

Control Functions and Simultaneous Equations Methods

By RICHARD BLUNDELL*, DENNIS KRISTENSEN AND ROSA L. MATZKIN

Economic models of agent's optimization problems or of interactions among agents often exhibit simultaneity. It is well known that any function in which an explanatory variable is partly determined by the dependent variable of the function, cannot be identified without additional information. Typically this additional information is provided by observable exogenous variables or functional structures. Here we focus on nonlinear simultaneous equations models with non-additive disturbances. Nonlinear non-additive specifications are a fundamental feature of economic models with unobserved heterogeneity, including multidimensional models of consumer and producer choice, models of Nash equilibrium in industrial organization and many models of labour market behavior.

In linear models with additive disturbances there are many alternative approaches to addressing endogeneity in estimation. Two-stage least squares, instrumental variable and control function approaches are three such commonly adopted approaches. In simple specifications of the linear simultaneous model all three approaches generate consistent, often analytically identical, estimators (Hausman (1987), for example). However, linear models with additive disturbances are highly restrictive representations of simultaneity in economic behavior. The marginal effect of endogenous variables is constant across all its values and homogenous across all individuals. Nonlinear non-additive models are more attractive but the conditions for the application of standard simultaneous equations estimators are quite different and somewhat more involved.

The instrumental variable approach in the nonlinear case typically proceeds by using an observable instrument for a single structural equation, independent of the structural errors

and functionally dependent with the endogenous explanatory variable. Newey and Powell (1989, 2003), Ai and Chen (2003), Hall and Horowitz (2003) and Darolles, Fan, Florens, and Renault (2011) follow this general approach.¹ Identification requires additional conditions on the relationship between the endogenous variable and the excluded instrument, and estimation involves solving an ill-posed inverse problem.

We take a different tack in this paper and consider approaches that specify equations for all endogenous variables determined within the system. We retain the focus on a single structural equation but by completing the simultaneous system, these approaches avoid the ill-posed inverse problem. They do however require the complete specification of the simultaneity between the endogenous variables.

A particularly convenient method, and one we direct our attention to here, is the control function approach (Heckman and Robb, 1985). This is available when the simultaneous system can be expressed in a triangular form where the variables entering satisfy certain conditional independence restrictions; the precise restriction is provided below. Linear simultaneous models with additive errors can always be expressed in triangular form with variables satisfying a conditional mean independence restriction. However, in nonlinear non-additive simultaneous systems, conditional independence requires an additional restriction, so-called *control function separability*. Blundell and Matzkin (2010) derive this condition and show that it completely characterizes simultaneous models where the control function approach can be used to estimate the structural function of interest. Chesher (2003), Imbens and Newey (2009), Hahn and Ridder (2011) and Kasy (2011), amongst others, develop identification and estimation results assuming triangularity. Here we review some of

* Department of Economics, UCL, and Institute for Fiscal Studies, email: r.blundell@ucl.ac.uk

Department of Economics, UCL, Institute for Fiscal Studies, email d.kristensen@ucl.ac.uk

Department of Economics, UCLA, email: matzkin@econ.ucla.edu.

¹See also Blundell, Chen and Kristensen (2007), Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007) and Chen and Pouzo (2012), Chen, Chernozhukov, Lee and Newey (2011).

the different ‘system’ approaches to estimation in nonlinear simultaneous equations models, focussing on the control function approach, and document the biases that can occur when inappropriately assuming a linear structural model.

We first characterize nonlinear simultaneous systems and briefly review different approaches to estimation. We contrast a number of specific approaches to estimation. In particular, those that use a conditional independence assumption with conditioning either on an observable variable or, as in the control function case, on an unobservable variable. We also note approaches that rely on exclusive instruments. In nonlinear simultaneous models, if the control function separability assumption is not satisfied, estimation using a control function approach may be highly misleading. A numerical example is used to show the potential biases from inappropriately adopting a control function approach.

I. Simultaneous Equation Systems

Consider the following simultaneous model describing the interactions between two outcome variables y_1 and y_2 ,

$$(1) \quad \begin{aligned} y_1 &= b_{10} + b_{11}y_2 + b_{12}z_1 + \varepsilon_1 \\ y_2 &= b_{20} + b_{21}y_1 + b_{22}z_2 + \varepsilon_2 \end{aligned}$$

satisfying $(1 - b_{11}b_{21}) \neq 0$. For ease of interpretation we assume all variables are scalar and the variables $\varepsilon_j \in \mathbb{R}$, and $z_j \in \mathbb{R}$ are continuously distributed, for $j = 1, 2$ and the z_j variables are exclusive to specific equations. The inverse form of the simultaneous system (1) may be written $\varepsilon_j = a_{j0} + a_{j1}y_1 + a_{j2}y_2 + a_{j3}z_j$ for $j = 1, 2$ with appropriate restrictions on the coefficients. Similarly, the reduced form can be derived explicitly and has a linear additive form.

Although convenient, linearity in endogenous variables implies that the marginal effect of the endogenous variable is constant across all its values. Moreover, the additivity in unobservables implies that the marginal effect of endogenous variables is homogeneous across individuals with the same $z = (z_1, z_2)$. If we wish to interpret the ε_j as unobserved heterogeneity relating to agents making the observed actions on y_1 and y_2 , such as in models for demand and supply or the interactions between agents, then it is very difficult to derive a separable structural model. Consequently, in models describ-

ing structural equations of economic behavior with unobserved heterogeneity we would typically wish to consider nonlinear non-additive systems of the form $y_1 = m^1(y_2, z_1, \varepsilon_1)$ and $y_2 = m^2(y_1, z_2, \varepsilon_2)$.

Parameters of Interest: In non-additive specifications we have to consider carefully the parameters of interest. Suppose we are interested in the feedback from y_2 to y_1 as described by the structural function m^1 . For this purpose we can ignore the presence of z_1 and write

$$(2) \quad \begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= m^2(y_1, z_2, \varepsilon_2) \end{aligned}$$

There are three parameters of interest we wish to highlight.

(a) The *average structural function* of m^1 , defined by Blundell and Powell (2003) as $G(y_2) = \int m^1(y_2, \varepsilon_1) f_{\varepsilon_1}(\varepsilon_1) d\varepsilon_1$.

(b) The *local average response function*, defined by Altonji and Matzkin (2005) as $\beta(y_2) = \int \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} f_{\varepsilon_1|Y_2=y_2}(\varepsilon_1) d\varepsilon_1$ which will be constant for all y_2 when m^1 is linear.

These two parameters refer to averages. We might also be interested in parameters defined at points in the distribution of unobservables. For example,

(c) The *quantile structural function*, $q^1(\tau, y_2)$, defined by Imbens and Newey (2009) as the τ th quantile of $m^1(y_2, \varepsilon_1)$ for y_2 fixed (so that the only source of randomness is ε_1). We may also be interested in derivatives of $q^1(\tau, y_2)$ w.r.t. y_2 ; see Chesher (2003) for identification results on this.

Characterizing the Simultaneous Model:

To characterize the simultaneous equations model we retain the simple structural system (2) where m^1 and m^2 are assumed to be continuously differentiable, and that, conditional on any value z_2 , the densities of $(\varepsilon_1, \varepsilon_2)$ and of (y_1, y_2) are continuous with convex support.

We make two further assumptions: (*Monotonicity*) the functions m^1 and m^2 are strictly monotone in ε_1 and ε_2 , respectively; (*Crossing*) $(\partial m^1 / \partial y_2)(\partial m / \partial y_1) < 1$. The monotonicity assumption guarantees that m^j can be inverted in ε_j , $j = 1, 2$. This assumption allows us to express the direct system of structural equations (2), defined by (m^1, m^2) , in terms of a structural inverse system of functions

$$(r^1, r^2)$$

$$(3) \quad \begin{aligned} \varepsilon_1 &= r^1(y_1, y_2) \\ \varepsilon_2 &= r^2(y_1, y_2, z_2) \end{aligned} \cdot$$

The crossing assumption is a weakening of the common situation where the value of the endogenous variables is determined by the intersection of a downwards and an upwards sloping function. Together with monotonicity, this assumption guarantees the existence of a unique value for (y_1, y_2) , given any z_2 . In other words, these assumptions guarantee the existence of a reduced form system of equations, defined by functions (h^1, h^2) , which map the vector of exogenous variables $(\varepsilon_1, \varepsilon_2, z_2)$ into the vector of endogenous variables (y_1, y_2)

$$(4) \quad \begin{aligned} y_1 &= h^1(z_2, \varepsilon_1, \varepsilon_2) \\ y_2 &= h^2(z_2, \varepsilon_1, \varepsilon_2) \end{aligned} \cdot$$

These assumptions also guarantee that the reduced form function h^j is monotone increasing in ε_j , $j = 1, 2$. These results are established in Blundell and Matzkin (2010, Lemma 1).

II. Approaches to Identification and Estimation

Suppose we are simply interested in the derivative of the unknown function m^1 in a structural model $y_1 = m^1(y_2, \varepsilon_1)$ where m^1 is strictly increasing in ε_1 and where it is suspected or known that y_2 is itself a function of y_1 .

The single equation approach proceeds by using an observable instrument, z_2 , independent of ε_1 and functionally dependent with y_2 . Identification takes the form of asking whether an integral equation has a unique solution. The answer requires restrictions on the conditional distribution of y_2 given z_2 . Estimation involves dealing with an ill-posed inverse problem. We refer to the references on nonparametric instrumental variables provided in the introduction for details.

The approaches we focus on in this paper involve describing the source of simultaneity, by specifying some function m^2 and unobservable ε_2 , such that (y_1, y_2) satisfies eq. (2), where m^2 is strictly increasing with respect to ε_2 and where, as in the single equation approach, z_2 is an observable variable that is excluded from m^1 . Identification in the system approach is analyzed in terms of conditions on the struc-

tural system composed by (m^1, m^2) and the distribution of $(\varepsilon_1, \varepsilon_2, z_2)$. Pointwise estimation of the derivative of m^1 with respect to y_2 can be performed without facing ill-posed inverse problems. Roehrig (1988), Benkard and Berry (2006), and Matzkin (2004, 2008, 2010) follow this approach.

A. Conditional Independence

One system approach to the identification and estimation of the derivative of the function m^1 in the system (2) proceeds by assuming that the system is observationally equivalent to an alternative system, of the form

$$(5) \quad \begin{aligned} y_1 &= m^1(y_2, \varepsilon_1) \\ y_2 &= s(z_2, \eta) \end{aligned} \cdot$$

The second equation is then used to determine a variable or sets of variables such that, conditioning the distribution of y_1 on y_2 and those additional variables has the effect of purging the dependence between y_2 and ε_1 . When such additional conditioning variables are observed, the method is called "conditioning on observables". When a conditioning variable is unobserved and is estimated in a first-stage, the estimated variable is usually called a "control function". In either case, the additional conditioning variable or variables can be interpreted as providing a proxy for the elements within ε_1 that are not distributed independently of y_2 . Conditioning on such "proxy" leaves the unobserved part of the equation independent of y_2 and the "proxy". (See Matzkin, 2004, for details.)

Conditioning on observables: Suppose that in the system (5) ε_1 is distributed independently of η conditional on z_2 . Given z_2 , $y_2 = s(z_2, \cdot)$ is a function of only η . Recall that when two random variables are independently distributed, any functions of those random variables are also independently distributed. Hence, η being independent of ε_1 given z_2 implies that y_2 , as a function of η , $y_2 = s(z_2, \cdot)$, is distributed independently of ε_1 , given z_2 . Hence, "conditioning on the observable z_2 " purges the dependence between y_2 and ε_1 .

Control function: Suppose instead that in the system (5) ε_1 is distributed independently of z_2 conditional on η . Then, in this case, the "proxy" that purges the dependence between y_2 and ε_2 is η . Conditional on η , $y_2 = s(\cdot, \eta)$ is a function

of z_2 . Since conditional on η , ε_1 is distributed independently of z_2 , the function $y_2 = s(\cdot, \eta)$ is also distributed independently of ε_1 , conditional on η . When the relation $y_2 = s(z_2, \eta)$ is such that η can be estimated, one can condition the first equation on the estimated η and proceed as if there existed no endogeneity. Chesher (2003) and Imbens and Newey (2009) follow this approach in the nonparametric nonadditive setup above. Conditions for pointwise identification and estimation of the derivatives of m^1 again avoid an ill-posed inverse problem. They obtain the expressions for m^1 and the derivative of m^1 as displayed in Section 2.1. They both exploit the indirect mapping from the observable variables (y_1, y_2, z_2) to the unobservable variables, $(\varepsilon_1, \eta) = (r^1(y_1, y_2), \tilde{s}(y_2, z_2))$.

When can a control function be used in models with simultaneity? If the simultaneous model m^1 and m^2 were linear and additive in ε_1 and ε_2 as in (1), the elements of m^1 could be easily identified (and estimated) recursively, following the control function approach. One can obtain the second equation in the form needed to estimate the control function by solving for y_2 in the first eq. of (1) and rewriting the system as a recursive one. This would yield (5) with $s(z_2, \eta)$ being linear and, conditional on η , y_2 is mean independent of ε_1 . As a consequence, estimation of the parameters in the structural equation for y_1 can proceed using the augmented regression approach (Dhrymes (1970), Telser (1964)). This *triangular* representation implies no further restrictions in the linear system.

If, however, the simultaneous equations model were either nonlinear or nonadditive in either ε_1 or ε_2 , it might not be possible to express the structural system (2) in the triangular form (5) with z_2 being independent of (ε_1, η) as needed for the implementation of the control function approach. So under what conditions are the two systems, (2) and (5), observationally equivalent?

Bundell and Matzkin (2010) provided an answer to this question. They defined a new concept, *control function separability* defined in terms of the indirect system (3):

Definition (CFS): *The structural inverse system (3) satisfies control function separability (CFS) if there exist functions $v : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $q : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that (a) $r^2(y_1, y_2, z_2) =$*

$v(q(z_2, y_2), r^1(y_1, y_2))$; (b) v is strictly increasing in its first argument, and; (c) q is strictly increasing in its second argument.

For CFS to be satisfied, $r^2(y_1, y_2, z_2)$ must be weakly separable in $r^1(y_1, y_2)$. Moreover, y_1 can affect the value of ε_2 only through $r^1(y_1, y_2)$. This implies that the reduced form for y_2 can be written as a function of the scalar unobservable control variable $\eta(\varepsilon_1, \varepsilon_2)$ and z_2 , where η is independent of z_2 . Under regularity conditions, Blundell and Matzkin (2010) establish that CFS completely characterizes observational equivalence between (2) and (5). In particular, if CFS holds, the functions m^1 and m^2 are identified and can be estimated using the two-step procedure described in Imbens and Newey (2009).

B. Exclusive Instruments

When the system of simultaneous equations is not observationally equivalent to one of the triangular systems considered in the previous section, identification can be obtained by imposing alternative restrictions on the system. One such set of restrictions is in terms of the inverse system (3) where, for example, we could impose the restriction that the second equation of (3) satisfies $\varepsilon_2 = r^2(y_1, y_2) + z_2$. In this case, Matzkin (2010) provides additional conditions on (r^1, r^2) and the distribution of $(\varepsilon_1, \varepsilon_2, z_2)$ under which, the derivative of m^1 with respect to y_2 can be read off the conditional density of the observable variables (y_1, y_2) conditional on the observable variable z_2 , at either one or two values of z_2 . The particular values of z_2 that provide such identification can also be read off from the conditional density of (y_1, y_2) given z_2 . Indirect estimation of such derivative is then obtained by substituting the conditional density by a nonparametric estimator for it. Under similar conditions, a minimum distance estimator for such derivative is also obtained, which has a closed form solution.

III. Numerical Illustration

Here we develop a simple bivariate nonlinear nonadditive system with inverse equations $\varepsilon_1 = \Lambda_1^{-1}(y_2) - \Lambda_2^{-1}(y_1)$, implying $m^1(y_2, \varepsilon_1) = \Lambda_1(\Lambda_2^{-1}(y_2) + \varepsilon_1)$, and $\varepsilon_2 = (1 - b_1(z_2))/b_2(z_2) \Lambda_1^{-1}(y_2) -$

$b_1(z_2)/b_2(z_2)\Lambda_1^{-1}(y_1) = g_2(z_2)$ where $\Lambda_1(x) = x^3$, $\Lambda_2(x) = \Phi(x)$ is the standard normal cdf, and $g(z_2) = z_2^2/2 - 1$. The functions $b_j(z_2)$ will be specified below. It is easily checked that the system is invertible if $b_2(z_2) \neq 0$ and that CFS is satisfied if, for some constant $c \in \mathbb{R}$, $b_1(z_2) = cb_2(z_2)$, in which case $y_2 = s(z_2, \eta) = \Lambda_2(g_2(z_2) + b_2(z_2)\eta)$, where $\eta = c\varepsilon_1 + \varepsilon_2$. In particular, y_2 is endogenous if $c \neq 0$. We then investigate how control function estimators of m^1 perform when CFS is satisfied or not, respectively.

We focus exclusively on biases incurred by different estimators and so choose the sample size to be $n = 10,000$ such that the variance of the different estimators can be ignored. For the two estimators in question, we evaluate their performance by plotting the true quantile structural function which takes the form $q^1(\tau, y_1) = \Lambda_1(\Lambda_2^{-1}(y_2) + q_{\varepsilon_1}(\tau))$, where $q_{\varepsilon_1}(\tau)$ is τ th quantile of ε_1 , against the ones implied by the estimators. The quantile structural functions are evaluated at three different quantiles, $\tau_1 = 0.2$, $\tau_2 = 0.5$ and $\tau_3 = 0.8$.

We first consider the case where CFS is satisfied with $b_2(z_2) = \sqrt{1/9 + 1/90 \times z^2}$ and $b_1(z_2) = 2b_2(z_2)$. Figure 1 plots the nonparametric structural quantile estimator of Imbens and Newey (2009) for this case. Here and in the following figures the drawn lines represent the true function and dotted lines the estimates for three different quantiles. We find that this estimator does well and closely tracks the population versions.

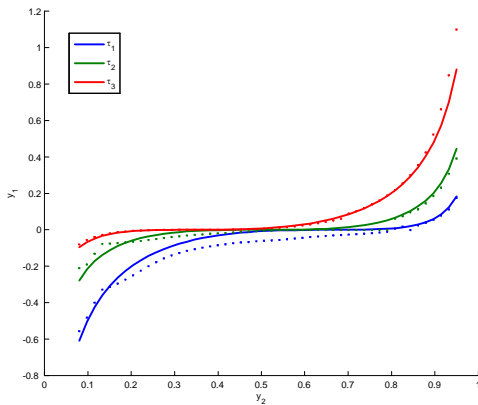


Figure 1. CFS holds

Next, we investigate how the Imbens and Newey (2009) estimator does when CFS does not hold in data. For this case,

$b_2(z_2)$ remains unchanged but now $b_1(z) = \sqrt{0.5 + 0.05 \times z^4 + \log(1 + z^2)}/10$. Figure 2 shows the performance of the Imbens-Newey estimator in this case. Biases are now present, in particular, in the upper quantiles.

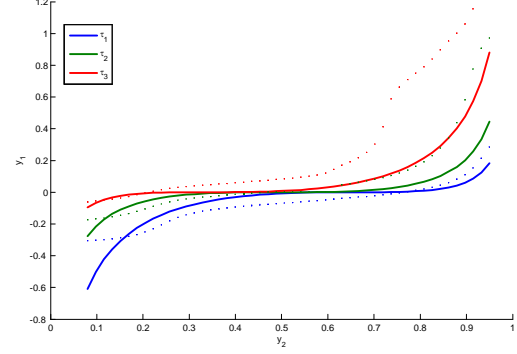


Figure 2. CFS is violated

Finally, Figure 3 shows the mean of the estimated quantile structural function based on the control function estimator assuming (incorrectly) that data is generated by a linear regression model, $\hat{q}_{2SLS}^1(\tau, y_1) = \hat{b}_{10} + \hat{b}_{11}y_2 + \hat{q}_{\varepsilon_1}(\tau)$, where \hat{b}_{10} and \hat{b}_{11} are two-stage least squares (2SLS) estimates and $\hat{q}_{\varepsilon_1}(\tau)$ is the τ quantile of the residuals from the 2SLS regression. We see that the 2SLS estimator is severely biased and over-estimates of the impact of y_2 on y_1 . This is not surprising given the highly non-linear features of the data-generating process which is ignored by the 2SLS estimator. In particular, the data-generating process has fat tails which affects predictions based on covariances but not quantiles.

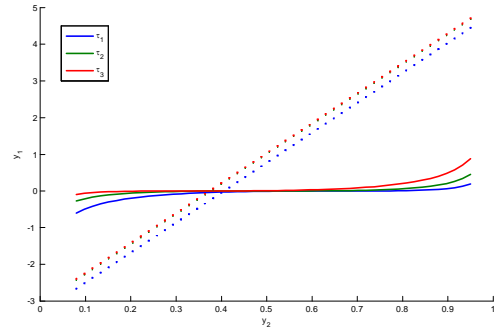


Figure 3. CFS holds, misspecified model used

IV. Summary and Conclusions

We have reviewed two broad alternative classes of ‘system’ approaches to estimation in nonlinear non-additive simultaneous equations

models. The first use a conditional independence assumption and second adopt exclusion restrictions on instrumental variables. In particular, we have focussed on a special case of the second approach, the *control function*, noting that in nonlinear models estimation using the control function estimator may be severely inconsistent unless strong restrictions on the simultaneous model are satisfied. A simulation model was then used to show the potential biases from inappropriately adopting a control function approach.

REFERENCES

- Altonji, J. and R.L. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053-1102.
- Benkard, C.L. and S.T. Berry (2006) "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models" , *Econometrica*, 74, 1429-1440.
- Blundell, R., X. Chen and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613-1669.
- Blundell, R. and R.L. Matzkin (2010): "Conditions for the Existence of Control Functions in Nonadditive Simultaneous Equation Models," mimeo, UCL and UCLA.
- Blundell, R. W. and Powell, J. L. (2003), "Endogeneity in Nonparametric and Semiparametric Regression Models", in M. Dewatripont, L. P. Hansen and S. J. Turnovsky (eds.) *Advances in Economics and Econometrics*, Vol.II CUP.
- Chen, X., V. Chernozhukov, S. Lee and W.K. Newey (2011): "Local Identification of Nonparametric and Semiparametric Models," mimeo, MIT.
- Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.
- Chernozhukov, V., G. Imbens and W.K. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics* 139, Pages 4-14.
- Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica* 71, 1405-1441.
- Darolles, S., Y. Fan, J.P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541-1565.
- Dhrymes, P. J. (1970), *Econometrics*. New York: Springer-Verlag.
- Hahn J.A. and G. Ridder (2011): "Conditional Moment Restrictions and Triangular Simultaneous Equations", *Review of Economics and Statistics*, 93, 683-689.
- Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics* 33, 2904-2929.
- Hausman, J.A. (1987) "Specification and Estimation of Simultaneous Equation Models", in *Handbook of Econometrics*, Volume I (eds. Z. Griliches and M.D. Intriligator), North-Holland.
- Heckman, J. J. and R. Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, (eds J. Heckman and B. Singer), CUP.
- Imbens, G. and W.K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica* 77, 1481-1512.
- Kasy, M. (2011): "Identification in triangular systems using control functions," *Econometric Theory*, 27, 663-671.
- Matzkin, R.L. (2004) "Unobservable Instruments," mimeo, Northwestern University.
- Matzkin, R.L. (2008) "Identification in Nonparametric Simultaneous Equations," *Econometrica*, Vol. 76, No. 5, 945-978.
- Matzkin, R.L. (2010) "Estimation of Nonparametric Models with Simultaneity," mimeo, UCLA
- Newey, W. and J. Powell (2003) "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*.
- Roehrig, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433-447.
- Telser, L. G. (1964), "Iterative Estimation of a Set of Linear Regression Equations," *Journal of the American Statistical Association*, 59, 845-862.