution are quite different than in the sample selection model, although there is a similarity in one method of identification (exclusion restrictions) referred to below.

many others. These articles propose that first-stage equations for the probability that $I = 1$ be obtained from semi-parametric or non-parametric methods, thereby reducing or eliminating the parametric assumptions on (20)–(21); that either the estimates of $Z\delta$ or p from the first stage be entered into the second stage and some type of semi-parametric method (kernels, pairwise differences, series estimation, etc.) be used to account for the unknown function h or h' in the estimation. A somewhat older approach that represents a halfway house between these semi-parametric methods and the conventional parametric, normal model, is one which frees up the bivariate distribution to allow it to be of a form that can capture more types of distribution shapes that the bivariate normal but still maintain a parametric form (e.g., Mroz and Guilkey, 1995). These models are relatively easy to estimate.[21]

In the absence of distributional assumptions, identification of the model requires an exclusion restriction, for it should be clear from (23) that if X and Z coincide, β could not be separated from h or h'.[22] The source of the collinearity problems that are often experienced in the application of the parametric, normal-based sample selection model are largely the result of either no exclusion restrictions or exclusion restrictions that are weak.

To date the new methods have been very little used and hence their potential in addressing the difficulties associated with the sample selection model have yet to be assessed. One exception is Newey et al. (1990) who applied one version of the semi-parametric method to the classic wage-labor-supply model of Heckman (1974). Interestingly, they found that selection bias adjustment made little difference to estimation of coefficients of the wage equation and that normality could not be rejected. This article may be the source of the view, noted earlier, that selection bias adjustments make little difference. However, much more empirical experience is needed to determine whether this result applies to other groups and datasets, and whether it applies to the enormous range of areas other than the wages of workers where sample selection issues arise before any general conclusion can be reached.[23]

The few applications that have been thus far reported continue to fail to emphasize or explore in depth the issue of exclusion restrictions for identification which come to the fore when distributional assumptions are relaxed. This has also been a problem in past applications of the sample selection model, where exclusion restrictions have been given little attention and have been treated quite casually. The econometric literature has not dealt in great detail with this issue because it is not intrinsically an econometric problem but rather

[21] The Mroz–Guilkey approach is closely related, in turn, to an approach of Heckman and Singer (1984) which was originally applied to hazard models but which is applicable to sample selection models as well. In all these approaches the bivariate distribution of two error terms is assumed to be composed of one error which is discrete multinomial and another is continuous. An issue of theoretical, but as yet unclear practical, importance in this literature is whether these distributions are viewed as the true distributions or only as approximations to the true distributions. The asymptotic distribution of the parameter estimates differs depending on which view is taken.

[22] The intercept cannot be identified in any case under most of these methods but can be estimated by extrapolation, assuming it is of interest.

[23] Vella (1998) presents an example where, unlike Newey et al. (1990), he argues that sample selection adjustments to a wage equation do make a substantive difference.

an economic and empirical problem of finding variables that plausibly affect selection but do not affect y directly. In this respect identification of the sample selection model turns out to have a close affinity to identification in the treatment-effects model, despite the differences in structure of the models noted earlier. In the typical consideration of IV in treatment-effects estimation, the search for instruments which are both (i) relevant in the sense of having a strong asymptotic correlation with the endogenous variables holding constant all the other exogenous variables and (ii) which are exogenous have exact parallels in the sample selection model in the search for exclusion restrictions (Z) which are strongly related to the endogenous variable I holding constant X and which are exogenous (independent of ε).[24] In short, then, the solution to identification in the sample selection model, at least if approached through the use of exclusion restrictions, is no more or less difficult than the conventional identification problem through exclusion restrictions that has preoccupied economists since 2SLS was developed and which continues to be a key source of attention in empirical work aiming at the estimation of causal effects. Both in that general literature and in the sample selection literature, an important lesson from much of the empirical work in the last decade is that exclusion restrictions and, more generally, identification cannot be treated cavalierly or in a mechanical fashion; any method which is applied by rote is likely to lead to unsatisfactory results. This is the lesson of the new literature on non-parametric and semi-parametric estimation as well, and it implies that the role of exclusion restrictions should occupy a much more central role in the estimation of sample selection models just as it has come to occupy that role in the treatment-effects literature.

Even given these generalizations, however, there has been much less work in exploring alternative exclusion restrictions in sample selection models compared to treatment-effects models. A body of empirical experience has yet to be built up on the sensitivity of results to different restrictions using the new, less-restrictive methods that have been developed. This should be a topic for research in the future.

3. Conclusions

This survey of econometric methods in labor economics and of recent developments in a few of those methods shows that both practitioners and econometricians are moving in the same direction but without as much contact and interchange as would be fruitful. Empirical work in labor is moving toward less restrictive, more robust, and simpler methods which attempt to isolate and highlight key sources of identification clearly where they can be made the subject of investigation and attention. New developments in econometrics are moving in exactly the same direction but the tools developed there have not spilled over into econometric practice. To do so it is necessary that a body of empirical experience be

[24] Vella (1998) points out that the selection term in (22) can be thought of as a generalized residual from the first-stage regression, which is also closely analogous to IV estimation, for IV can also be formulated by including a first-stage residual in the second-stage equation.

built up so that rules of thumb can be developed, the more useful techniques weeded out from the plethora of those that have been proposed, and incorporated into the commonly used software packages. More work on assessing when and where the relaxation of restrictions makes a difference should be part of this endeavor.

References

Ahn, H. and J. Powell (1993), "Semiparametric estimation of censored selection models with a nonparametric selection mechanism", Journal of Econometrics 58: 3–29.

Aldrich, J. and F. Nelson (1984), Linear probability, logit, and probit models (Sage, Newbury Park).

Amemiya, T. (1981), "Qualitative response models: a survey", Journal of Economic Literature 19: 1483–1536.

Arabmazar, A. and P. Schmidt (1982), "An investigation of the robustness of the tobit estimator to nonnormality", Econometrica 50: 1055–1063.

Ashenfelter, O. and A. Krueger (1994), "Estimates of the economic return to schooling from a new sample of twins", American Economic Review 84: 1157–1173.

Ashenfelter, O. and R. Layard, eds. (1986), Handbook of labor economics, Vols. I and II (North-Holland, Amsterdam).

Barnow, B., G. Cain and A. Goldberger (1980), "Issues in the analysis of selectivity bias", in: E. Stromsdorfer and G. Farkas, eds., Evaluation studies review annual, Vol. 5 (Sage, Newbury Park).

Behrman, J., M. Rosenzweig and P. Taubman (1994), "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment", Journal of Political Economy 102: 1131–1174.

Berkovec, J. and S. Stern (1991), "Job exit behavior of older men", Econometrica 59: 189–210.

Berkson, J. (1944), "Application of the logistic function to bio-assay", Journal of the American Statistical Association 39: 357–365.

Berkson, J. (1951), "Why I prefer logits to probits", Biometrics 7: 327–339.

Björklund, A. and R. Moffitt (1987), "The estimation of wage and welfare gains in self-selection models." Review of Economics and Statistics 69: 42–49.

Blundell, R. and A. Duncan (1998), "Kernel regression in empirical microeconomics", Journal of Human Resources 33: 62–87.

Bound, J., D. Jaeger and R. Baker (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak", Journal of the American Statistical Association 90: 443–450.

Brown, C. and R. Moffitt (1983), "The effect of ignoring heteroskedasticity on estimates of the tobit model", Technical working paper no. 27 (NBER, Cambridge, MA).

Choi, K. (1990), "The semiparametric estimation of the sample selection model using series expansion and the propensity score", Unpublished manuscript (University of Chicago).

Cogan, J. (1981), "Fixed costs and labor supply", Econometrica 49: 945–964.

Cragg, J. (1971), "Some statistical models for limited dependent variables with application to the demand for durable goods", Econometrica 39: 829–844.

Domenich, T. and D. McFadden (1995), Urban travel demand (North-Holland, Amsterdam).

Finney, D.J. (1947), Probit analysis (Cambridge University Press, Cambridge).

Fisher, R.A. (1925), Statistical methods for research workers (Oliver and Boyd, Edinburgh).

Gallant, R. and D. Nychka (1987), "Semi-nonparametric maximum likelihood estimation", Econometrica 55: 363–390.

Geweke, J. and M. Keane (1999), "Mixture of normals probit models", in: C. Hsiao, K. Lahiri, L. Lee and H. Pesaran, eds., Analysis of panels and limited dependent variable models: an edited volume in honor of G.S. Maddala (Cambridge University Press, Cambridge, UK) in press.

Geweke, J., M. Keane and D. Runkle (1994), "Alternative computational approaches to inference in the multinomial probit model", Review of Economics and Statistics 76: 609–632.

Goldberger, A. (1964), Econometric theory (Wiley, New York).

Goldberger, A. (1983), "Abnormal selection bias", in: S. Karlin, T. Amemiya and L. Goodman, eds., Studies in econometrics, time series, and multivariate statistics (Academic Press, New York).

Hajivassiliou, V. (1993), "Simulation estimation methods for limited dependent variable models", in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., Handbook of statistics, Vol. 11 (North-Holland, Amsterdam).

Hajivassiliou, V. and P. Ruud (1994), "Classical estimation methods for limited dependent variable models using simulation", in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. IV (North-Holland, Amsterdam).

Härdle, W. and O. Linton (1994), "Applied nonparametric methods", in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. IV (North-Holland, Amsterdam).

Hausman, J. (1980), "Effects of wages, taxes, and fixed costs on women's labor force participation", Journal of Public Economics 14: 161–194.

Hausman, J. and D. McFadden (1984), "Specification tests for the multinomial logit model", Econometrica 52: 1219–1240.

Heckman, J. (1974), "Shadow prices, market wages, and labor supply", Econometrica 42: 679–693.

Heckman, J. (1978), "Dummy endogenous variables in a simultaneous equation system", Econometrica 47: 931–959.

Heckman, J. (1979), "Sample selection bias as a specification error", Econometrica 47: 153–161.

Heckman, J. (1993), "What has been learned about labor supply in the past twenty years?" American Economic Review 83: 116–121.

Heckman, J. (1997), "Instrumental variables", Journal of Human Resources 32: 441–462.

Heckman, J. and R. Robb (1985), "Alternative methods for evaluating the impact of interventions", in: J. Heckman and B. Singer, eds., Longitudinal analysis of labor market data (Cambridge University Press, Cambridge, UK).

Heckman, J. and B. Singer (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", Econometrica 52: 271–320.

Heckman, J. and J. Snyder (1996), "Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators", Working paper 5785 (NBER, Cambridge, MA).

Honoré, B. and J. Powell (1994), "Pairwise difference estimators of censored and truncated regression models", Journal of Econometrics 64: 241–278.

Horowitz, J. (1993a), "Semiparametric estimation of a work-trip choice model", Journal of Econometrics 58: 49–70.

Horowitz, J. (1993b), "Semiparametric and nonparametric estimation of quantal response models", in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., Handbook of statistics, Vol. 11 (North-Holland, Amsterdam).

Hurd, M. (1979), "Estimation in truncated samples when there is heteroskedasticity", Journal of Econometrics 11: 247–258.

Imbens, G. and J. Angrist (1994), "Identification and estimation of local average treatment effects", Econometrica 62: 467–475.

Keane, M. (1992), "A note on identification in the multinomial probit model", Journal of Business and Economic Statistics 10: 193–200.

Keane, M. (1993), "Simulation estimation for panel data models with limited dependent variables", in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., Handbook of statistics, Vol. 11 (North-Holland, Amsterdam).

Keane, M. (1994), "A computationally practical simulation estimator for panel data", Econometrica 62: 95–116.

Keane, M. and R. Moffitt (1998), "A structural model of multiple welfare program participation and labor supply", International Economic Review 39: 553–589.

Lee, L. (1979), "Identification and estimation in binary choice models with limited (censored) dependent variables", Econometrica 47: 977–996.

Lerman, S. and C. Manski (1981), "On the use of simulated frequencies to approximate choice probabilities", in:

C. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT Press, Cambridge, MA).

Leung, S. and S. Yu (1996), "On the choice between sample selection and two-part models", Journal of Econometrics 72: 107–128.

McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior", in: P. Zarembka, ed., Frontiers of econometric analysis (Academic Press, New York).

McFadden, D. (1981), "Econometric models of probabilistic choice", in: C. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT Press, Cambridge, MA).

McFadden, D. (1989), "A method of simulated moments for estimation of discrete response models without numerical integration", Econometrica 57: 996–1026.

Maddala, G.S. (1995), "Specification tests in limited dependent variable models", in: G.S. Maddala, P.C.B. Phillips and T.N. Srinivasan, eds., Advances in econometrics and quantitative economics (Blackwell, Oxford, UK).

Manser, M. (1999), "Existing labor market data: current and potential research uses", in: J. Haltiwanger, M. Manser and R. Topel, eds., Labor statistics measurement issues (The University of Chicago Press, Chicago, IL) in press.

Manski, C. (1975), "Maximum score estimation of the stochastic utility model of choice", Journal of Econometrics 3: 205–228.

Manski, C. (1988), "Identification of binary response models", Journal of the American Statistical Association 83: 729–738.

Manski, C. (1991), "Regression", Journal of Economic Literature 29: 34–50.

Manski, C. (1994), "The selection problem", in: C. Sims, ed., Advances in econometrics (Cambridge University Press, Cambridge, UK).

Manski, C. and S. Thompson (1986), "Operational characteristics of maximum score estimation", Journal of Econometrics 32: 85–108.

Matzkin, R. (1992), "Nonparametric and distribution-free estimation of the binary crossing and binary choice models", Econometrica 60: 239–270.

Melenberg, B and A. Van Soest (1996), "Parametric and semi-parametric modeling of vacation expenditures", Journal of Applied Econometrics 11: 59–76.

Moffitt, R. (1984), "The estimation of a joint wage-hours labor supply model", Journal of Labor Economics 2: 550–566.

Mroz, T. (1987), "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions", Econometrica 55: 765–800.

Mroz, T. and D. Guilkey (1995), "Discrete factor approximations for use in simultaneous equation models with both continuous and discrete endogenous variables", Working paper no. 95-02 (Department of Economics, University of North Carolina, Chapel Hill, NC).

Nelson, F. (1984), "Efficiency of the two-step estimator for models with endogenous sample selection", Journal of Econometrics 24: 181–196.

Newey, W.K. (1988), "Two-step series estimation of sample selection models", Unpublished manuscript (Princeton University).

Newey, W., J. Powell and J. Walker (1990), "Semiparametric estimation of selection models: new results", American Economic Review 80: 324–328.

Pagan, A. and F. Vella (1989), "Diagnostic tests for models based on individual data: a survey", Journal of Applied Econometrics 4: S29–S59.

Pakes, A. and D. Pollard (1989), "Simulation and the asymptotics of optimization estimators", Econometrica 57: 1027–1058.

Pencavel, J. (1986), "Labor supply of men", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics (North-Holland, Amsterdam).

Powell, J. (1984) "Least absolute deviations estimation for the censored regression model", Journal of Econometrics 25: 303–325.

Powell, J. (1986a) "Censored regression quantiles", Journal of Econometrics 32: 143–155.
Powell, J. (1986b) "Symmetrically trimmed least squares estimation of tobit models", Econometrica 54: 1435–1460.
Powell, J. (1994), "Estimation of semiparametric models", in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. IV (North-Holland, Amsterdam).
Quandt, R. (1956), "A probabilistic model of consumer behavior", Quarterly Journal of Economic Behavior 70: 507–536.
Robinson, P.M. (1993), "Root-N consistent semiparametric regression", Econometrica 56: 931–954.
Staiger, D. and J. Stock (1997), "Instrumental variables regression with weak instruments", Econometrica 65: 557–586.
Stafford, F. (1986), "Forestalling the demise of empirical economics: the role of microdata in labor economics research", in: O. Ashenfelter and R. Layard, eds., Handbook of labor economics (North-Holland, Amsterdam).
Stern, S. (1997), "Simulation-based estimation", Journal of Economic Literature 35: 2006–2039.
Theil, H. (1981), Principles of econometrics (Wiley, New York).
Thurstone, L. (1927), "A law of comparative judgment", Psychological Review 34: 273–286.
Tobin, J. (1958), "Estimation of relationships for limited dependent variables", Econometrica 26: 24–36.
Vella, F. (1998), "Estimating models with sample selection bias", Journal of Human Resources 33: 127–169.