NONPARAMETRIC STRUCTURAL MODELS

Rosa L. Matzkin*
Department of Economics
Northwestern University

November 2006

The interplay between economic theory and econometrics comes to its full force when analyzing structural models. These models are used in Industrial Organization, Marketing, Public Finance, Labor Economics, and many other fields in economics. Structural econometric methods make use of the behavioral and equilibrium assumptions specified in economic models to define a mapping between the distribution of the observable variables and the primitive functions and distributions that are used in the model. Using these methods, one can infer elements of the model, such as utility and production functions, that are not directly observed. This allows one to predict behavior and equilibria outcomes under new environments and to evaluate the welfare of individuals and profits of firms under alternative policies, among other benefits.

To provide an example, suppose that one would like to predict the demand for a new product. Since the product has not previously been available, no direct data exists. However, one could use data on the demand for existent products together with a structural model, as shown and developed by McFadden (1974). Characterize the new product and the existent competing products by vectors of common attributes. Assume that consumers derive utility from the observable and unobservable attributes of the products, and that each chooses the product that maximizes his/her utility of those attributes among the existent products. Then, from the choice of consumers among existent products, one can infer their preferences for the attributes, and then predict what the choice of each of them would be in a situation when a new vector of attributes, corresponding to the new product, is available. Moreover, one could get a measure of the difference in the welfare of the consumers when the new product is available.

Economic theory seldom has implications regarding the parametric structures that functions and distributions may possess. The behavioral and equilibrium specifications made in economic models typically imply shape restrictions, such as monotonicity, concavity, homogeneity, weak separability, and additive separability, and exclusion restrictions, but not parametric specifications, such as linearity of conditional expectations, or normal distributions for unobserved variables. Nonparametric methods, which do not require specification of parametric structures for the functions and distributions in a model are ideally fitted, therefore, to analyze structural models, using as few a-priory restrictions as possible. Nonparametric techniques have been applied to many models, such as discrete choice models, tobit models, selection models, and duration models. We will concentrate here, however, on the basic models and, on those, indicate some of the latest works that have dealt with identification and estimation.

Nonparametric structural econometric models

As with parametric models, a nonparametric econometric model is characterized by a vector X of variables that are determined outside the model and are observable, a vector ε of variables that are determined outside the model and are unobservable, a vector Υ of outcome variables, which are determined within the model and are unobservable, and a vector Y of outcome variables that are determined within the model and are observable. These variables are related by functional relationships, which determine the causal structure by which Υ and Y are determined from X and ε. The functional relationships are characterized by some functions that are known and some that are unknown. Similarly, some distribution may be known, some are unknown, and the others should be derived from the functional relationships and the known and unknown functions and distributions. Let $\underline{h}^*$ denote the vector of all the unknown functions in the model, $\underline{F}^*$ denote the vector of all unknown distributions, and $\zeta^* = (\underline{h}^*, \underline{F}^*)$. In contrast to parametric models, in nonparametric models, none of the coordinates of ζ^* is assumed known up to a finite dimensional parameter. Only restrictions such as continuity or values of the conditional expectations are assumed. The specification of the model should be such that from any vector $\zeta = (\underline{h}, \underline{F})$, satisfying those same restrictions that ζ^* is assumed to satisfy, one is able to derive a distribution for the observable variables, $F_{Y,X}(\cdot, \cdot; \zeta)$.

Nonparametric identification

When specifying a econometric model, we may be interested in testing it, or we may be interested in estimating $\zeta^* = (\underline{h}^*, \underline{F}^*)$ or some feature of ζ^*, such as only one of the elements of $\underline{h}^*$, or even the value of that element at one point. Suppose that interest lies on estimating a particular feature, $\Psi(\zeta^*)$, of ζ^*. The first question one must answer is whether that feature is identified. Let Ω denote the set of all possible values that $\Psi(\zeta)$ may attain, when ζ is restricted to satisfy the properties that ζ^* is assumed to satisfy. Given $\psi \in \Omega$, define $\Gamma_{Y,X}(\psi)$ to be the set of all probability distributions of (Y, X) that are consistent with ψ. This is the set of all distributions that can be generated by a ζ, satisfying the properties that ζ^* is assumed to satisfy, and with $\Psi(\zeta) = \psi$. We say that two values $\psi, \psi' \in \Omega$ are *observationally equivalent* if

$$\left[\Gamma_{Y,X}(\psi) \cap \Gamma_{Y,X}\left(\psi'\right)\right] \neq \varnothing$$

That is, they are observationally equivalent if there exist a distribution of the observable variables that could have been generated by two vectors ζ and ζ' with $\Psi(\zeta) = \psi$ and $\Psi\left(\zeta'\right) = \psi'$. The feature $\psi^* = \Psi(\zeta^*)$ is said to be *identified* if there is no $\psi \in \Omega$ such that $\psi \neq \psi^*$ and ψ is observationally equivalent to ψ^*. That is, $\psi^* = \Psi(\zeta^*)$ is identified if a change from ψ^* to $\psi \neq \psi^*$ cannot be compensated by a change in other unknown elements of ζ, so that a same distribution of observable variables could be generated by both, vectors ζ^* and ζ with $\psi^* = \Psi(\zeta^*)$ and $\psi = \Psi(\zeta)$.

When ψ^* can be expressed as a continuous functional of the distribution of observable variables (Y, X), one can estimate ψ^* nonparametrically by substituting the distribution by a nonparametric estimator for it.

Additive and nonadditive models with exogenous explanatory variables

The current literature on nonparametric econometrics methods considers additive and nonadditive models. In *additive models*, the unobservable variables ε are specified as affecting the value of Y though an additive function. Hence, for some functions m and v and some unobservable η

$$\begin{aligned} Y &= m(X) + v(X, \varepsilon) \\ &= m(X) + \eta \end{aligned}$$

In these models, the object of interest is typically the function m. Depending on the restrictions that one may impose on η, m may denote a conditional expectation, a conditional quantile, or some other function. Many methods exist to estimate conditional means and conditional quantiles nonparametrically. Prakasa Rao (1983), Härdle and Linton (1994), Pagan and Ullah (1999), Matzkin (1994), X. Chen (2005), and Koenker (2005), among others, survey parts of this literature. Some nonparametric tests for these models include Wooldridge (1992), Yatchew (1992), Hong and White (1995), and Fan and Li (1996).

In *nonadditive models,* one is interested in analyzing the interaction between the unobservable and observable explanatory variables. These models are specified, for some function m as

$$Y = m(X, \varepsilon)$$

Nonparametric identification and estimation in models of this type was studied in Roehrig (1988), Olley and Pakes (1996), Brown and Matzkin (1998), Matzkin (1999, 2003), Imbens and Newey (2003), and Athey and Imbens (2006), among others.

Dependence between observable and unobservable explanatory variables

In econometric models, it is often the case that in an equation of interest, some of the explanatory variables are endogenous; they are not distributed independently of the unobservable explanatory variables in that same equation. This typically occurs when restrictions such as agent's optimization and equilibrium conditions generate interrelationships among observable variables and unobservable variables, ε, that affect a common observable outcome variable, Y. In such cases, the distribution of the observable outcome and observable explanatory variables does not provide enough information to recover the causal effect of those explanatory variables on the outcome variable, since changes in those explanatory variables do not leave the value of ε fixed. A typical example of this is when Y denotes quantity demanded for a product, X denotes the price of the product, and ε is an unobservable demand shifter. If the price that will make firms produce a certain quantity increases with quantity, this change in ε will generate an increment in the price X. Hence, the observable effect of a change in price in demanded quantity would not correspond to the effect of changing the value of price when the value ε stays constant. Another typical example arises when analyzing the effect of years of education on wages. An unobservable variable, such as ability, affects wages and also affects the decision about years of education.

Estimation techniques for additive and nonadditive functions of endogenous variables

The estimation techniques that have been developed to estimate nonparametric models with endogenous explanatory variables typically make use of additional information, which provides some exogenous variation on either the value of the endogenous variable or on the value of the unobservable variable. The common procedures are based on conditional independence methods and on instrumental variable methods. In the first set of procedures, independence between the unobservable and observable explanatory variables in a model is typically achieved by either *conditioning on observable* variables, or *conditioning on unobservable* variables. A *control function* approach (Heckman and Robb (1985)) models the unobservable as a function, so that conditioning on that function purges the dependence between the explanatory observable and unobservable variables in the model. Instrumental variable methods derives identification from an independence condition between the unobservable and an external variable (an instrument) or function, which might be estimable.

Conditioning on unobservable variables often requires the estimation of those unobservable variables. Two step procedures, where they are first estimated, and then used as additional regressors in the model of interest have been developed for additive models by Ng and Pinkse (1995) and Pinkse (2000), and Newey, Powell and Vella (1999), among others. Two step procedures for nonadditive models have been developed by Altonji and Matzkin (2001), Blundell and Powell (2003), Chesher (2003), and Imbens and Newey (2003), among others. Conditional moment estimation methods or quasi-maximum likelihood estimation methods can also be used (see, for example, Ai and Chen (2003).) Altonji and Ichimura (2000), Altonji and Matzkin (2001, 2005), and Matzkin (2004), among others, considered conditioning on observables for estimation of nonadditive models with endogenous explanatory variables. Matzkin (2004) provides insight into the sources of exogeneity that are generated when conditioning on either observables or unobservbles, and which allow identificiation and estimation in nonadditive models. In particular, if $Y = m(X, \varepsilon)$, with m strictly increasing in ε, and ε is independent of X conditional on W, she shows that there exists functions $s(W, \eta)$ and $r(W, \delta)$ such that δ is independent η conditional on W, $X = s(W, \eta)$, and $\varepsilon = r(W, \delta)$. Hence,

$$Y = m(X, \varepsilon) = m(s(W, \eta), r(W, \delta))$$

.

Instrumental variable methods for additive models were considered by Newey and Powell (1989, 2003), Ai and Chen (2003), Darolles, Florens and Renault (2002), and Hall and Horowitz (2003), among others. To develop estimators for m in the model

$$Y_1 = m(Y_2) + \varepsilon \qquad\qquad E[\varepsilon|Z] = 0$$

they use the moment condition

$$E[Y_1|Z=z] = \int m(y_2)\ f_{Y_2|Z=z}(y_2)\ dy_2$$

that depends on the functions $E[Y_1|Z=z]$ and $f_{Y_2|Z=z}(y_2)$, which can be estimated nonparametrically. For nonadditive models, of the form

$$Y_1 = m(Y_2, \varepsilon) \qquad \varepsilon \text{ independent of } Z$$

where m is strictly increasing in ε, Chernozhukov and Hansen (2005), Chernozhukov, Imbens and Newey (2004) developed estimation methods using the moment condition that for al τ

$$\tau = E[1(\varepsilon < \tau)] = E[1(\varepsilon < \tau)|Z]$$

from which m can be estimated using the conditional moment restriction

$$E[1(Y < m(W, \varepsilon)) - \varepsilon|Z] = 0$$

Matzkin (2006) considered the model

$$\begin{aligned} Y_1 &= m_1(Y_2, \varepsilon) \\ Y_2 &= m_2(Y_1, Z, \eta) \end{aligned}$$

where Z is distributed independently of (ε, η). She establishes restrictions on the functions m_1 and m_2 and on the distribution of (ε, η, Z) under which

$$\left[\frac{\partial r_1(y_1, y_2)}{\partial y_2}\right]^{-1} \left[\frac{\partial r_1(y_1, y_2)}{\partial y_1}\right]$$

can be expressed as a function of $f_{Y_1,Y_2|Z=z^*}(y_1, y_2)$, where r_1 is the inverse of m_1 with respect to ε, and the value z^* of the instrument Z is easily identified. (See also Matzkin (2005a, 2006).)

Estimation of averages and average derivatives

Nonparametric estimators are notorious by their slow rate of convergence, which worsens as the dimension of the number of arguments of the nonparametric function increases. A remedy for this is to consider averages of the nonparametric function. The average derivative method in Powell, Stock and Stoker (1989) and the partial integration methods of Newey (1994) and Linton and Nielsen (1995), for example, show how rates of convergence can increase by averaging a nonparametric function or its derivatives. Those results have been extended to cases where the explanatory variables are endogenous, using additional variables to deal with the endogeneity. Examples are Blundell and Powell (2003)'s *average structural function*, Imbens and Newey (2003)'s *average quantile function*, and Altonji and Matzkin (2001, 2005)'s *local average response* function.

Suppose, for example, that the model of interest is

$$Y_1 = m\left(Y_2, \varepsilon\right)$$

and W is such that Y_2 and ε are independent conditional on W. Then, Blundell and Powell (2003) average structural function is

$$G(y_2) = \int m\left(y_2, \varepsilon\right) \quad f_\varepsilon(\varepsilon)\; d\varepsilon$$

which can be derived from a nonparametric estimator for the distribution of (Y_1, Y_2, W) as

$$G(y_2) = \int E\left(Y_1 | Y_2 = y_2, W = w\right)\; f_W(w)\; dw$$

Imbens and Newey (2003) quantile structural function is defined as

$$r\left(y_2, y_1\right) = \Pr\left(m\left(Y_2, q_{\varepsilon_1}(\tau)\right) \leq y_1 | Y_2 = y_2\right)$$

which can be estimated by

$$r\left(y_2, y_1\right) = \int \Pr\left(Y_1 \leq y_1 | Y_2 = y_2, W = w\right)\; f_W(w)\; d\nu$$

Altonji and Matzkin (2001, 2005) local average response function is

$$\beta(y_2) = \int \frac{\partial m\left(y_2, \varepsilon\right)}{\partial y_2}\; f_{\varepsilon|Y_2=y_2}(\varepsilon)\; d\varepsilon$$

which can be derived from a nonparametric estimator for the distribution of (Y_1, Y_2, W) as

$$\beta(y_2) = \int \frac{\partial E\left(Y_1 | Y_2 = y_2, W = w\right)}{\partial y_2}\; f_{W|Y_2=y_2}(w)\; dw$$

Conclusions

The literature on nonparametric structural models has been rapidly developing in recent years. The new methods allow one to analyze counterfactuals without making use of parametric assumptions. Estimation of some features of the model rather than the functions themselves may reduce the curse of dimensionality therefore providing improved properties and reducing the need for large data sets.

References

AI, C. and X. CHEN (2003), "Efficient Estimation of Models with Conditional Moments Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1843.

ALTONJI, J.G. AND H. ICHIMURA (2000): “Estimating Derivatives in Nonseparable Models with Limited Dependent Variables,” mimeo.

ALTONJI, J.G. AND R.L. MATZKIN (2001): “Panel Data Estimators for Nonseparable Models with Endogenous Regressors", NBER Working Paper T0267.

ALTONJI, J.G. AND R.L. MATZKIN (2005): “Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors,” *Econometrica,* 73, 3, 1053-1102.

ATHEY, S. and G. IMBENS (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica,* 74, 2, 431-497.

BLUNDELL, R. and J. L. POWELL (2003) "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress,* Volume II, edited by M. Dewatripont, L.P. Hansen, and S.J.Turnovsky, Cambridge University Press, Cambridge, U.K.

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 Paper #1175.

CHEN, X. (2006) "Large Sample Sieve Estimation of Semi-Nonparametric Models," forthcoming in *Handbook of Econometrics,* Vol. 6, edited by E. Leamer and J.J. Heckman.

CHERNOZHUKOV, V. AND C. HANSEN (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73, 245-261.

CHERNOZHUKOV, V., G. IMBENS, AND W. NEWEY (2004): “Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," mimeo, Department of Economics, M.I.T.

CHESHER, A. (2003): “Identification in Nonseparable Models,” *Econometrica*, Vol. 71, No. 5.

DAROLLES, S., J.P. FLORENS, and E. RENAULT (2002), ”Nonparametric Instrumental Regression,” mimeo, IDEI, Toulouse.

FAN, Y. and Q. LI (1996) "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica,* 64, 4.

FLORENS, J.P. (2003) "Inverse Problems and Structural Econometrics: The Example of Instrumental Variables," in *Advances in Economics and Econometrics, Theory and Applications,* Vol. II, edited by M. Dewatripont, L.P. Hansen, and S. Turnovsky, Cambridge: Cambridge University Press.

HALL, P. and J.L. HOROWITZ (2003), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," Center for Microdata Methods and Practice, WP 102/03.

HARDLE, W. and O. LINTON (1994), "Applied Nonparametric Methods," in *Handbook of Econometrics,* vol. 4, R.F. Engel ad D.F. McFadden, eds., Amsterdam: Elsevier, Ch. 38.

HECKMAN, J.J. and R. ROBB (1985) "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, edited by J.J. Heckman and B. Singer, Econometric Society Monograph 10, Cambridge University Press, Cambridge, U.K.

HONG Y. and H. WHITE (1995), "Consistent Specification Testing via Nonparametric Series Regression," *Econometrica*, 1133-1159.

IMBENS, G.W. AND W.K. NEWEY (2003) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," mimeo, M.I.T.

KOENKER, R.W. (2005) *Quantile Regression*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.

LINTON, O.B. and J.B. NIELSEN (1995) "A Kernel Method of Estimating Structured Nonparametric Regression Based n Marginal Integration," *Biometrika*, 82, 93-100.

MATZKIN, R.L. (1999) "Nonparametric Estimation of Nonadditive Random Functions," mimeo.

MATZKIN, R.L. (2003) "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica,* 71, 1339-1375.

MATZKIN, R.L. (2004) "Unobservable Instruments," mimeo, Northwestern University.

MATZKIN, R.L. (2005a) "Identification in Nonparametric Simultaneous Equations," mimeo, Northwestern University.

MATZKIN, R.L. (2005b) "Heterogenous Choice," invited lecture, Ninth World Congress of the Econometric Society, forthcoming in *Advanced in Economics and Econometrics, Theory and Applications, Ninth World Congress*, edited by Richard Blundell, Whitney Newey, and Torsten

Persson, Cambridge University Press.

MATZKIN, R.L. (2006) "Estimation of Nonparametric Simultaneous Equations," mimeo, Northwestern University.

McFADDEN, D. (1974) “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka, ed. *Frontiers in Econometrics,* pp105-142, New York: Academic Press.

NG, S. and J. PINKSE (1995) "Nonparametric Two-Step Estimation of Unknown Regression Functions whern the Regressors and the Regressin Error are not Independente," mimeo, CIREQ.

NEWEY, W.K. (1994) "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233-253.

NEWEY, W. and J. POWELL (1989) “Instrumental Variables Estimation of Nonparametric Models,” mimeo.

NEWEY, W. and J. POWELL (2003) “Instrumental Variables Estimation of Nonparametric Models,” *Econometrica,* 71, 5, 1565-1578.

NEWEY, W.K., J.L. POWELL, and F. VELLA (1999): “Nonparametric Estimation of Triangular Simultaneous Equations Models”, *Econometrica* 67, 565-603.

OLLEY, G.S. and A. PAKES (1996) "The Dynamics of Productivity in the Telecommunications Equipement Industry," *Econometrica,* 64, 6, 1263-1297.

PAGAN A. and A. ULLAH (1999) Nonparametric Econometrics, Cambridge University Press, Cambridge, U.K.

PINKSE, J. (2000): "Nonparametric Two-Step Regression Estimation when Regressors and Errors are Dependent," *Canadian Journal of Statistics,* 28-2, 289-300.

POWELL, J.L., J.H. STOCK, and T.M. STOKER (1989) "Semiparametric Estimation of Index Coefficients," *Econometrica*, 51, 1403-1430.

PRAKASA RAO, B.L.S. (1983) *Nonparametric Functional Estimation,* Academic Press.

ROEHRIG, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica,* 56:2, 433-447.

WOOLDRIDGE, J. (1992) "Nonparametric Regression Tests Based on an Infinite Dimensional Least Squares Procedure," *Econometric Theory*, 8, 435-451.

YATCHEW, A.J. (1992) "Nonparametric Regression Tests Based on an Infinite Dimensional Lease Squares Procedure," *Econometric Theory,* 8, 435-451.