Semi and Nonparametric Models in Econometrics

Part 3: IV methods in nonparametric and/or nonlinear models

Xavier d'Haultfoeuille
CREST-INSEE

Outline

Outline

IV models

- So far we have seen many techniques for estimating parametric or semiparametric nonlinear models with exogenous regressors:

$$Y = \varphi(X, \varepsilon), \; q(f_{X,\varepsilon}) = 0.$$

where $f_{X,\varepsilon}$ is the density of (X, ε) and q is a restriction such as
  - Mean-independence: $q(f_{X,\varepsilon})(x) = \int u f_{X,\varepsilon}(x, u) du$;
  - Quantile-independence: $q(f_{X,\varepsilon})(x) = F^{-1}_{\varepsilon|X}(\tau|x)$...
- Can we extend our results to models where X is endogenous, but instruments Z are available, so that:

$$Y = \varphi(X, \varepsilon), \; q(f_{Z,\varepsilon}) = 0 \;?$$

A search for "universal solution"

- The linear model, where the situation is simple, provides insights on general solutions to handle IV estimation in more complex cases.
- In the linear case, three equivalent ways can be used to define β_0.
- Two of them will extend to nonlinear and/or nonparametric models. However, they are not equivalent anymore, neither in terms of identification nor of estimation.
- We consider hereafter nonparametric models. In general, semiparametric identification / estimation can be easily treated as particular cases.

Outline

Introduction

- Consider the IV linear model:

$$Y = X'\beta_0 + \varepsilon, \quad E(Z\varepsilon) = 0.$$

- In this model, there are three equivalent ways to define β_0:
  1. through a projection;
  2. through an estimating equation;
  3. through a control variable approach.

Projection

- In the first, we project linearly X on Z:

$$X = \Gamma_0 Z + \nu, \text{ with } E(Z\nu) = 0$$

- Then, instead of regressing Y on X, we regress Y on $\widehat{X} = \Gamma_0 Z$:

$$\begin{aligned} Y &= X'\beta_0 + \varepsilon \\ &= \widehat{X}'\beta_0 + \nu'\beta_0 + \varepsilon. \end{aligned}$$

- In this regression, $\widehat{X}$ is exogenous because $E(Z\nu) = 0$ and $E(Z\varepsilon) = 0$.

- Identification is ensured as soon as the regressors $\widehat{X}$ are linearly independent. One can show that this is equivalent to $E(ZX')$ being full rank.

Estimating equation

- The second way is to write:

$$E(ZY) = E(ZX')\beta_0,$$

  and solve the linear equation to find β_0.
- Identification directly follows from the standard condition of $E(ZX')$ being full rank.
- Note that this strategy is not considered for estimating β_0 because this is computationally more demanding (we cannot rely on OLS estimation) and because it does not lead to a proper estimator when the model is overidentified (since the system $\widehat{E}(ZY) = \widehat{E}(ZX')\beta$ has usually no solution in β).

The control variable approach

- The third way to identify β_0 is to project ε on $\nu(= X - \Gamma_0 Z)$:
$$\varepsilon = \nu'\delta_0 + \zeta, \text{ with } E(\nu\zeta) = 0.$$
- Then we regress Y on (X, ν):
$$Y = X'\beta_0 + \nu\delta_0 + \zeta.$$
- Regressors are exogenous because $E(\nu\zeta) = 0$ and
$$\begin{aligned} E(X\zeta) &= E((\Gamma_0 Z + \nu)\zeta) \\ &= \Gamma_0 E(Z(\varepsilon - \nu\delta_0)) \\ &= 0. \end{aligned}$$
- The intuition behind is that by exogeneity of Z, the residual ν contains all the endogeneity of X. Once we control for ν in the regression, X is exogenous.
- Identification is ensured as soon as X and ν are not linearly dependent (equivalent to $E(ZX')$ being full rank).

Outline

Why this fails in general

- Consider a general model $Y = \varphi(X, \varepsilon)$ with $q(f_{Z,\varepsilon}) = 0$. Then $Y = \varphi(\widehat{X} + \nu, \varepsilon)$ but in general there exists no ζ such that
$$\varphi(\widehat{X} + \nu, \varepsilon) = \varphi(\widehat{X}, \zeta), \text{ with } q(f_{\widehat{X},\zeta}) = 0.$$
- This works in the linear model where $\varphi(X, \varepsilon) = X'\beta_0 + \varepsilon$ but not in general when φ is nonlinear.
- Example: suppose that $\varphi(X, \varepsilon) = \alpha_0 + X\beta_0 + X^2\gamma_0 + \varepsilon$ and the regression of X on Z is heteroskedastic:
$$X = \widehat{X}(1 + \widetilde{\nu}), \quad \text{with } \widetilde{\nu} \perp\!\!\!\perp \widehat{X}.$$
Then:
$$Y = \alpha_0 + \widehat{X}\beta_0 + \widehat{X}^2\gamma_0 + \left[\widehat{X}^2\widetilde{\nu}^2\gamma_0 + \left(\varepsilon + \widehat{X}(\beta_0 + 2\widehat{X}\gamma_0)\widetilde{\nu}\right)\right]$$
The first term into the brackets is correlated with $\widehat{X}$ and induces a bias in the regression.

Another example

- In quantile IV models, some people propose (1) to regress X on Z and (2) to run a quantile regression of Y on the projector $\widehat{X}$.
- However, this is valid only under the very weird condition

$$q_\tau(\varepsilon_\tau + (X - \widehat{X} - q_\tau(X - \widehat{X}))\beta_\tau | Z) = 0,$$

- This does not hold in general, even when $q_\tau(\varepsilon_\tau | Z) = 0$ and $q_\tau(X - \widehat{X}|Z) = q_\tau(X - \widehat{X})$ because in general,

$$q_\tau(U + V) \neq q_\tau(U) + q_\tau(V).$$

Outline

A first generalization to additive model

- Instead of $Y = X'\beta_0 + \varepsilon$, consider the nonparametric additive model (see Newey and Powell, 2003, and Darolles et al., 2005)

$$Y = \varphi(X) + \varepsilon, \text{ with } E(\varepsilon|Z) = 0.$$

- Then one can identify $\varphi(.)$ throught the estimating equation:

$$E(Y|Z) = E(\varphi(X)|Z)$$

or, equivalently, the conditional moment condition

$$E(Y - \varphi(X)|Z) = 0.$$

A first generalization to additive model

- The identifying condition is

$$E(g(X)|Z) = 0 \Rightarrow g(X) = 0. \tag{1}$$

- This is known as the *completeness condition* (because of the link with complete statistics).
- Condition (1) is far less intuitive than in the linear case. Suppose for instance that $X = Z + U$:
    - Then if $U \sim \mathcal{N}(0, \sigma^2)$, the completeness condition holds;
    - But if $U \sim \mathcal{U}[-1/2, 1/2]$, it fails to hold because there are periodic functions for which

$$\int_{-1/2}^{1/2} g(z+u)du = 0 \quad \forall z.$$

- Not much is known about this condition: see Newey and Powell (2003) and D'Haultfœuille (2009) for sufficient conditions.

A first generalization to additive model

- Note that this model is not well suited when Y is limited, and $Y = g(\mu(X) + \varepsilon)$. On the other hand, X can be limited.
- As for estimation, this is a rather difficult problem since we have to solve an infinite dimensional inverse problem.
- A simple solution is to rely on sieve estimation: consider for instance polynomials of X and solve for the moment conditions:

$$E\left[Z^k\left(Y - \sum_{j=0}^{k_n} \lambda_j X^j\right)\right] = 0 \text{ with } 1 \leq k \leq K(\geq k_n).$$

A second generalization to nonadditive model

- Consider a nonadditive model (see Chernozhukov and Hansen, 2005):

$$Y = \varphi(X, \varepsilon), \quad Z \perp\!\!\!\perp \varepsilon, \tag{2}$$

and $\varphi(x, .)$ is strictly increasing.

- We can suppose without loss of generality (provided that ε is continuous) that $\varepsilon \sim \mathcal{U}[0, 1]$.
- Then, for all $\tau \in (0, 1)$,

$$\begin{aligned} \tau &= P(\varepsilon \leq \tau) = P(\varepsilon \leq \tau | Z) \\ &= P(\varphi(X, \varepsilon) \leq \varphi(X, \tau) | Z) = P(Y \leq \varphi(X, \tau) | Z). \end{aligned}$$

- Thus, $\varphi(., \tau)$ solves the conditional moment conditions:

$$E\left(\mathbb{1}\{Y \leq \varphi(X, \tau\} - \tau | Z\right) = 0.$$

A second generalization to nonadditive model

- Semiparametric example: quantile IV model. Suppose that

$$Y = X'\beta_\varepsilon, \text{ with } \varepsilon \perp\!\!\!\perp Z, \; \varepsilon \sim \mathcal{U}[0,1] \text{ and } u \mapsto x'\beta_u \text{ increasing.}$$

- Then

$$E\left[\mathbb{1}\{Y \leq X'\beta_\tau\}|Z\right] = E\left[\mathbb{1}\{\varepsilon \leq \tau\}|Z\right] = \tau.$$

- Model (2) generalizes the previous one but still cannot handle limited Y. For a binary threshold model for instance, $Y = \mathbb{1}\{X'\beta_0 + \varepsilon \geq 0\}$ so that $\varphi(x, \varepsilon) = \mathbb{1}\{x'\beta_0 + \varepsilon \geq 0\}$ is not strictly increasing in ε.
- Estimation is still more difficult than previously because $g \mapsto E\left(\mathbb{1}\{Y \leq g(X, \tau)\} - \tau|Z\right)$ is not linear: see Horowitz and Lee (2007) for details.

Outline

A first generalization to additive model

- Consider the model (see Newey, Powell and Vella, 1999):

$$\left\{\begin{array}{lcl} Y & = & \varphi(X)+\varepsilon \\ X & = & \psi(Z)+\nu \end{array}\right. \quad Z \perp\!\!\!\perp (\nu,\varepsilon).$$

- Then, because $Z \perp\!\!\!\perp (\varepsilon,\nu)$ implies $X \perp\!\!\!\perp \varepsilon|\nu$:

$$\begin{array}{rcl} E(Y|X,\nu) & = & \varphi(X)+E(\varepsilon|\nu,X) \\ & = & \varphi(X)+E(\varepsilon|\nu,Z) \\ & = & \varphi(X)+E(\varepsilon|\nu). \end{array}$$

- We can recover Y by 1) regressing nonparametrically X on Z to obain ν and 2) regress nonparametrically Y on (X,ν).
- Note that we still cannot use it for limited Y...
- Compared to the previous approach, we require more on the instrument: $Z \perp\!\!\!\perp \varepsilon$ is replaced by $Z \perp\!\!\!\perp (\varepsilon,\nu)$.

A first generalization to additive model

- To estimate the second step of this model, we have to recover φ in the nonparametric additive model $E(Y|X,\nu) = \varphi(X) + g(\nu)$.
- A first point is that a normalization is needed since $\varphi(X) + g(\nu) = [\varphi(X) - c] + [g(\nu) + c]$. Suppose w.l.o.g. that $E(g(\nu)) = 0$.
- A first solution is *marginal integration*, which is based on the following equality:

$$\int E(Y|X,\nu = u)dF_\nu(u) = \varphi(X) + E(g(\nu)) = \varphi(X).$$

  Then: 1) estimate by a kernel estimator $E(Y|X,\nu)$ and 2) define $\widehat{\varphi}(.)$ by:

$$\widehat{\varphi}(x) = \frac{1}{n}\sum_{i=1}^{n} \widehat{E}(Y|X = x, \nu = \nu_i)$$

- Another solution is to rely on sieves: regress Y on (separate) functions of X and functions of ν.

A second generalization to nonadditive model

- Consider the nonadditive model (see Imbens and Newey, 2009):

$$\left\{ \begin{array}{rcl} Y & = & \varphi(X,\varepsilon) \\ X & = & \psi(Z,\nu) \end{array} \right. \qquad Z \perp\!\!\!\perp (\nu,\varepsilon)$$

- Suppose also that $\psi(Z,.)$ is strictly increasing. Without loss of generality, we can suppose that ν is uniform.
- In this model it is difficult to recover φ directly. However, we can recover other quantities of interest, e.g. the *quantile structural function* $\tau \mapsto q_\tau(Y,x)$, i.e. the quantile function of $\varphi(x,\varepsilon)$, under the assumption that

$$\text{Supp}(\nu|X) = \text{Supp}(\nu).$$

- This condition is restrictive. It imposes in particular a rank condition between X and Z and a large support condition on Z.

A second generalization to nonadditive model

- First, let $\psi_2^{-1}(z,.)$ denote the inverse of $\psi(z,.)$. Then

$$\begin{aligned} F_{X|Z}(x|Z) &= P(X \leq x|Z) = P(\psi(Z,\nu) \leq x|Z) \\ &= P(\nu \leq \psi_2^{-1}(Z,x)|Z) \\ &= \psi_2^{-1}(Z,x). \end{aligned}$$

  Thus, $\nu = \psi(Z,X) = F_{X|Z}^{-1}(X|Z)$.

- Second, because as previously, $X \perp\!\!\!\perp \varepsilon|\nu$,

  $$P(Y \leq y|X=x,\nu) = P(\varphi(x,\varepsilon) \leq y|X=x,\nu) = P(\varphi(x,\varepsilon) \leq y|\nu).$$

- Thus, integrating ν over $[0,1]$ yields:

$$\int_0^1 P(Y \leq y|X=x,\nu=u)du = P(\varphi(x,\varepsilon) \leq y),$$

  and we can recover $q_\tau(Y,x)$ by finding y^* such that

$$\textstyle\int_0^1 P(Y \leq y^*|X=x,\nu=u)du = \tau.$$

A second generalization to nonadditive model

- This generalization is convenient because it imposes no restriction on φ. Thus, we can handle limited Y.
- A parametric particular case is Rivers and Vuong (1989) method for probit models.
- On the other hand, $\psi(Z,.)$ is strictly increasing, which imposes X to be continuous...
- To sum up, we have solutions for either Y limited and X continuous or Y continuous and X limited, but not when Y and X are limited.
- To date, there is no "universal" solution for this problem. Particular solutions do exist, however: see Lewbel (2000), Vytlacil and Yildiz (2007)...