# Endogeneity and Instruments in Nonparametric Models: Comments to papers by Darolles, Florens and Renault, and Blundell and Powell

Manuel Arellano

Seattle, August 11, 2000

*Topic:*

Extending nonlinear models with endogenous regressors to semi & nonparametric contexts.

The papers are of high quality and extend the literature in significant ways.

*Structural form and reduced form approaches*

The structural approach aims at estimating behavioral or technological relationships. Endogeneity arises when a model suggests dependence between explanatory variables and unobservables. When this occurs, *instrumental variables* is the leading method of identification in econometrics.

(Despite the modest record of instances in which there is an instrument that is convincing a priori and empirically useful.)

Structural parameters are estimated for policy evaluation, or simply as a way of providing economically interesting description.

The conventional approach to policy evaluation is to simulate the (counter-factual) policy of interest using a previously estimated structural model.

But recognition that certain policy impacts may be identified under weaker assumptions than structural parameters, has created renewed interest in *reduced form approaches* to estimation of policy impacts (Angrist, 2000, and Ichimura & Taber, 2000, are recent examples).

Sometimes the identification of policy impacts in reduced form approaches is also based on instrumental variables. In this context, instruments have been used in conjunction with (or instead of) controlled experiments, selection, or matching techniques.

The papers in this session are, nevertheless, concerned with the identification and estimation of structural (functional or finite dimensional) parameters. So for my discussion I will have in mind structural models rather than reduced forms for policy impacts.

## A) Darolles-Florens-Renault (DFR)

### Summary

DFR consider nonparametric estimation of $g(x)$ in a structural equation of the form

$$\begin{aligned} y &= g(x) + u \\ E(u \mid z) &= 0 \end{aligned}$$

when $x$ and $z$ are continuous r.v.'s. An earlier treatment of this problem was provided by Newey and Powell (1989).

The function $g(x)$ satisfies:

$$E(y \mid z) = \int g(x)\, dF(x \mid z). \tag{1}$$

If $x$ and $z$ are *discrete* with finite support, $g(x)$ is of finite dimension and the inverse of the linear functional operator (1) is continuous in $E(y \mid z)$. In such case, provided a rank condition is satisfied (which requires the support of $z$ to be at least as large as the support of $x$), $g(x)$ is identified and consistent estimation is straightforward.

However, if (1) is an infinite dimensional operator, its inverse is not continuous in general (called an *ill-posed inverse problem*). Lack of continuity implies that the availability of consistent estimates of $E(y \mid z)$ and $F(x \mid z)$ does not guarantee consistent estimation of $g(x)$.

*Identification*

The function $g$ is identified if the solution to the integral equation (1) is unique.

In turn, this is equivalent to the statistical *completeness* of $F(x \mid z)$ in $z$.

When $x$ and $z$ have no elements in common (ie. all $x$ are endogenous), DFR prove that $g(x)$ is identified if and only if $\lambda_j > 0$ for all $j$, where $\{\lambda_j\}$ is the sequence of eigenvalues of the double conditional expectation operator $E\{E[g(x) \mid z] \mid x\}$.

*Estimation*

Replace the original problem (1) with the transformed double expectation operator

$$E\left[E\left(y \mid z\right) \mid x\right] = E\left\{E\left[g\left(x\right) \mid z\right] \mid x\right\}. \qquad (2)$$

This is still ill-posed but it has the same argument as $g(x)$. For this formulation, DFR note that the problem can be overcome replacing (2) with the approximate problem (Tikhonov regularization):

$$E\left[E\left(y \mid z\right) \mid x\right] = E\left\{E\left[g^{*}\left(x\right) \mid z\right] \mid x\right\} + \alpha g^{*}\left(x\right). \quad (3)$$

As $\alpha \to 0$ the approximate problem approaches the original one, but for fixed $\alpha$ (3) is well-posed.

## Comments

This is an exciting paper that opens up a new class of applications by providing a workable theory for nonparametric structural equations.

The paper raises many questions that still have no answer. Some of these issues are listed by the authors in the conclusions.

A specially relevant issue for applied work is whether DFR's asymptotic normality result will lead to a practical asymptotic inference framework for functionals of the estimator of $g(x)$.

The following comments are just meant to add to the list of open issues suggested by the paper.

1) *Nonlinear implicit structural equations*. The literature on parametric nonlinear structural models (beginning with the work of Amemiya) considered the more general formulation

$$f(y,x) \equiv f(w) = u \quad (4)$$

$$E(u \mid z) = 0. \quad (5)$$

In a structural context, implicit equations seem more attractive. The model $y = g(x) + u$ is a natural formulation in a regression context, but for a structural equation it gives a very uneven treatment of observables and unobservables.

• $f(y, x) = u$ may reflect a direct interest in $f$. An example is an Euler equation of the form

$$U'(c_{t+1})r_{t+1} - U'(c_t) = u_{t+1}. \tag{6}$$

where $U'(.)$ denotes the marginal utility of consumption. Parameters of interest could be the coefficients of relative risk aversion for different values of $c$ estimated in a non-parametric way (Gallant and Tauchen, 1989).

In this example there is no left-hand side variable but the model imposes a particular structure to the implicit function: additivity and monotonicity in $U'(.)$. Structural models often impose not only IV conditions but also restrictions on the shape of functions. The analysis of DFR suggests the interest to explore alternative *economically based types of regularization* for specific models.

• Another situation of interest is when the starting point is an *invertible response function*

$$y = H(x, u), \tag{7}$$

which can be represented as $f(y, x) = u$, together with the assumption

$$E\left[c(u) \mid z\right] = 0$$

for some function $c(.)$. Identification of $f^*(y, x) = c\left[f(y, x)\right]$ up to scale affords calculation of the following derivative effects with respect to the structural function $H$:

$$-\left(\frac{\partial f(y, x)}{\partial y}\right)^{-1} \frac{\partial f(y, x)}{\partial x}. \tag{8}$$

• If $(w, z)$ is *discrete* with finite support the analysis of the implicit IV model is straightforward. The model specifies

$$\sum_{j=1}^{J} f(\xi_j) \Pr(w = \xi_j \mid z = \zeta_\ell) \; (\ell = 1, ..., L) \tag{9}$$

where $w \in \{\xi_1, ..., \xi_J\}$ and $z \in \{\zeta_1, ..., \zeta_L\}$.

In matrix form:

$$P\theta = 0 \tag{10}$$

where $P$ is an $L \times J$ matrix of conditional probabilities, and $\theta$ is the $J \times 1$ vector with elements $\theta_j = f(\xi_j)$. The order condition for identification of $\theta$ up to scale is $L \geq J - 1$ and the rank condition is $rank(P) = J - 1$.

This is a standard GMM problem: Letting $r_j = \mathbf{1}(w = \xi_j)$ and $m_\ell = \mathbf{1}(z = \zeta_\ell)$, we can write

$$E\left[m_\ell\left(\theta_1 r_1 + ... + \theta_J r_J\right)\right] = 0 \; (\ell = 1, ..., L) \tag{11}$$

which is in the form of a system of $L$ simultaneous equations with instruments $m_\ell$ in equation $\ell$.

• Another special case is a model including a subset of $z$ in $f$ so that $f(w, z_1) = u$, in which the endogenous r.v.'s $w$ are discrete but $z = (z_1, z_2)$ are continuous. This is equivalent to the semi-parametric conditional moment restriction:

$$E\left(\sum\nolimits_{j=1}^{J} \theta_j(z_1) r_j \mid z\right) = 0 \tag{12}$$

where $w \in \{\xi_1, ..., \xi_J\}$, $\theta_j(z_1) = f(\xi_j, z_1)$, and $r_j = \mathbf{1}(w = \xi_j)$.

2) *Testing for overidentification & underidentification.* The IV model can be regarded as a restriction on the cdf of $w \mid z$

$$\int f(w)dF(w \mid z) = 0.$$

Sometimes the focus is not in estimating $f(w)$ (or $y - g(x)$) but in testing the restrictions on $F(w \mid z)$. From this point of view $f(w)$ becomes a nuisance parameter function.

• In the discrete case an invariant $\chi^2$ test statistic of the overidentifying restrictions (with $L - J + 1$ d.f.) is readily available -but of no use in the continuous case-. This is given by

$$\min_{\theta} n\widehat{p}'(I \otimes \theta) \left[ (I \otimes \theta')\widehat{V}(I \otimes \theta) \right]^{-1} (I \otimes \theta')\widehat{p}$$

where $\widehat{p} = vec(\widehat{P})$ denotes a vector of sample frequencies, and $\widehat{V}$ is the estimated sampling variance of $\widehat{p}$.

• Testing for *underidentification* in the discrete case is also straightforward: One would test the null of underidentification ($rank(P) < J - 1$) against the alternative of identification ($rank(P) = J - 1$). A statistic of this kind provides a natural diagnostic of the extent to which structural parameter estimates are well identified.

Related to this, I wondered if DFR's identification conditions using the spectral decomposition, and the results on the asymptotic properties of estimated eigenvalues (Darolles-Florens-Gourieroux, 1998) could lead to a nonparametric test for underidentification in the continuous case.

**B) Blundell-Powell**

**Control functions with additive errors**

*Newey-Powell-Vella (1999) (NPV)* considered a nonparametric structural equation with an explicit reduced form:

$$y = g(x) + u$$
$$x = \pi(z) + v$$

and the assumptions

$$E(u \mid z, v) = E(u \mid v) \tag{13}$$
$$E(v \mid z) = 0. \tag{14}$$

These assumptions were chosen for convenience. In effect, they imply

$$E(y \mid x, v) = g(x) + E(u \mid x, v) \tag{15}$$
$$= g(x) + E(u \mid z, v) = g(x) + E(u \mid v) = g(x) + h(v).$$

In this way the problem of nonparametric estimation of $g(x)$ is assimilated to the problem of estimating the regression function $E(y \mid x, v)$ subject to an additive structure.

*Discussion of the assumptions.* Note that (13)-(14) do not imply $E(u \mid z) = 0$:

$E(u \mid z) = E[E(u \mid z, v) \mid z] = E[E(u \mid v) \mid z] = E[h(v) \mid z$

A sufficient condition for $E[h(v) \mid z] = 0$ is that $v$ is independent of $z$. Mean independence does not guarantee that $E[h(v) \mid z] = 0$ unless $h(v)$ is linear in $v$.

Alternatively, if we begin with the assumptions $E(u \mid z) = 0$ and $E(v \mid z) = 0$, in general (13) or (15) are not satisfied.

The CF assumption can be very useful in applied work, but one should insist that the IV condition $E\left(u \mid z\right)=0$ also holds.
• Having a structural equation in which instruments are correlated with errors because of a simplifying assumption may jeopardize the interpretability of the structural parameters.
• From the point of view of econometric practice, it is better to regard the CF assumption as *a specialization of the IV assumption* than to pretend that one is no more, no less general than the other.
ie. CF as an approach in which estimation of $g(x)$ is *helped* by an explicit semiparametric modelling of the reduced form.
• This will typically require aiming for a reduced form with errors that are independent of instruments.

As an example, suppose that for scalar $x$, $v$ is heteroskedastic with $\sigma^2(z)=E\left(v^2 \mid z\right)$, but $v^{\dagger}=\sigma^{-1}(z)v$ is independent of $z$. In such case, the assumption

$$E\left(u \mid z, v^{\dagger}\right)=E\left(u \mid v^{\dagger}\right)$$

is compatible with $E\left(u \mid z\right)=0$, but (13) will imply in general correlation between $u$ and $z$.

The control $v$ can be generalized further, eg. to include a Box-Cox-like transformation of $x$. The idea is that the approach works well when there is a reduced form equation for some transformation of $x$ with errors that are independent of $z$.

## Control functions in discrete choice models

*Blundell-Powell (2000) (BP)* is a very nice paper which shows how the CF approach can be particularly helpful in models with non-additive errors. BP consider the model

$$y = \mathbf{1}(x\beta + u > 0) \tag{16}$$

$$x = \pi(z) + v \tag{17}$$

$$E(v \mid z) = 0 \tag{18}$$

together with the assumption

$$u \mid x, v \sim u \mid v. \tag{19}$$

In this way

$$\Pr(y = 1 \mid x, v) = \Pr(-u \leq x\beta \mid x, v) = \Pr(-u \leq x\beta \mid v)$$

so that

$$E(y \mid x, v) = F(x\beta, v)$$

where $F(., v)$ is the conditional cdf of $-u$ given $v$.

As in the case of NPV the problem of estimating a structural equation is assimilated to the problem of estimating the regression function $E(y \mid x, v)$ subject to restrictions.

• In the NPV case it was sufficient to assume that $u$ was mean independent of $x$ given $v$, and $E(y \mid x, v)$ had an additive structure.

• In the discrete choice case full independence of $x$ given $v$ is required, and $E(y \mid x, v)$ has a multiple index structure.

The difference between the two models is due to the fact that (16) is not additive or invertible in $u$.

An interesting feature of the BP method is that *the marginal cdf of* $u$ evaluated at $x\beta$ can be obtained by averaging $F(x\beta, v)$ over $v$ whose cdf is identified:

$$\Pr(-u \leq x\beta) \equiv G(x\beta) = \int F(x\beta, v)\, dF_v.$$

This is useful because the function $G(x\beta)$ is arguably a parameter of interest for policy evaluation in this context.

**Comments**

1) If $(u, v)$ are independent of $z$ then

$$u \mid z, v \sim u \mid v. \tag{20}$$

Moreover, in view of (17), $u \mid x, v \sim u \mid z, v \sim u \mid v$.

However, the CF assumptions by themselves do not imply independence or even lack of correlation between $u$ and $z$: If (20) holds, in general $u$ will not be independent of $z$ unless $v$ is independent of $z$:

$$F(u \mid z) = \int F(u \mid z, v) dF_v(v \mid z) = \int F(u \mid v) dF_v(v \mid z)$$
$$\neq \int F(u \mid v) dF_v(v) = F(u).$$

So, the previous remarks also apply in this context: If $x\beta + u$ represents a latent structural equation, one would expect to select instruments on a priori grounds that suggest some form of independence with $u$.

The conclusion is that the CF approach is best regarded not as a competing identification strategy to IV assumptions but as a complementary modelling strategy for the reduced form of the model.

This strategy is specially useful in discrete choice and related models, for which IV assumptions by themselves do not appear to be sufficient to identify parameters of interest.

2) *Discrete choice with the IV assumption* $u \mid z \sim u$. According to this model:

$$\begin{aligned}\Pr(y = 1 \mid z) &= \int \mathbf{1}(x\beta + u > 0) dF_{ux}(u, x \mid z) \\ &= \int E\left[\mathbf{1}(x\beta + u > 0) \mid z, u\right] dF_u(u)\end{aligned}$$

It would be useful to have a *non-identification theorem* for this model in the absence of further assumptions.

3) *Relation to Lewbel (1996).* Lewbel considered discrete choice models in which there is a special exogenous continuous explanatory variable $z_1$ such that

$$y = \mathbf{1}(\gamma z_1 + w\delta + u > 0)$$
$$E(zu) = 0$$
$$u, w \mid z_1, z_2 \sim u, w \mid z_2.$$

This implies that $z_1$ is excluded from the reduced form of $w$. Under these assumptions (and normalizing $\gamma = 1$) Lewbel showed that

$$E(z_2 x')\delta = E\left(z_2 \frac{y - \mathbf{1}(z_1 > 0)}{f(z_1 \mid z_2)}\right),$$

and proposed a 2SLS procedure to estimate $\delta$.

As suggested by Lewbel, by combining this approach with the CF approach it is possible to rely on a less restrictive CF assumption when $z_1$ is in the equation. The idea is to relax the BP CF assumption by considering

$$u \mid z_1, w, v \sim u \mid w, v$$

where $z = (z_1, z_2)$, $w = \pi(z) + v$, $E(v \mid z) = 0$, and suppose for simplicity that all variables in $w$ are endogenous.

Then it turns out that Lewbel's methodology can be applied using $\pi(z)$ as instruments instead of $z_2$, and $f(z_1 \mid \pi(z))$ instead of $f(z_1 \mid z_2)$, under the assumption that $E[\pi(z)u] = 0$.

4) *Two-sample estimation.* Suppose we have two independent samples on $(y, z)$ and $(x, z)$, respectively, but the joint distribution of $y$ and $x$ is not observed (as in Arellano-Meghir, 1992, or Angrist-Krueger, 1992).

It is interesting to compare IV and CF in this context to highlight the different data requirements in the two approaches.

• In the IV approach only the marginal distributions of $y$ and $x$ given $z$ are needed for identification. So $y \mid z$ can be obtained from one sample and $x \mid z$ from the other.

• In the CF approach, however, $(y, x)$ need to be observed in the same sample to be able to do a nonparametric regression of $y$ on $x$ and $v$.

5) *Empirical application.* It would be nice to test for heteroskedasticity in the log income equation, and if present to obtain a kernel estimate of $Var(y_2 \mid z) \equiv \sigma^2(z)$. Then consider $v^\dagger = v/\sigma(z)$ as an alternative CF, provided $v^\dagger \perp z$ is not rejected.