Modeling Heterogeneity

Arthur Lewbel Boston College

October, 2005

1 Introduction

My goal here is to provide some synthesis of recent results regarding unobserved heterogeneity in nonlinear and semiparametric models, using as a context Matzkin (2005a) and Browning and Carro (2005), which were the papers presented in the Modeling Heterogeneity session of the 2005 Econometric Society World Meetings in London. These papers themselves consist of enormously heterogeneous content, ranging from high theory to Danish milk, which I will attempt to homogenize.

The overall theme of this literature is that, in models of individual economic agents, errors at least partly reflect unexplained heterogeneity in behavior, and hence in tastes, technologies, etc.,. Economic theory can imply restrictions on the structure of these errors, and in particular can generate nonadditive or nonseparable errors, which has profound implications for model specification, identification, estimation, and policy analysis.

2 Statistical vs Structural Models of Heterogeneity

Using one of Browning and Carro's models to fix ideas, suppose we have a sample of observations of a dependent variable Y such as a household's purchases of organic whole milk, and a vector of covariates X, such as the prices of alternative varieties of milk and demographic characteristics of the consuming household. The heterogeneity we are concerned with here is unobserved heterogeneity, specifically the behavioral variation in Y that is not explained by variation in X. By behavioral, I mean variation that primarily reflects actual differences in actions, tastes, technologies, etc.,. across the sampled economic agents, rather than measurement or sampling errors. For simplicity, assume for this discussion that our data are independent, identically distributed observations of Y and X, without any complications associated with sample selection, censoring, or measurement errors.

One view of heterogeneity, which could be called the statistical model, is that unobserved heterogeneity is completely defined by $F(Y \mid X)$, that is, the conditional distribution of Y given X, since this gives the probability of any value of Y given any value of X. This can be easily estimated nonparametrically, and by this view the only purpose or reason to construct a model is dimension reduction or otherwise improving the efficiency of our estimate of F.

This may be contrasted with what I'll call the Micro Econometrician's view, which is that there is a behavioral or structural model $Y = g(X, \theta, U)$, which may include parameters θ and has unobservables (heterogeneity) U. We could also include endogeneity, by allowing U to correlate with X or by including Y into the function g, but ignore that complication for now.

We wish to understand and estimate this structural model because it is based on some economic model of behavior and its parameters may have implications for economic policy. The structural model g is more informative than F, because it tells us how F would change if some underlying parameters change. For example, θ could include production function parameters, which would then allow us to predict how the distribution of outputs across firms might change in response to an improvement in technology, or in the milk example θ provides information about the distribution of tastes across households, which we might use to predict the response to introduction of a new milk product, or the returns to a marketing compaign aimed at changing tastes.

To illustrate, the Micro Econometrician might assume a linear random coefficients model $Y = a + bX + \varepsilon = (a + e_1) + (b + e_2)X$, where the errors e_1 and e_2, representing unobserved variation in tastes, are distributed independently of each other and of X, and are normal so $e_j \backsim N(0, \sigma_j^2)$ and $\varepsilon = e_1 + e_2 X$. This then implies the statistical model $F(Y \mid X) = N(a + bX, \sigma_1^2 + \sigma_2^2 X^2)$.

Given a scalar, continuously distributed Y, there is a trick Matzkin employs to construct a function that could be structural. Define an unobserved error term U by $U = F(Y \mid X)$, and a function G by $Y = G(X, U) = F^{-1}(U \mid X)$. The idea is then to assume that the true behavioral model $Y = g(X, \theta, U)$ is just $Y = G(X, U)$. By construction this error term U is independent of X, since it has a standard uniform distribution regardless of what value X takes on. We could further transform U by any known monotonic function, e.g., converting U to a standard normal instead of a standard uniform. Matzkin then adds economically motivated assumptions regarding G and U for identification or to interpret G as structural, e.g., if Y is output G might be interpreted as a production function with U being an unobserved input or factor of production, and with constant returns to scale inplying that G is linearly homogeneous in X and U.

Taking $Y = G(X, U) = F^{-1}(U \mid X)$ to be a structural model is a very tempting construction. It says that, once we abandon the idea that errors must appear additively in models, we can assume that all unobserved heterogeneity takes the form of independent uniform errors! This would appear to conflict with Browning's claim that the way we model errors is often too simplistic, that is, there is typically more (and more complicated) unobserved heterogeneity in the world than in our models.

The source of this conflict is that the clean error construction $Y = G(X, U)$ requires that G be the inverse of F, but a model $Y = g(X, \theta, U)$ that has some economically imposed structure is typically not the inverse of F. Equivalently, the construction of G assumes that the unobserved heterogeneity U equals conditional quantiles of Y, while a structural model g may assume errors of another form. For example, in the random coefficients model $Y = a + bX + (e_1 + e_2 X)$ the error $\varepsilon = e_1 + e_2 X$ appears additively, and imposing constant returns to scale in X,ε implies only $a = 0$. In contrast, with an independent uniform error U the model is $Y = G(X, U) = a + bX + (\sigma_1^2 + \sigma_2^2 X^2)^{1/2} \Phi^{-1}(U)$ where Φ^{-1} is the inverse of the cumulative standard normal distribution function. I chose this example for its simplicity, but even in this model one can see that imposing structure like additivity in U or constant returns to scale in X, U would place very different restrictions on behavior than imposing additivity in ε or constant returns to scale in X, ε.

Browning and Carro consider structural models we could write generically as latent additive error models, $Y = g\,(h(x) + U)$. Knowing that we can always write $Y = G(X, U)$ with U independent of X, we might hope that after imposing a relatively mild restriction like latent additivity, the resulting errors U in $Y = g\,(h(x) + U)$ could still be close to independent. However, using panels that are long enough to permit separate estimation for each household, Browning and Carro finds empirically that the distribution of U depends strongly on X in complicated ways.

3 Multiple Equations, Coherence, and Invertibility

Now consider multiple equation models, so Y is now a vector of endogenous variables. Once we allow for nonadditivity or nonlinearity of errors or endogenous regressors, the issue of coherency arises. An incoherent model is one in which, for some values of the errors or regressors, there is either no corresponding value for Y, or multiple values for Y (when the incoherence only takes the form of multiple solutions, the problem is sometimes called incompleteness). See Heckman (1978), Gourieroux, Laffont, and Monfort (1980), Blundell and Smith (1994), and Tamer (2003).

For example, consider the simple system of equations $Y_1 = I(Y_2 + U_1 \geq 0)$, $Y_2 = \alpha Y_1 + U_2$, where I is the indicator that equals one if its argument is true and zero otherwise. This system might arise as the reaction function of two players in a game, one of whom chooses a discrete strategy Y_1, and the other a continuous strategy Y_2. This innocuous looking model is incoherent. The system has multiple solutions, implying both $Y_1 = 0$ and $Y_1 = 1$, when $-a \leq U_1 + U_2 < 0$, and it has no solution, with neither $Y_1 = 0$ nor $Y_1 = 1$ satisfying the system, if $0 \leq U_1 + U_2 < -a$. This example is from Lewbel (2005), who provides general conditions for coherency of models like these. Examples of incoherent or incomplete economic models similar to this example include Bresnahan and Reiss (1991) and Tamer (2003).

When we write down econometric models, coherence is usually taken for granted to such an extent that many researchers are unaware of the concept. This may be due to the fact that linear models, optimizing models, and triangular models are all generally coherent, or typically incomplete in only harmless ways, such as potential ties in optimizing solutions that occur with probability zero.

A coherent model is one that has a well defined reduced form, that is, it's a model for which a function G exists such that $Y = G(X, U)$ for all values that X and U can take on. Closely related to coherence is invertibility. A model is invertible if a function H exists such that $U = H(Y, X)$, so a coherent model can be solved for Y and an invertible model can be solved for U. Invertibility is convenient for estimation, e.g., it implies one can do maximum likelihood if the U distribution is parameterized, or one can construct moments for GMM estimation if U is uncorrelated with instruments. For structural models, if U represents unobserved taste or technology parameters, then invertibility generally means that we can estimate the distribution of these tastes or technologies.

Invertibility depends on how the structural model, and hence the corresponding errors, are defined. The probit model $Y = I(a + bX + e \geq 0)$ with standard normal e independent of X is, like most latent variable models, coherent but not invertible in the latent error e. However, this model can be rewritten as $Y = \Phi(a + bX) + \varepsilon$ which is invertible in the heteroskedastic error $\varepsilon = Y - E(Y \mid X)$.

With a continuously distributed vector Y, Matzkin uses an extension of the idea of inverting the distribution function to construct coherent, invertible models, with errors U that are independent of X. Let $Y = (Y_1, Y_2)$. Define errors $U = (U_1, U_2)$ by $U_1 = F_1(Y_1 \mid X)$, and $U_2 = F_2(Y_2 \mid X, U_1)$, where F_1 and F_2 are conditional distribution functions of Y_1 and Y_2. Then let $G_j = F_j^{-1}$ to obtain the coherent, invertible, triangular model $Y = G(X, U)$ defined by $Y_1 = G_1(X, U_1)$ and $Y_2 = G_2(X, U_1, U_2)$. The extension to Y_3 and up is immediate. As before, the question that arises is whether this construction of G is just a statistical trick or a model that represents underlying behavior, and whether economic modeling restrictions can be imposed on this G without reintroducing substantial dependence of U on X (and without introducing incoherence).

Consider the example where Y is a consumption bundle, that is, a K vector of quantities of goods that an individual consumes, determined by maximizing a utility function subject to a linear budget constraint. Then X is a vector consisting of the prices of the K goods, the consumer's income or total expenditures on goods, and observed characteristics of the consumer that affect or are otherwise correlated with the consumer's utility function. In this model, U is interpreted as a vector of unobserved utility function parameters. A necessary condition for coherence and invertibility of this model is that U have $K - 1$ elements. Roughly, this is because the demand functions for $K - 1$ of the goods each have an error term corresponding to an element of U, and then the budget constraint pins down the quantity of the $K'th$ good.

Actually achieving coherency and invertibility in this application is surprising difficult, because utility maximization imposes constraints on $Y = G(X, U)$, such as symmetry and negative semidefiniteness of the Slutsky matrix, that must be satisfied for every value that U can take on. See, e.g., van Soest, Kapteyn, and Kooreman (1993). In particular, if consumers are price takers with fixed preferences, then the utility parameters U should be independent of prices, but semidefiniteness imposes inequality constraints on demand functions and hence upon U that, except for very special classes of preferences, will generally be functions of prices. Similarly, Lewbel (2001) shows that additive demand errors must generally be heteroskedastic. Another example that Browning mentions is that popular utility functions for empirical demand system estimation do not coherently encompass ordinary heterogeneity models such as random coefficients.

Brown and Matzkin (1998) provide one rather restrictive demand model that is coherent and invertible, which are demands derived from a random utility function of the form $V(Y) + \widetilde{Y}'U$, where $\widetilde{Y}$ denotes the K vector Y with one element Y_K removed. Beckert and Blundell (2004) provide some generalizations of this model. See also the stochastic revealed preference results summarized in McFadden (2005). Matzkin's paper combines these concepts, considering coherent, invertible demands with unobservables U independent of X, reparameterized using the above technique of sequentially inverting distribution functions, and gives conditions for their nonparametric identification.

4 Nonparametric Identification

There is a small but growing literature on identification with incomplete (and hence in that way incoherent) models, which is closely related to set identification concepts, but the vast bulk of the identification literature assumes coherent, complete models. I will focus on ordinary point identification assuming coherency, but note that the unknown objects we are identifying can be entire functions, not just finite parameter vectors.

The statistical model $F(Y \mid X)$ tells us everything we can know from the data about the response in Y to a change in X. We can therefore define identification as the ability to recover unknown structural parameters or functions from the statistical model. For example, in the normal random coefficients model $Y = (a+e_1)+(b+e_2)X$ identification means obtaining the parameter values, a, b, σ_1, and σ_2 from the distribution function F. With $F(Y \mid X) = N(a+bX, \sigma_1^2+\sigma_2^2 X^2)$, the slope and intercept of the function $E(Y \mid X)$ are a and b, while the slope and intercept of the function $var(Y \mid X^2)$ are σ_1 and σ_2, so we have identification. More generally, in the random coefficients model the joint distribution of the two random variables $a+U_1$ and $b+U_2$ is nonparametrically identified assuming they are independent of X, a fact that Browning exploits in some of his work.

Brown (1983) and Roehrig (1988) provided general conditions for nonlinear and nonparametric identification of structural functions. Assume a coherent, invertible model with error vector U that is independent of X, and invert the structural model $Y = G(X, U)$ to obtain $U = H(Y, X)$. The model H or equivalently G is nonparametrically identified if there does not exist any other observationally equivalent model, that is, any alternative inverse model $\widetilde{H}(Y, X)$ such that $\widetilde{U}$ defined by $\widetilde{U} = \widetilde{H}(Y, X)$ is also independent of X, where $\widetilde{H}$ also possesses whatever properties were used to define or characterize H. These properties could be parameterizations or functional restrictions such as additivity, separability, homogeneity, monotonicity, or latent index constructions.

Such identification generally requires some normalizations of parameters or of functions, for example, we must rule out trivial cases where $\widetilde{H}$ is just a simple monotonic transformation of H. With some assumptions, Roehrig (1988) expressed this identification condition as a restriction on the rank of a matrix of derivatives of $\partial H(Y,X)/\partial Y$, $\partial H(Y,X)/\partial X$, $\partial \widetilde{H}(Y,X)/\partial Y$, and $\partial \widetilde{H}(Y,X)/\partial X$, and Brown (1983) earlier provided a similar expression based on parameterization.

Recently, Benkard and Berry (2004) found, by means of a counterexample, a flaw in this characterization of identification. The source of the problem is that one must consider both $H(Y,X)$ and the distribution function of U, $F_U(U) = F_U(H(Y,X))$. Matzkin (2005b) fixes the Brown and Roehrig results, showing that the matrix one actually needs to check the rank of depends on functions of derivatives of F_U and $F_{\widetilde{U}}$ as well as the above derivatives of H and $\widetilde{H}$.

Matzkin (2005a) applies this result to a range of coherent demand models. Some cleverness is required to find sufficient functional restrictions and normalizations to obtain nonparametric identification. This generally involves limiting the number and dimension of unknown functions by using additivity and separability restrictions. One example she considers is utility derived demand systems with K goods and a linear budget constraint using the utility function $V(\widetilde{Y}) + \widetilde{Y}'U + Y_K$ with U independent of X. She shows nonparametric identification of V and the distribution of U in this model, essentially using this new nonparametric identification machinery to recast and extend results from Brown and Matzkin (1998).

Another example she considers is utility derived demand systems with two goods and a linear budget constraint, which for coherence and invertibility requires that U equal a scalar U_1, and assuming U_1 is independent of X. A general utility function $V(Y_1, Y_2, U_1)$ is not identified from demand functions in this case, essentially because a general three dimensional function cannot be recovered from two demand equations. Matzkin obtains nonparametric identification in this example by reducing the dimensionality of the problem, specifically by assuming the utility function has the additively separable form $V_1(Y_1, Y_2) + V_2(Y_1, U_1)$, where V_1, and V_2 are unknown functions that she shows can be nonparametrically identified from $F(Y \mid X)$ (essentially corresponding to nonparametric Marshallian demand functions), again assuming a suitably normalized U that is independent of X.

Matzkin also considers stronger restrictions on utility to obtain identification in cases where either total expenditures or prices are not observed. The former is closely related to nonlinear hedonic modeling (see, e.g., Ekeland, Heckman, and Nesheim 2002) while the latter is concerned with Engel curve estimation. Regarding the latter, it would be useful to derive the connections between these new identification theorems and results in the older demand system literature on identification of Engel curve rank (Gorman 1981, Lewbel 1991), and identification of equivalence scales using Engel curve data and the related use of variation in characteristics to help identify price effects (Barten 1964, Muellbauer 1974, Lewbel 1989).

5 Discrete Choice Models

In his presentation, Browning observed, "Conventional schemes have usually been devised by statisticians to introduce hetereogeneity in such a way as to allow us to immediately take it out again." This is particularly true in discrete choice models, where estimators, and associated assumptions about all types of errors (not just heterogeneity) are often chosen based on tractibility rather than behavioral realism. A particularly egregious example is the widespread use of the linear probability model to deal with measurement errors, endogeneity, or fixed effects.

Browning and Carro consider parametric dynamic discrete choice models allowing for substantial heterogeneity, including both fixed effects and coefficients that vary across individuals. They note that in short (fixed T) panels, the presence of individual specific parameters implies that unbiased estimators for transition probabilities do not exist and they propose, as an alternative to maximum likelihood estimation, a bias adjusted maximum likelihood estimator and a minimum integrated mean squared error estimator. These results and estimators are essentially applications of recent theoretical advances in the larger literature on the incidental parameters problem. See, e.g, Hahn and Newey (2004) and Woutersen (2004). However, it is informative to see just how much these adjustments matter in empirical work, and more importantly, how strong the evidence is for all kinds of unobserved heterogeneity

Turning to less parametric models, let Z be a scalar variable that is observed along with the vector X. Assume Y is conditionally independent of Z, conditioning on X. Assume now that we do *not* observe Y, but instead observe a discrete binary D given by the latent variable threshold crossing model $D = I(Y - Z > 0)$ where I is the indicator function that equals one when its argument is true and zero otherwise. For example, D could indicate the purchase of a good, where Z is a price faced by the consumer and Y is the consumer's unobserved reservation price. More generally, Y could be an observed benefit and Z the observed cost of some decision D, or Z could be any observed variable that monotontically affects the probability of choosing $D = 1$, and Y is all other observed and unobserved variables that determine D.

In this model $E(D \mid Z, X) = 1 - \Pr(Y \leq Z \mid X, Z) = 1 - F(Z \mid X)$, where as before F denotes the conditional distribution of the (now latent) Y conditioning on X. It follows that the distribution F is identified for all Y on the support of Z, and is therefore identified everywhere if the support of Z is large enough to contain the support of Y. Here Z is a special regressor as in Lewbel (2000), who uses this construction to estimate parameters in a model of Y.

Most of Matzkin's discrete choice identification results are variants of this construction. In particular, any set of assumptions that are used to identify a model $Y = g(X, U)$ where X and a continuous Y are observed can now alternatively be used to identify the latent variable discrete choice model $D = I(Z - g(X, U) \geq 0)$ where we instead observe D, Z, and X. Lewbel (2000) focused on linear $g(X, U)$ with errors U that are heteroskedastic or correlated with some elements of X (such as endogenous or mismeasured regressors). Lewbel, Linton, and McFadden (2004) use this model to estimate nonparametric conditional moments of Y. Matzkin considers cases such as nonparametric G assuming an independent U. Lewbel and Matzkin also both extend these ideas to multinomial choice, assuming we can observe a separate Z for each possible choice.

6 Endogenous Regressors

Matzkin next considers identification of some general nonparametric models with endogenous regressors. In addition to the special regressor as above, another general technique for identifying such models is control function methods (See Lewbel 2004 for a comparison of the two methods in the above discrete choice context when g is linear). The control function approach has a long history going back at least to Heckman (1978), but in its modern form assumes we observe vectors Y, X, and instruments Z where $Y = g(X, U)$ and $X = r(Z, e)$ and U and e are correlated. The key assumption here is that U and e are independent of Z, so the source of endogeneity is only through e, that is, X and U are correlated *only* because e and U are correlated. This then implies that X and U are conditionally independent, conditioning on e, so instead of identifying the g model based on $F(Y \mid X)$, we can instead use $F(Y \mid X, e)$. Essentially, the control function method consists of fixing the endogeneity problem by nonparametrically including e as an additional regressor in the model for Y. For example, If $Y = r(X'\beta + U)$, $X = Z'\alpha + e$ and $U = e\gamma + \varepsilon$ where ε is independent of e and Z, then $Y = r(X'\beta + e\gamma + \varepsilon)$ and ε must be independent of e and X, so including both X and e as regressors results in a model that no longer suffers from endogeneity.

We now have a whole range of tools that can be mixed and matched. Essentially, any of the earlier described techniques could be used in the first step to define and identify e in $X = r(Z, e)$, and again any might be used in the second step to define and identify U. For example, in the first stage if the endogenous X is continuous then e could be defined by inverting the conditional distribution function of X given Z, or if the endogenous X is discrete then the distribution function of e could be obtained using a special regressor in the X model. Either method could also be used in the model for Y given X and e. Other variants can also be applied in either step, for example, semiparametric parameters of the Y model could be obtained when Y is discrete from a nonparametric regression of Y on X and e as in Blundell and Powell (2004). Matzkin (2004) suggests exchanging the roles of Z and e, letting e be an observed covariate and z be an unobserved instrument, which then removes the need to identify and estimate e. Other examples and variations include Altonji and Matzkin (2005), Chesher (2003), and Imbens and Newey (2003).

With such an assortment of models to choose from, there is little excuse for continued use of flawed specifications such as the linear probability model for dealing with problems like endogeneity and heterogeneity.

7 Conclusions

In classical econometrics, virtually all errors were treated (either essentially or literally) as measurement errors. Models such as McFadden (1973) that took errors seriously as heterogeneity parameters were the exception rather than the rule. However, it is clear that the errors in most microeconomic relationships are so large and systematic that they must to some extent represent true structural heterogeneity (though misspecification and mismeasurement may also loom large). Indeed, it is common in microeconomic models to see R^2 statistics far below a half, that is, often most of the variation in Y is unexplained individual variation. The errors matter more than the regressors! This is a humbling realization for the researcher, but it demonstrates the crucial importance of developing specification, identification, and estimation methods that incorporate realistic models of unobserved individual variation.

References

[1] ALTONJI, J. AND R. L. MATZKIN (2005), “Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors,” Econometrica, 73, 1053-1102.

[2] BARTEN, A. P. (1964), ”Family Composition, Prices, and Expenditure Patterns,” in Econometric Analysis for National Economic Planning: 16th Symposium of the Colston Society, P. Hart, G. Mills, and J. K. Whitaker, eds., 277-292, London: Butterworth.

[3] BENKARD, C. L. AND S. BERRY (2004), "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988), Cowles Foundation Discussion Paper #1482.

[4] BECKERT, W. AND R. BLUNDELL (2004), Invertibility of Nonparametric Stochastic Demand Functions, Birkbeck Working Papers in Economics and Finance No. 406.

[5] BLUNDELL, R. AND J. L. POWELL (2004), “Endogeneity in Semiparametric Binary Response Models,” Review of Economic Studies, 71, 581-913.

[6] BLUNDELL, R. AND R. J. SMITH (1994), ”Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables,” Journal of Econometrics, 64 355-373.

[7] BRESNAHAN, T. F. AND P. C. REISS (1991), "Empirical Models of Discrete Games." Journal of Econometrics, 48, 57-81

[8] BROWN, B. W. (1983), "The Identification Problem in Systems Nonlinear in the Variables" Econometrica, 51, 175-196.

[9] BROWN, D. J. AND R. L. MATZKIN, (1998), "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," Cowles Foundation Discussion Papers 1175, Cowles Foundation, Yale University.

[10] BROWNING, M. AND J. CARRO (2005), "Heterogeneity in Dynamic Discrete Choice Models," 2005 Econometric Society World Meetings London.

[11] CHESHER, A. (2003), "Identification in nonseparable models," Econometrica, 71, 1405-1441.

[12] EKELAND, I, J. HECKMAN, AND L. NESHEIM (2002), “Identifying Hedonic Models,” American Economic Review, 92, 304-309.

[13] GORMAN, W. M. (1981), "Some Engel Curves," in Essays in the Theory and Measurement of Consumer Behaviour in Honor of Sir Richard Stone, ed. by Angus Deaton, Cambridge: Cambridge University Press.

[14] GOURIEROUX, C., J. J. LAFFONT, AND A. MONFORT (1980), ”Coherency Conditions in Simultaneous Linear Equations Models with Endogenous Switching Regimes,” Econometrica, 48, 675-695.

[15] HAHN, J. AND W. NEWEY, (2004), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Data Models," Econometrica 72, 1295-1319.

[16] HECKMAN, J. J. (1978), ”Dummy Endogenous Variables in Simultaneous Equation Systems,” Econometrica, 46 931-960.

[17] IMBENS, G. AND W. NEWEY (2003). "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," unpublished manuscript.

[18] LEWBEL, A. (1989), "Identification and Estimation of Equivalence Scales Under Weak Separability," Review of Economic Studies, 56, 311-316.

[19] LEWBEL, A. (1991), "The Rank of Demand Systems: Theory and Nonparametric Estimation," Econometrica, 59, 711-730.

[20] LEWBEL, A. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity and Instrumental Variables," Journal of Econometrics 97, 145-177.

[21] LEWBEL, A. (2001), "Demand Systems with and without Errors," American Economic Review, 91, 611-618

[22] LEWBEL, A. (2004) "Simple Estimators for Hard Problems," unpublished manuscript.

[23] LEWBEL, A. (2005), "Coherence of Structural Models Containing a Dummy Endogenous Variable," unpublished manuscript.

[24] LEWBEL, A., O. LINTON, AND D. MCFADDEN (2004) "Estimating Features of a Distribution From Binomial Data," unpublished manuscript.

[25] MATZKIN, R.L. (2004), "Unobservable Instruments," Unpublished Manuscript.

[26] MATZKIN, R.L. (2005a), "Heterogeneous Choice," 2005 Econometric Society World Meetings London.

[27] MATZKIN, R.L. (2005b), "Identification in Simultaneous Nonparametric Equations," Unpublished Manuscript.

[28] MCFADDEN, D. L., (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in Frontiers in Econometrics, edited by P. Zarembka, New York: Academic Press.

[29] MCFADDEN, D. L. (2005), ”Stochastic Revealed Preference: A Synthesis," Economic Theory, 26, 245-264.

[30] MUELLBAUER, J. (1974) ‘Household composition, Engel curves and welfare comparisons between households’. European Economic Review, 5, 103-22.

[31] ROEHRIG, C. S. (1988), "Conditions for Identification in Nonparametric and Parametric Models," Econometrica, 56, 433-447.

[32] TAMER, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," Review of Economics Studies, 70, 147-165.

[33] VAN SOEST, A., KAPTEYN, A. AND P. KOOREMAN (1993), "Coherency and Regularity of Demand Systems with Equality and Inequality Constraints." Journal of Econometrics. vol. 57.

[34] WOUTERSEN, T. (2004), "Robustness Against Incidental Parameters," unpublished manuscript.