Nonparametric Instrumental Variables Estimation

Whitney K. Newey

January 2012

INTRODUCTION

In many economic models, objects of interest satisfy conditional expectation restrictions.

May come from first-order conditions for choices by individuals or firms.

Need to estimate these to test theory or predict the effect of policy changes like taxes.

Economics does not restrict functional form, motivating nonparametric estimators.

Challenging to recover high order nonlinearities in structure with conditional moment restrictions because:

– Data provides information about reduced forms.

– Reduced forms are conditional expectations.

– Conditional expectations "smooth out" nonlinearities.

Global curvature is recovered in nonlinear empirical examples.

Important to go beyond linear model.

BRIEF SURVEY

Newey and Powell (1989, 2003): Model, identification, series estimator, consistency.

Ai and Chen (2003): Root-n consistency for semiparametric estimator.

Blundell, Chen, and Kristensen (2007), Darrolles, Florens, and Renault (2011), Hall and Horowitz (2005): Nonparametric convergence rates.

Horowitz (2006): Testing linearity.

Horowitz (2007): Initial results on asymptotic normality.

Chernozhukov and Hansen (2004), Chernozhukov, Imbens, and Newey (2007), Horowitz and Lee (2007): Nonseparable and quantile model.

Santos (2011), Severini and Tripathi (2012): Root-n consistent estimation.

Santos (2012): Inference without identification.

Chen, Chernozhukov, Lee, and Newey (2012): Local identification of nonparametric models.

THE MODEL

$$y = g_0(x) + \varepsilon, E[\varepsilon|z] = 0$$

y is left-hand side endogenous variable, x is right hand side endogenous variable, z is instrument, g_0 is an unknown structural function, ε is a disturbance.

Can allow for exogenous covariates z_1 by including in x and z.

Economic model may imply $E[y - g_0(x)|z] = 0$, as in consumption CAPM, where y is rate of return, g_0 is intertemporal marginal rate of substitution, z is information available at time decision is made.

Nonseparable model, neither more or less general, is $y = \tilde{g}_0(x, \varepsilon)$, z and ε independent.

Control function, neither more or less general, has $E[\varepsilon|x, v] = E[\varepsilon|v]$ for observable or estimable v.

For an economic example consider a supply demand model where

$$q = g_0(p, z_1) + \varepsilon, p = h_0(q, z_2) + \eta, E[\varepsilon|z_1, z_2] = E[\eta|z_1, z_2] = 0.$$

Here $g_0(p, z_1) + \varepsilon$ is the demand function and the inverse supply $h_0(q, z_2) + \eta$.

Let τ be a percentage tax that is paid by the purchaser.

The equilibrium quantity that would result is the solution $q(z_1, z_2, \varepsilon, \eta)$ to

$$\tilde{q} = g_0((1 + \tau)[h_0(\tilde{q}, z_2) + \eta], z_1) + \varepsilon.$$

The effect of the tax on average quantity would be

$$E[\tilde{q} - q].$$

Knowledge of the structural function g_0 is essential here.

IDENTIFICATION

Taking conditional expectations of both sides of $y = g_0(x) + \varepsilon$ with respect to z gives

$$E[y|z] = E[g_0(x)|z] = \int g_0(x) f(x|z) dx,$$

where $f(x|z)$ is the conditional pdf of x given z.

$E[y|z]$ and $f(x|z)$ are identified nonparametric reduced forms.

Identification question is whether there is a unique solution to $E[y|z] = \int g(x) f(x|z) dx$.

Note that $g_0(x)$ and $g(x)$ both solve this equation if and only if

$$E[g_0(x) - g(x)|z] = E[g_0(x)|z] - E[g(x)|z] = E[y|z] - E[y|z] = 0.$$

Thus, $g_0(x)$ is identified if and only if $\delta(x) = 0$ is the only function satisfying $E[\delta(x)|z] = 0$.

Completeness of conditional expectation of functions of x conditional on z.

A nonparametric generalization of the rank condition from linear models.

Equivalent to a rank condition if x and z only take on finite number of values.

$x \in \{x_1, ..., x_J\}$ and $z \in \{z_1, ..., z_K\}$.

Let $\pi_{jk} = \Pr(X = x_j | Z = z_k)$.

Completeness holds if and only if

$$rank(\pi) = J.$$

Necessary order condition is $K \geq J$.

Newey and Powell (1989,2003).

Identification is that $\delta(x) = 0$ is the only function satisfying $E[\delta(x)|z] = 0$.

Unlike parametric case, where the rank condition is testable, completeness is not testable, Canay, Santos, and Shaikh (2011).

Intuitively, reduced form "matrix" is infinite dimensional with eigenvalues that go to zero.

Cannot distinguish empirically a sequence of eigenvalues that goes to zero from a sequence that has one zero.

But completeness is "generic," i.e. holds for "most" $f(x|z)$, if it holds for one, Andrews (2011), Chen, Chernozhukov, Lee, and Newey (2012).

Like finite outcome case, where"most" π have $rank(\pi) = J$ if $K \geq J$.

Similarly like linear model where "most" reduced form matrices have full rank if the order condition is satisfied and all the instrumental variables affect the endogenous variable.

ESTIMATION

A challenge is that the solution g to

$$E[y|z] = \int g(x)f(x|z)dx.$$

is discontinuous in $E[y|z]$. Thus, replacing $E[y|z]$ and $f(x|z)$ by estimators and solving for $\hat{g}$ need not give consistent $\hat{g}$.

A solution to this problem is to "regularize," to control $\hat{g}$.

Penalizing the size of $\hat{g}$ and its derivatives regularizes.

Series approximations also regularize.

Focus on series approximations because they are simple.

Like Horowitz (2011) without density weighting; see also Newey and Powell (1989).

Series estimate.

Let $p_{1J}(x), ..., p_{JJ}(x)$ be approximating functions

$$g(x) \approx \sum_{j=1}^{J} \gamma_j p_{jJ}(x).$$

Plug in the approximation into $E[y|z] = E[g_0(x)|z]$ to obtain

$$E[y|z] \approx \sum_{j=1}^{J} \gamma_j E[p_{jJ}(x)|z].$$

Replace $E[p_{jJ}(x)|z]$ by nonparametric estimates $\widehat{E[p_{jJ}|z]}$ and do least squares to get

$$\hat{\gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_J)' = \arg\min_{\gamma} \hat{S}(\gamma), \hat{S}(\gamma) = \sum_i \{y_i - \sum_{j=1}^{J} \gamma_j \widehat{E[p_{jJ}|z_i]}\}^2,$$

for data (y_i, x_i, z_i), $(i = 1, ..., n)$. A nonparametric 2SLS estimator of $g_0(x)$ is

$$\hat{g}(x) = \sum_{j=1}^{J} \hat{\gamma}_j p_j(x).$$

Can also use series approximation to construct $\widehat{E[p_{jJ}|z]}$.

For $q^K(z) = (q_{1K}(z), ..., q_{KK}(z))'$, $q_i^K = q^K(z_i)$, can take

$$\widehat{E[p_{jJ}|z]} = q^K(z)'(\sum_{i=1}^{n} q_i^K q_i^{K\prime})^{-1} \sum_{i=1}^{n} q_i^K p_{jJ}(x_i).$$

Then estimator is just the same as 2SLS with right-hand side variables $p_{1J}(x), ..., p_{JJ}(x)$ and instrumental variables q_i^K.

Should not think of this as just a parametric estimator, or flexible functional form (e.g. translog), because J and K can be chosen to adjust to the conditions in the data.

What makes series estimators nonparametric is that the number of terms can vary across applications, with more terms included to account for more nonlinearlity.

This view of nonparametric series estimators highlights the need for J and K to adjust to conditions in the data, and thus the need for good data based methods of selecting J and K.

Discuss briefly below.

Can allow for larger J, and hence more local variation, at the expense of introducing additional bias.

This can be done by adding a penalty, which is another form of regularization to make the estimator consistent in spite of the ill-posed inverse problem.

Can construct $\tilde{\gamma}$ as

$$\tilde{\gamma} = \arg\min_{\gamma}\{\hat{S}(\gamma) + \gamma'\hat{\Lambda}\gamma\},$$

where $\hat{\Lambda}$ is a positive definite matrix that is well conditioned relative to the leading term.

Can allow J quite large and still get consistency of the estimator.

Choice of $\hat{\Lambda}$: Identity is not a good idea; does not account for location and scale of γ.

In sample Tikhonov regularization sets $\hat{\Lambda}$ equal to the second moment matrix of $(p_{1J}(x_i), ..., p_{jJ}(x_i))$. This penalizes using the size of $\tilde{g}(x) = \sum_{j=1}^{J} \tilde{\gamma}_j p_j(x)$.

Instead might want to penalize size of derivatives. Leads to more complicated $\hat{\Lambda}$. See Blundell, Chen, and Kristensen (2007).

NONLINEARITY AND INSTRUMENT STRENGTH

The ability to estimate nonlinearities in nonparametric IV is linked to the strength of the instrument.

We show this in a Gaussian example where x and z are joint standard normal with correlation coefficient ρ.

Here we can easily construct a nonparametric estimator and see how the growth of J must be limited to get consistency.

More interestingly, it leads to a clear link between the strength of the instruments and the variability of coefficients of nonlinear terms.

Also, given the intellectual history of IV, the Gaussian example seems potentially interesting.

Let $(p_1(x), p_2(x), ...)$ denote the Hermite polynomials, that are orthonormal, satisfying

$$E[p_j(x)^2] = 1, E[p_j(x)p_k(x)] = 0, j \neq k.$$

It is known that

$$E[p_j(x)|z] = \rho^j p_j(z).$$

This formula exemplifies the ill-posed inverse problem, in that $E[p_j(x)^2] = 1$ for all j, so $p_j(x)$ is not converging zero, but $E[p_j(x)|z]$ does go to zero in the sense that $E[\{E[p_j(x)|z]\}^2] = \rho^{2j}$ goes to zero.

There is also a clear relationship between the ability to estimate nonlinear terms and the strength of the instrument.

The model is $y = g_0(x) + \varepsilon$, $E[\varepsilon|z] = 0$.

Simple series estimator (using known ρ) is $\hat{g}(x) = \sum_{j=1}^{J} \hat{\gamma}_j p_j(x)$, where

$$\hat{\gamma}_j = \rho^{-j} \sum_{i=1}^{n} y_i p_j(z_i)/n.$$

Note that

$$Var(\hat{\gamma}_j) \leq \left(\rho^2\right)^{-j} E[y_i^2 p_j^2(z)]/n \leq \left(\rho^2\right)^{-j} \max_z E[y_i^2|z]/n.$$

Note also that ρ^2 is the r-squared from the reduced form regression of x on z.

Thus the smaller is ρ^2 the larger will be the variance of $\hat{\gamma}_j$ for large j relative to small j.

For example, if the reduced form r^2 is .1, then the variance of the quadratic term may be 10 times as large as the linear term.

Have similar calculation when x and z are uniform with $f(x|z)$ analytic, Kress (1999, Theorem 5.20).

Angrist and Krueger (1991): x is years of schooling and z is quarter of birth.

The reduced form r-squared is very small and the variance of the linear IV estimator is quite large.

Previous calculation would predict that variance of a quadratic term is extremely large. We understand this is what happens in that application when a quadrtic term is included.

Result is roughly consistent with previous calculation, even though x and z not Gaussian in this application.

Blundell, Chen, and Kristensen (2007): x is total expenditure and z is income.

The reduced form r-squared is quite large there, about .25.

Important nonlinearities can be estimated in that application.

INFERENCE

Important to be able to tell in an application whether nonlinearities are important.

Previous suggests that if doing linear IV and have small reduced form r-squared and large variance for linear IV then no need to check.

Would be convenient if you could just treat the estimator as if it were parametric and use 2SLS standard errors.

In other nonparametric estimation settings can do this, for both kernel and series estimators, Newey (1994, 1997).

Intuition is this accounts for variance, and asymptotic normality results assume bias goes away faster than standard deviation.

Leads to simple suggestion for choosing J, which is to increase until estimator does not change much relative to the (parametric) standard error.

Beyond the scope of this paper to do the econometric theory but does work in a simple Monte Carlo example.

To focus on inference we consider a simple example where x and z are Gaussian, $g_0(x)$ is zero, and ρ is either .25 or .1.

We take for approximating functions simple power series, where $p_{jJ}(x) = x^{j-1}$ and $q_{kK}(z) = z^{k-1}$.

We consider the coverage probability for an asymptotic 90 percent confidence interval for the value $g_0(0)$ of the structural function at zero.

We use as standard error the Hansen (1982), White (1982) heteroskedasticity consistent standard errors for IV.

Here are the results:

$\rho^2 = .25$

n	J	K	$Cover$
100	1	2	.896
500	2	6	.899
2500	3	12	.905

$\rho^2 = .1$

n	J	K	$Cover$
100	1	2	.914
1200	2	6	.908
8000	3	15	.919

CONCLUSIONS

–Challenging to estimate higher order nonlinearities in conditional moment models with low reduced form r-squareds.

–Can find important nonlinearities in applications.

–Shape restrictions might help.

–Need for method for selecting order of approximation to give good distribution results.

–Simple example suggests that just using parametric standard error formulas may work fine, as it does in other series regression applications.