13 Endogeneity and Nonparametric IV

13.1 Nonparametric Endogeneity

A nonparametric IV equation is

$$\begin{aligned} Y_i &= g(X_i) + e_i \qquad (1) \\ E(e_i \mid Z_i) &= 0 \end{aligned}$$

In this model, some elements of X_i are potentially endogenous, and Z_i is exogenous.

We have studied this model in 710 when g is linear. The extension to nonlinear g is not obvious.

The first – and primary – issue is identification. What does g mean?

Let

$$\lambda(z) = E(Y_i \mid Z_i = z)$$

and take the conditional expectation of (1):

$$\begin{aligned} \lambda(z) &= E(Y_i \mid Z_i = z) \\ &= E(g(X_i) + e_i \mid Z_i = z) \\ &= E(g(X_i) \mid Z_i = z) \\ &= \int g(x) f(x \mid z)\, dx. \end{aligned}$$

The functions $\lambda(z)$ and $f(x \mid z)$ are identified. The unknown nonparametric $g(x)$ is a solution to the integral equation

$$\lambda(z) = \int g(x) f(x \mid z)\, dx.$$

The difficulty is that the solution $g(x)$ is not necessarily unique. The mathematical problem is that the solution g is not necessarily continuous in the function f. The non-uniqueness of g is called the "ill-posed inverse problem".

A solution is to restrict the space of allowable functions g. For example, the linear model $g(x) = x'\beta$ is linear, then the above equation reduces to

$$\begin{aligned} \lambda(z) &= \beta' \int x f(x \mid z)\, dx \\ &= \beta' E(X_i \mid Z_i = z). \end{aligned}$$

Idenfication of β in the linear model exploits this simple relationship.

13.2 Newey-Powell-Vella's Triangular Simultaneous Equations

Newey, Powell, Vella (1989, Econometrica)

The model is

$$\begin{aligned} Y_i &= g(X_i) + e_i \\ E(e_i \mid Z_i) &= 0 \end{aligned} \tag{2}$$

plus a reduced-form equation for X_i

$$\begin{aligned} X_i &= \Pi(Z_i) + u_i \\ E(u_i \mid Z_i) &= 0 \end{aligned} \tag{3}$$

Thus $\Pi(z)$ is the conditional mean of X_i given $Z_i = z$. The vectors X_i and Z_i may overlap, so X_i can contain both endogenous and exogenous variables.

NPV then take expectations of (2) given X_i and Z_i

$$\begin{aligned} E(Y_i \mid X_i, Z_i) &= E(g(X_i) + e_i \mid X_i, Z_i) \\ &= g(X_i) + E(e_i \mid X_i, Z_i) \end{aligned}$$

Since X_i is endogenous, the latter conditional expectation is not zero. In general, this cannot be simplified further. But NPV observe the following. From (3), X_i is a function of Z_i and u_i, so conditioning on X_i and Z_i is equivalent to conditioning on u_i and Z_i. Hence

$$E(Y_i \mid X_i, Z_i) = g(X_i) + E(e_i \mid u_i, Z_i)$$

Next, suppose that Z_i is strongly exogenous in the sense that

$$E(e_i \mid u_i, Z_i) = E(e_i \mid u_i) = g_2(u_i)$$

That is, conditional on u_i, Z_i provides no information about the mean of the error e_i. In this case we have the simplification

$$E(Y_i \mid X_i, Z_i) = g(X_i) + g_2(u_i)$$

which implies

$$\begin{aligned} Y_i &= g(X_i) + g_2(u_i) + \varepsilon_i \\ E(\varepsilon_i \mid u_i, X_i) &= 0 \end{aligned}$$

This is an additive regression model, with the regressor u_i unobserved but identified.

Is g identified?

Since (2) is a (reduced-form) regression, Π is identified. Thus u_i is identified.

Then the functions g and g_2 are identified so long as X_i and u_i are distinct. NPV discussed several identification conditions. One is:

Theorem: If there is no functional relationship between x and u, then $g(x)$ is identified up to an additive constant.

The additive constant qualitification is required in all additive nonparametric models.

The authors propose the following series estimator:

1. Estimate $\hat{\Pi}_L(z) = \hat{\theta}'_L Z_{Li}$ using a series in Z_i with L terms, say
    (a) $\hat{\theta}_L = (Z'_L Z_L)^{-1} Z'_L X$ where Z_L are the basic functions of Z
    (b) Residual $\hat{u}_i = X_i - \hat{\Pi}(X_i)$
2. Create a basis transformation for X_i, and a separate one for $\hat{u}_i$, with K coefficients.
    (a) spline functions of X_i
    (b) spline functions of $\hat{u}_i$,
3. Least-squares regression of Y_i on these basis functions, obtain $\hat{g}$ and $\hat{g}_2$

NPV show that this estimator is consistent and asymptotically normal. The conditions require that the functions g and Π be sufficiently smooth (enough derivatives), and that the number of terms K and L diverge to infinity in a controlled way. The regularity conditions are not particularly helpful.

It is not clear how K and L should be selected in practice. A reasonable suggestion is to select L by cross-validation on the reduced form regression, and then select K by cross-validation on the second-stage regression. The trouble is that the two stages are not orthogonal, so the MSE of the second stage is affected by the first stage, so it is unlikely that the CV criterion will correctly reflect this.

We are primarily interested in the estimate $\hat{g}$ of g (the structural form equation). The estimates of $(\hat{g}, \hat{g}_2)$ in the second stage are asymptotically normal, but affected by the first stage (the generated regressors problem).

One solution is to write out the correct asymptotic covariance matrix for the two-step estimator as discussed in NPV

The other, easier approach is to view the problem as a one-step estimator. Stack the moment equations from each step. Then the two-step estimates are equivalent to a one-step estimator – just-identified GMM on the stacked equations. The covariance matrix may be calculated for the estimates using the standard GMM formula.

The authors include an application to wage/hours profile.

13.3 Newey and Powell's Estimator

Newey and Powell (Econometrica, 2003) propose a nonparametric method which avoids the strong exogeneity assumption, but imposes restrictions on allowable g.

Return to the base model

$$\begin{aligned} Y_i &= g(X_i) + e_i \\ E(e_i \mid Z_i) &= 0 \end{aligned}$$

and tthe integral equation

$$E(Y_i \mid Z_i) = \int g(x) f(x \mid Z_i)\, dx$$

To identify g, NP point out that one solution is to assume that g lives in a compact space. Their paper is based on this assumption, and impose this on their estimates of g.

Next, suppose that $g(x)$ can be approximated using a series approximation. Write this as

$$g(x) \simeq g_K(x) = \gamma_K' p_K(x)$$

where γ_K is a K vector of parameters and $p_K(x)$ is a K vector of basis functions. Compactness of g can be impose by assuming that γ_K is bounded. They use $\gamma_K' W_K \gamma_K \leq C$ where W_K is a specific weight matrix and C is a pre-determined constant.

Substituting the series expasion into the integral equation,

$$\begin{aligned} E(Y_i \mid Z_i) &\simeq \gamma_K' \int p_K(x) f(x \mid Z_i)\, dx \\ &= \gamma_K' E(p_K(X_i) \mid Z_i) \\ &= \gamma_K' h_K(Z_i) \end{aligned}$$

where

$$h_K(z) = E(p_K(X_i) \mid Z_i = z)$$

is the K vector of conditional expectations of the basis function transformations of X_i.

We thus have the regression models

$$\begin{aligned} Y_i &= \gamma_K' h_K(Z_i) + v_i \\ E(v_i \mid Z_i) &= 0 \end{aligned}$$

and

$$\begin{aligned} p_K(X_i) &= h_K(z) + \eta_i \\ E(\eta_i \mid Z_i) &= 0 \end{aligned}$$

NW suggest a two-step estimator.

1. Select the basis functions $p_K(x)$
2. Non-parametrically regress each element of $p_K(x)$ on Z_i using series methods. The estimates are collected in to the vector $\hat{h}_K(z)$.
3. Regress Y_i on $\hat{h}_K(z)$ (least squares) to obtain $\hat{\gamma}_K$.
4. The estimate of interest is $\hat{g}(x) = \hat{\gamma}_K' p_K(x)$

Identification requires that g (and thus $\hat{g}$) satisfy a compactness condition. NW recommend that this be imposed on $\hat{g}$ by restricting the estimate $\hat{\gamma}_K$ to satisfy $\hat{\gamma}_K' W_K \hat{\gamma}_K \leq C$. (This can be easily imposed by constrained LS regression.) As the constant C is arbitrary it is unclear what this means in practice.

NW discuss conditions for consistency of $\hat{g}$.

The estimator $\hat{\gamma}_K$ is a two-step estimator in the generate regressor class. Thus the conventional standard errors for $\hat{\gamma}_K$ (and thus $\hat{g}$) are incorrect.

The method is extended and applied to Engel curve estimation by Blundell, Chen and Kristensen (Econometrica, 2007). They extend the analysis of identification, and include a proof of asymptotic normality, how to calculate standard errors, and computational implication issues.

Other important related papers are referenced include Hall and Horowitz (Annals of Statistics, 2005), and Darolles, Florens and Renault, "Nonparametric Instrumental Regression" (early version 2002, current version 2009, still unpublished).

This general topic is clearly very important to econometrics and underdeveloped.

13.4 Ai and Chen (Econometrica, 2003)

Take the conditional moment restriction model

$$E\left[\rho\left(Z, \alpha_0\right) \mid X\right] = 0$$

(Notationally, we have switched the Z and X from the previous section, and I do this to follow Li-Racine, who simply followed the notation in the original papers.) Here, Z is the "endogenous" variables, and X are exogenous. The function ρ is a residual (or moment) equation). E.g. $\rho(Z) = y - z_1'\alpha$ in a linear framework). The are interested in the semiparametric framework in which $\alpha = (\theta, g)$ where θ is parametric and g is nonparametric. Their focus is on efficient estimation of the parametric component θ. (In this sense their work takes a different focus from the papers earlier reviewed, which focused on the nonparametric component.)

An example is the partially linear regression model with an endogenous regressor, where the focus is on the parametric component.

For the moment consider estimation of α assuming that it is parametric

Define the conditional mean and variance of $\rho(Z,\alpha)$ for generic values of α :

$$\begin{aligned} m(x,\alpha) &= E\left[\rho(Z,\alpha) \mid X=x\right] \\ \sigma^2(x) &= var\left[\rho(Z,\alpha) \mid X=x\right] \end{aligned}$$

Note that at the true value α_0

$$m(x,\alpha_0) = 0$$

for all x.

If the functions m and σ^2 were known and α were parametric, a reasonable estimator for α would be found by minimizing the squared error criterion

$$\sum_{i=1}^{n} \frac{m(X_i,\alpha)^2}{\sigma^2(X_i)}.$$

For simplicity, suppose $\sigma^2(x) = 1$, then the criterion simplifies to

$$\sum_{i=1}^{n} m(X_i,\alpha)^2 = m(\alpha)' m(\alpha)$$

where $m(\alpha)$ is the vector of stacked $m(X_i,\alpha)$.

As m is unknown this is infeasible. We can replace m with an estimate.

For any fixed α estimate m by a series regression. That is, approximate

$$m(x,\alpha) \simeq p_K(x)'\pi_K(\alpha)$$

where $p_K(x)$ is a K vector of basis functions in x. Then

$$m(\alpha) = P\pi_K(\alpha)$$

where P is the matrix of the regressors $p_K(X_i)$.

Let $\rho(\alpha)$ be the vector of stacked $\rho(Z_i,\alpha)$.

We estimate $\pi_K(\alpha)$ by LS of $\rho(\alpha)$ on P :

$$\hat{\pi}_K(\alpha) = \left(P'P\right)^{-1} P'\rho(\alpha)$$

The estimate of $m(\alpha)$ is

$$\hat{m}(\alpha) = P\hat{\pi}_K(\alpha) = P\left(P'P\right)^{-1} P'\rho(\alpha)$$

and the squared error criterion is estimated by

$$\begin{aligned}\hat{m}(\alpha)'\hat{m}(\alpha) &= \rho(\alpha)'P\left(P'P\right)^{-1}P'P\left(P'P\right)^{-1}P'\rho(\alpha) \\ &= \rho(\alpha)'P\left(P'P\right)^{-1}P'\rho(\alpha)\end{aligned}$$

which is a GMM criterion. If α were parametric, it could be estimated by minimizing this criterion. Indeed this is conventional GMM using the instrument set P under the assumption of homoskedasticity.

Now, as $\alpha = (\theta, g)$ includes a nonparametric component, Ai and Chen suggest replacing g by a series approximation:

$$g(z) \simeq q_L(z)'\beta_L$$

where $q_L(z)$ is an L vector of basis functions in z.

The moment equation

$$\rho(z, \theta, g(z)) \simeq \rho\left(z, \theta, q_L(z)'\beta_L\right)$$

is then a function of the parameters θ and β_L. For fixed (θ, β_L) define the $n \times 1$ vector $\rho(\theta, \beta_L)$ of stacked elements $\rho(Z_i, \theta, q_L(Z_i)'\beta_L)$. Replacing $\rho(\alpha)$ with $\rho(\theta, \beta_L)$ we have the revised GMM criterion

$$\begin{aligned}&\rho(\theta, \beta_L)'P\left(P'P\right)^{-1}P'\rho(\theta, \beta_L) \\ =\ &\sum_{i=1}^{n}\rho\left(Z_i, \theta, q_L(Z_i)'\beta_L\right)p_K(X_i)\left(\sum_{i=1}^{n}p_K(X_i)p_K(X_i)'\right)^{-1}\sum_{i=1}^{n}p_K(X_i)\rho\left(Z_i, \theta, q_L(Z_i)'\beta_L\right)\end{aligned}$$

The estimates $(\hat{\theta}, \hat{\beta}_L)$ then minimize this function.

This is not Ai and Chen's preferred estimator. For efficiency, they suggest estimating $\hat{\sigma}^2(X_i)$, the conditional variance, and using the weighted criterion.

This criterion is

$$\sum_{i=1}^{n}\frac{\left(p_K(X_i)'\left(P'P\right)^{-1}P'\rho(\theta, \beta_L)\right)^2}{\hat{\sigma}^2(X_i)} = \rho(\theta, \beta_L)'P\left(P'P\right)^{-1}\left(P'\hat{D}^{-1}P\right)\left(P'P\right)^{-1}P'\rho(\theta, \beta_L).$$

This is GMM with the eficient weight matrix $\left(P'P\right)^{-1}\left(P'\hat{D}^{-1}P\right)\left(P'P\right)^{-1}$.

Ai and Chen demonstrate that the estimate $\hat{\theta}$ is root-n asymptotically normal and asymptotically efficient (in the semiparametric sense)