American Economic Review: Papers & Proceedings 2013, 103(3): 550–556
http://dx.doi.org/10.1257/aer.103.3.550

LESSONS FROM NONPARAMETRIC METHODS IN HISTORICAL PERSPECTIVE[‡]

Nonparametric Instrumental Variables Estimation[†]

By Whitney K. Newey*

In many economic models, objects of interest are functions that satisfy conditional moment restrictions. These models may come from first-order conditions for choices by individuals or firms. They are also natural generalizations of linear instrumental variable (IV) models, where there is an unknown function rather than unknown parameters, and disturbances have conditional mean zero rather than just being uncorrelated with instruments. In many economic models these unknown functions have structural interpretations, making it essential to estimate them in order to test economic theory or predict policy effects. Economics does not restrict functional forms, motivating nonparametric models where objects of interest are unknown functions.

In this paper we describe a simple approach to estimating these models. It is based on series approximation to the unknown structural function and reduced forms, similar to Newey and Powell (1989, 2003); Blundell, Chen, and Kristensen (2007); and Horowitz (2011). This approach leads to nonparametric IV (NPIV) estimation. We also describe a way to modify the objective function for this estimator that allows for more approximating terms at the expense of introducing additional bias. We also suggest a method of inference based on these estimators.

It can be challenging to recover nonlinearities from conditional moment restrictions. The difficulty comes because the data identifies only reduced form conditional expectations, and conditional expectations "smooth out" nonlinearities, leading to what is known as the "ill-posed inverse problem." In this paper we quantify this problem in a Gaussian example. We find a link between the reduced form R^2, that quantifies the strength of the instrument, and the variance of the coefficients of nonlinear terms. When the reduced form R^2 is very low, and a linear IV slope estimate is quite variable, say as in Angrist and Krueger (1991), the variance of coefficients of any nonlinear terms will be very high. It will be difficult to find any nonlinearity in such settings. On the other hand, when the reduced form R^2 is larger, say as in the Engel curve estimation in Blundell, Chen, and Kristensen (2007), one can estimate nonlinear terms with some precision.

We also discuss a simple approach to inference, namely treating the series estimator as if it were parametric and using corresponding standard errors and the usual Gaussian approximation. In other nonparametric settings, it is known that this approach is asymptotically valid for either kernel or series estimators (Newey 1994, 1997). The intuition here is that the parametric calculation accounts correctly for the variance of the estimator and that the bias shrinks faster than the standard deviation in the asymptotic approximation. Recent work has shown that this approach works for some nonparametric IV series estimators, Chernozhukov and Chetverikov (2013). We give a small Monte Carlo example showing that it is promising in a Gaussian example.

The nonparametric IV problem and its identification, series estimation, and consistency

[‡]*Discussants:* Richard Blundell, University College London; Dennis Kristensen, University College London; Rosa Matzkin, University of California-Los Angeles.

*Department of Economics, MIT, 50 Memorial Drive, E52-373c, Cambridge, MA 02139, (e-mail: wnewey@mit.edu). The National Science Foundation (Grant SES 1132399) provided support for the work for this paper. Helpful comments were provided by X. Chen, V. Chernozhukov, R. Matzkin, and J. Powell.

[†] To view additional materials, and author disclosure statement(s),visit the article page at http://dx.doi.org/10.1257/aer.103.3.550.

was treated in Newey and Powell (1989, 2003). Blundell, Chen, and Kristensen (2007); Darrolles et al. (2011); and Hall and Horowitz (2005) derived nonparametric convergence rates. Horowitz (2006) gave some results on inference, as did Ai and Chen (2003) for root-n consistent estimators of parameters of semiparametric models. Santos (2012) developed identification robust inference. Chen et al. (2012) have analyzed local identification of nonlinear models. Many other papers have made interesting contributions to this literature. This paper pursues the modest goal of briefly discussing the model and its identification and offering some thoughts on series estimation, instrument strength, nonlinearities, and inference.

I. The Model

The model we consider is

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

where y is the left-hand-side endogenous variable, x is a vector of right-hand-side possibly endogenous variables, z is a vector of instruments, g_0 is an unknown structural function, and ε is a disturbance. This model includes as a special case nonparametric regression, where $x = z$ and $g_0(x) = E[y|x]$. It generalizes nonparametric regression to allow some of the regressors x to be correlated with the disturbance ε. Exogenous covariates can be allowed for by including a subvector z_1 of z in x.

An economic model may imply $E[\varepsilon|z] = 0$. In various consumption capital asset pricing models we have such restrictions. For example, y could be the rate of return on an asset, x consumption in successive time periods, $g_0(x)$ the intertemporal marginal rate of substitution, and z information available when consumption is decided, such as lagged returns and consumption, e.g., Hansen and Singleton (1982). The conditional moment restriction $E[\varepsilon|z] = 0$ can also be viewed as a generalization to nonlinear models of the fundamental condition of orthogonality between instruments and disturbances. For example, Berry and Haile (2009) use it for nonparametric demand models for imperfect competition.

Other nonparametric models that allow correlation between regressors and disturbances have also been developed. These include nonseparable models where $y = m_0(x, \varepsilon)$, and z and ε are independent. See Chernozhukov and Hansen (2005); Chernozhukov, Imbens, and Newey (2007); and Matzkin (2008). Also nonparametric models with control functions have been considered, in Newey, Powell, and Vella (1999); Blundell and Powell (2003); Chesher (2003); and Imbens and Newey (2009). In these models the control function is an observable or estimable variable v (often a reduced form residual) such that ε and x are independent, or mean independent, conditional on v. Nonseparable and control function models are neither more or less general than the conditional moment restrictions model. They may allow for ε to enter nonlinearly but impose different restrictions on ε, x, and z.

It should be well understood that structural models are needed to test economic hypotheses and evaluate policies. For example consider a supply and demand model where q is quantity and p is price, satisfying

$$q = g_0(p, z_1) + \varepsilon,$$

$$p = h_0(q, z_2) + \eta,$$

$$E[\varepsilon|z_1, z_2] = E[\eta|z_1, z_2] = 0.$$

Here $g_0(p, z_1) + \varepsilon$ is the demand function and the inverse supply is $h_0(q, z_2) + \eta$. 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, where the average is taken across observed markets would be $E[\tilde{q} - q]$. Knowledge of the structural functions g_0 and h_0 are essential to identifying this policy effect.

II. 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. We can think of $E[y|z]$ and $f(x|z)$ as identified nonparametric reduced forms. The 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

$$\begin{aligned} &E[g_0(x) - g(x)|z] \\ &= E[g_0(x)|z] - E[g(x)|z] \\ &= E[y|z] - E[y|z] \\ &= 0. \end{aligned}$$

Thus, $g_0(x)$ is identified if and only if $\delta(x) = 0$ is the only function satisfying $E[\delta(x)|z] = 0$. This is completeness of the conditional expectation of functions of x conditional on z.

In a linear model completeness is equivalent to the usual rank conditions. When the data is discrete with finite support, completeness is also equivalent to a rank condition. Suppose x and z only take on a finite number of values, say $x \in \{x_1, \ldots, x_J\}$ and $z \in \{z_1, \ldots, z_K\}$. Let $\pi_{jk} = \Pr(X = x_j | Z = z_k)$. Then completeness holds if and only if $rank(\pi) = J$. Here a necessary order condition for identification is that the number of support points K for z is no smaller than the number of support points J for x. These identification results were given in Newey and Powell (1989, 2003).

In fully nonparametric models (that are infinite dimensional), completeness is not testable, as pointed out by Canay, Santos, and Shaikh (2012). In these models the reduced form is like an infinite dimensional matrix with eigenvalues that have a limit point at zero. Nonidentification occurs when at least one of the eigenvalues equals zero. The problem with testing this hypothesis is that one cannot distinguish empirically a model with a zero eigenvalue from one where the eigenvalues have a limit point of zero. However, completeness is generic, in the sense that it holds for "most" $f(x|z)$, if it holds for one, Andrews (2011), Chen et al. (2012). This is like the discrete, finite support case where most π matrices have full column rank if the order condition is satisfied.

III. Series Estimation

The unknown function g_0 solves $E[y|z] = \int g(x) f(x|z)dx$. This type of equation is known in the mathematics literature as an integral equation of the first kind; e.g., see Kress (1999). It is well known that this equation is ill-posed, meaning that the solution is not continuous in the functions $E[y|z]$ and $f(x|z)$. This creates potential problems in estimation since the discontinuity means that a consistent estimator of g_0 need not result from plugging in consistent estimators of $E[y|x]$ and $f(x|z)$, and approximately solving this equation.

A solution to this problem is to form the estimator in such a way that the ill-posed inverse problem does not affect consistency. This approach is known as "regularization" in the mathematics literature. There are various ways to regularize. One way is to use series estimation, that specifies the number of terms in a linear approximation e.g., see Kress (1999). There are also other ways. For simplicity we focus first on series estimation and then mention other regularization approaches.

A series estimator is based on approximating the unknown function by a linear combination of known functions. Let $p_{1J}(x), \ldots, p_{JJ}(x)$ be approximating functions, such as power series or regression splines. We assume that we can approximate any $g(x)$ by a linear combination, as in

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

where $\gamma_1, \ldots, \gamma_J$ are linear combination coefficients. To estimate using this approximation we can plug the approximation for g into $E[y|z] = \int g(x) f(x|z)\, dx$ to obtain $E[y|z] \approx \sum_{j=1}^{J} \gamma_j E[p_{jJ}(x)|z]$. We can replace $E[p_{jJ}(x)|z]$ by nonparametric estimators $\widehat{E[p_{jJ}|z]}$ and do least squares, estimating $\gamma_1, \ldots, \gamma_J$ by choosing $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_J)'$ to minimize

$$\hat{S}(\gamma) = \sum_i \left\{ y_i - \sum_{j=1}^{J} \gamma_j \widehat{E[p_{jJ}|z_i]} \right\}^2,$$

where the data are (y_i, x_i, z_i), $(i = 1, \ldots, n)$. An estimator of $g_0(x)$ is then

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

We can also use a series estimator for $\widehat{E[p_{jJ}|z]}$. For another vector of approximating functions $q^K(z) = (q_{1K}(z), \ldots, q_{KK}(z))'$ and $q_i^K = q^K(z_i)$, let

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

With this choice of first stage $\widehat{E[p_{JJ}|z]}$, the estimator $\hat{\gamma}$ is the same as two-stage least squares (2SLS) with right-hand-side variables $p_{1J}(x_i), \ldots, p_{JJ}(x_i)$ and instrumental variable vector q_i^K.

This is a NPIV estimator by virtue of first and second stage series approximations. Note that this is not just a parametric estimator with some flexible functional form (e.g., translog), because any function can be approximated arbitrarily well as J grows. One way to think about this in practice is that the number of terms can vary across applications, with more terms included to account for more nonlinearity. The point of series estimators is not just that the number of terms can grow with the sample size to approximate any function. It is also that the number of terms can be chosen larger or smaller based on the conditions in different datasets.

This view of nonparametric series estimators highlights the importance of choosing the number of terms to adjust to conditions in the data, i.e., the need for good data based methods of selecting the approximating model. One approach that seems to work well for inference in practice is to increase the number of terms until estimated standard errors (which will tend to grow as number of terms increase) are large relative to how objects of interest change as more approximating terms are added. Blomquist and Newey (2002) adopted this approach in an application. A good topic for future research is its theoretical properties. Of course this approach depends on having estimated standard errors. We discuss below how these can be constructed for NPIV.

Another way to regularize the estimator is to add a penalty term to the second stage sum of squared residuals. Let Λ be some positive definite $J \times J$ matrix and α a nonnegative constant. Consider an estimator $\tilde{\gamma}$ minimizing

$$\hat{S}(\gamma) + \alpha \cdot \gamma'\Lambda\gamma,$$

where $\hat{S}(\gamma)$ is the 2SLS objective function from above. Here $\alpha \cdot \gamma'\Lambda\gamma$ is a penalty term that will tend to lower the variance of the estimator, because Λ is not random. The bigger is α the more weight the penalty has and so the less the variance and the larger the bias, with α shrinking to zero as sample size grows to ensure consistency. Various forms of these estimators were considered by Newey and Powell (1989, 2003) and Blundell, Chen, and Kristensen (2007), to which we refer the interested reader. An advantage of such methods is that J can be allowed to be much larger than for series estimates and still get consistency. A larger value of J will allow estimation of some higher-order nonlinearities, while the penalty term helps control the variance of $\hat{g}$.

IV. Nonlinearity and Instrument Strength

The ability to uncover nonlinearities with NPIV is linked to the strength of the instruments. Here we explain this in a Gaussian example where x_i and z_i are joint standard normal with correlation coefficient ρ. This simple setting allows us to relatively easily see the link. This example also seems important, given the long interest in Gaussian models.

Let $(p_1(x), p_2(x), \ldots)$ denote the Hermite polynomials, where $p_1(x) = 1$, $p_j(x)$ is a polynomial of order $j - 1$,

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

The formula for $p_j(x)$ can be found in Abramowitz and Stegun (1965). It is known that

$$E[p_j(x_i)|z_i] = \rho^j p_j(z_i).$$

This formula exemplifies the way conditional expectations generally "smooth out" nonlinearities. Here the conditional expectation shrinks the Hermite polynomial toward zero, with more shrinkage for higher order terms where j is larger, and more shrinkage when ρ is smaller. Also, this example shows the ill-posed inverse problem. Here $g_0(x) + p_1(x)$, $g_0(x) + p_2(x), \ldots$ is a sequence of "structures" that do not converge to the truth $g_0(x)$ while the corresponding "reduced forms" $E[g_0(x) + p_j(x_i)|z_i] = E[y|z] + \rho^j p_j(z)$ do converge to the true reduced form $E[y|z]$.

We can also see the impact of the ill-posed inverse problem in the relationship between structural and reduced form coefficients. To explain let $\overline{\gamma}_j = E[g_0(x_i)p_j(x_i)]$ denote the population regression coefficient from regressing $g_0(x_i)$ on $p_1(x_i), \ldots, p_J(x_i)$, where we assume $E[g_0(x_i)^2]$ exists. Here $\overline{\gamma}_j$ has this simple form because $E[p_j(x_i)^2] = 1$ and the $p_j(x_i)$ are uncorrelated across j. It is known that $g_0(x)$ will be a linear combination of all the approximation

functions, $g_0(x_i) = \sum_{j=1}^{\infty} \overline{\gamma}_j p_j(x_i)$. Similarly, for $\overline{\pi}_j = E[E[y_i|z_i]p_j(z_i)]$ the reduced form will be a linear combination of all the $p_j(z)$ functions, $E[y|z_i] = \sum_{j=1}^{\infty} \overline{\pi}_j p_j(z_i)$. Here $\overline{\gamma}_j$ and $\overline{\pi}_j$ are the structural and reduced form coefficients respectively.

There is a simple relationship between $\overline{\pi}_j$ and $\overline{\gamma}_j$. Switching the roles of x_i and z_i, we have $E[p_j(z_i)|x_i] = \rho^j p_j(x_i)$. Multiplying through by ρ^{-j} we have $p_j(x_i) = \rho^{-j} E[p_j(z_i)|x_i]$ so that

$$\begin{aligned}\overline{\gamma}_j &= \rho^{-j} E[g_0(x_i) E[p_j(z_i)|x_i]] \\ &= \rho^{-j} E[g_0(x_i) p_j(z_i)] \\ &= \rho^{-j}\, \overline{\pi}_j.\end{aligned}$$

Thus, the reduced form coefficient $\overline{\pi}_j$ must be inflated by ρ^{-j} to get the corresponding structural coefficient $\overline{\gamma}_j$.

There is also a corresponding relationship between coefficient estimators, that helps explain the difficulty of estimating nonlinearities. Note that in this example a simple, unbiased estimator of the reduced form coefficient $\overline{\pi}_j$ is $\hat{\pi}_j = \sum_{i=1}^{n} y_i p_j(z_i)/n$. A corresponding estimator of the associated structural coefficient is $\hat{\gamma}_j = \rho^{-j}\hat{\pi}_j$. Thus, the structural estimator is also inflated by ρ^{-j}. Consequently, while the variance of the reduced form estimator can be quite stable across different values of j, the variance of the structural estimator will grow with j, making it more difficult to estimate the coefficients of higher-order terms.

For a concrete example, suppose the structural function is constant, the disturbance is homoskedastic, and the data are i.i.d. with $g_0(x) = c$ and $Var(\varepsilon|z) = \sigma^2$. Then it is easy to see that $Var(\hat{\pi}_j) = (\sigma^2 + c^2)/n$. Therefore, it follows that

$$Var(\hat{\gamma}_j) = \rho^{-2} Var(\hat{\gamma}_{j-1}), j \geq 3.$$

When $g_0(x)$ is not constant and ε is heteroskedastic this relationship will not hold, but $Var(\hat{\gamma}_j)$ will still increase rapidly with j.

This example shows how the strength of the instrument is related to the ability to estimate nonlinearities in the Gaussian x and z case. Note that ρ^2 is the R^2 from the population reduced form regression of x on z. This quantifies the strength of the instrument. As j increases the variance of $\hat{\gamma}_j$ goes up by a factor of ρ^{-2}. Thus, the stronger the instrument the lower the variance of estimators of coefficients of higher-order terms relative to coefficients of lower order terms.

For instance, in the Angrist and Krueger (1991) returns to schooling application, the reduced form R^2 for the 1930–1939 cohort, partialling out covariates, is 0.00133. The above example would predict that the variance of the coefficient of the quadratic coefficient $\hat{\gamma}_3$ is about 750 times the variance of the linear coefficient $\hat{\gamma}_2$, or that the confidence interval for the quadratic term is about 25 times as large as the linear term. Given that the linear IV already produces a wide confidence interval in this application, this calculation suggests that it would be very difficult indeed to find any nonlinearities when using quarter of birth as an instrument for schooling. Of course, the quarter of birth and years of schooling application does not have Gaussian z and x, but with such a low reduced form R^2 it may still be difficult to uncover nonlinearities.

With stronger instruments it is possible to uncover important nonlinearities in applications. Blundell, Chen, and Kristensen (2007) consider an application where x is the log of total expenditure and z is the log of income. In that application the reduced form R^2 is about 0.25. This R^2 would lead to the variance of the quadratic coefficient estimator being only about four times bigger than the variance of the linear coefficient estimator. Since the linear IV estimate is quite precise in those data one would expect to be able to uncover nonlinearities with NPIV. They do.

The severity of the ill-posed inverse problem in the Gaussian case is not extreme. It is known that if x and z are each uniformly distributed and $f(x|z)$ is analytic, one gets approximations similar to the Gaussian case, where structural coefficients are obtained by scaling up reduced form coefficients by something that is at least as large as ρ^{-j} for some $0 < \rho < 1$, see Theorem 5.20 of Kress (1999). Since any continuously distributed x and z can be normalized to be uniform, it takes some sort of nonsmoothness in $f(x|z)$ to ameliorate the ill-posed inverse problem. Such things as nonexistence of higher order derivatives or thick tails of distributions can do this, e.g., see Blundell, Chen, and Kristensen (2007).

V. Inference

A simple approach to inference with NPIV is to treat the series estimator as if it were parametric and use corresponding standard errors and the usual Gaussian approximation. Here this would correspond to using the Hansen (1982) and White (1982) estimator of the variance of $\hat{\gamma}$ as if it was 2SLS. This approach is asymptotically valid for nonparametric series regression, as shown in Newey (1997). It works there because the parametric calculation accounts correctly for the variance and the bias shrinks faster than the standard deviation in the asymptotic approximation. One might expect it to work for NPIV for the same reasons. Recent work has shown that it does work for some nonparametric IV series estimators, Chernozhukov and Chetverikov (2013).

Here we give a small Monte Carlo example with encouraging results for NPIV inference in the Gaussian example. We let x and z be bivariate Gaussian, $g_0(x)$ be constant, and ρ^2 be either 0.25 or 0.1. We take for approximating functions simple power series, where $p_{jJ}(x) = x^{j-1}$ and $q_{kK}(z) = z^{k-1}$. We use the Hansen (1982) and White (1982) estimator of the asymptotic variance of $\hat{\gamma}$ as a 2SLS estimator. We consider the coverage probability for an asymptotic 90 percent confidence interval for the value $g_0(E[x_i])$ of the structural function at the mean of x_i. We do 5,000 Monte Carlo replications. We choose various combinations of (n, J, K) so that the sample size should be big enough to lower the variance of $\hat{g}(E[x_i])$ as J grows.

For $\rho^2 = 0.25$ we find that for (n, J, K) triplets of $(100, 1, 2)$, $(500, 2, 6)$, and $(2{,}500, 3, 12)$, the coverage probabilities are 0.896, 0.899, and 0.905, respectively. For $\rho^2 = 0.1$ we find that for (n, J, K) triplets of $(100, 1, 2)$ and $(1{,}200, 2, 6)$, the coverage probabilities are 0.914 and 0.908, respectively. These results are consistent with the 2SLS standard errors giving a correct asymptotic approximation. We did find that, with larger sample sizes, nominal coverage was closer to 0.90 for larger K values.

These findings are consistent with the parametric standard errors giving a correct asymptotic approximation. We do note the large sample sizes required for this result as J increases. It could be that Santos' (2012) inference method, that is robust to nonidentification, gives better approximaions for smaller sample sizes. We leave this topic to future research.

REFERENCES

Abramowitz, Milton, and Irene A. Stegun, editors. 1965. *Handbook of Mathematical Functions: With Formulas, Graphs and Mathematical Tables.* Dover: Dover Publications.

Ai, Chunrong, and Xiaohong Chen. 2003. "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions." *Econometrica* 71 (6): 1795–1843.

Andrews, Donald W. K. 2011. "Examples of L^2-Complete and Boundedly-Complete Distributions." Yale University Cowles Foundation Discussion Paper 1801.

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106 (4): 979–1014.

Berry, Steven T., and Philip A. Haile. 2009. "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers." Yale University Cowles Foundation Discussion Paper 1718.

Blomquist, Soren, and Whitney Newey. 2002. "Nonparametric Estimation with Nonlinear Budget Sets." *Econometrica* 70 (6): 2455–80.

Blundell, Richard, Xiaohong Chen, and Dennis Kristensen. 2007. "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves." *Econometrica* 75 (6): 1613–69.

Blundell, Richard, and James L. Powell. 2003. "Endogeneity in Nonparametric and Semiparametric Regression Models." In *Advances in Economics and Econometrics, Theory and Applications,* Volume II, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, 312–57. New York: Cambridge University Press.

Canay, Ivan A., Andres Santos, and Azeem M. Shaikh. 2012. "On the Testability of Identification in Some Nonparametric Models with Endogeneity." CeMMAP Working Paper CWP18/12.

Chen, Xiaohong, Victor Chernozhukov, Sokbae Lee, and Whitney K. Newey. 2012. "Local Identification of Nonparametric and Semiparametric Models." CeMMAP Working Paper CWP37/12.

Chernozhukov, Victor, and Denis Chetverikov. 2013. "Asymptotic Normality of Linear Functionals of Nonparametric IV Estimators." Unpublished.

Chernozhukov, Victor, and Christian Hansen. 2005. "An IV Model of Quantile Treatment Effects." *Econometrica* 73 (1): 245–61.

Chernozhukov, Victor, Guido W. Imbens, and Whitney K. Newey. 2007. "Instrumental Variable Estimation of Nonseparable Models." *Journal of Econometrics* 139 (1): 4–14.

Chesher, Andrew. 2003. "Identification in Nonseparable Models." *Econometrica* 71 (5): 1405–41.

Darolles, S., Y. Fan, J. P. Florens, and E. Renault. 2011. "Nonparametric Instrumental Regression." *Econometrica* 79 (5): 1541–65.

Hall, Peter, and Joel L. Horowitz. 2005. "Nonparametric Methods for Inference in the Presence of Instrumental Variables." *Annals of Statistics* 33 (6): 2904–29.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4): 1029–54.

Hansen, Lars Peter, and Kenneth J. Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50 (5): 1269–86.

Horowitz, Joel L. 2006. "Testing a Parametric Model against a Nonparametric Alternative with Identification through Instrumental Variables." *Econometrica* 74 (2): 521–38.

Horowitz, Joel L. 2011. "Applied Nonparametric Instrumental Variables Estimation." *Econometrica* 79 (2): 347–94.

Imbens, Guido W., and Whitney K. Newey. 2009. "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity." *Econometrica* 77 (5): 1481–1512.

Kress, Rainer. 1999. *Linear Integral Equations.* 2nd Ed. New York: Springer-Verlag.

Matzkin, Rosa L. 2008. "Identification in Nonparametric Simultaneous Equations Models." *Econometrica* 76 (5): 945–78.

Newey, Whitney K. 1994. "Kernel Estimation of Partial Means and a General Variance Estimator." *Econometric Theory* 10 (2): 233–53.

Newey, Whitney K. 1997. "Convergence Rates and Asymptotic Normality for Series Estimators." *Journal of Econometrics* 79 (1): 147–68.

Newey, Whitney K., and James L. Powell. 1989. "Instrumental Variable Estimation of Nonparametric Models." Unpublished.

Newey, Whitney K., and James L. Powell. 2003. "Instrumental Variable Estimation of Nonparametric Models." *Econometrica* 71 (5): 1565–78.

Newey, Whitney K., James L. Powell, and Francis Vella. 1999. "Nonparametric Estimation of Triangular Simultaneous Equations Models." *Econometrica* 67 (3): 565–603.

Santos, Andres. 2012. "Inference in Nonparametric Instrumental Variables with Partial Identification." *Econometrica* 80 (1): 213–75.

White, Halbert. 1982. "Instrumental Variables Regression with Independent Observations." *Econometrica* 50 (2): 483–99.

This article has been cited by:

1. Byunghoon Kang. 2020. INFERENCE IN NONPARAMETRIC SERIES ESTIMATION WITH SPECIFICATION SEARCHES FOR THE NUMBER OF SERIES TERMS. *Econometric Theory* **26**, 1-35. [Crossref]
2. Tadao Hoshino. 2019. Sieve IV Estimation of Cross-Sectional Interaction Models with Nonparametric Endogenous Effect. *SSRN Electronic Journal* . [Crossref]
3. Xiaohong Chen, Timothy M. Christensen. 2018. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quantitative Economics* **9**:1, 39-84. [Crossref]
4. Nir Billfeld, Moshe Kim. 2018. Semiparametric Wavelet-Based Jpeg IV Estimator for Endogenously Truncated Data. *SSRN Electronic Journal* . [Crossref]
5. Nir Billfeld, Moshe Kim. 2018. Semiparametric Fourier-Dependent Sieve IV Estimator (SPIV) For Truncated Data. *SSRN Electronic Journal* . [Crossref]
6. Carlos Felipe Balcázar, Lidia Ceriani, Sergio Olivieri, Marco Ranzani. 2017. Rent-Imputation for Welfare Measurement: A Review of Methodologies and Empirical Findings. *Review of Income and Wealth* **63**:4, 881-898. [Crossref]
7. Rodrigo Adao, Arnaud Costinot, Dave Donaldson. 2017. Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade. *American Economic Review* **107**:3, 633-689. [Abstract] [View PDF article] [PDF with links]
8. Xiaohong Chen, Timothy Christensen. 2017. Optimal Sup-Norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric IV Regression. *SSRN Electronic Journal* . [Crossref]
9. Steven G. Dieterle, Andy Snell. 2016. A simple diagnostic to investigate instrument validity and heterogeneous effects when using a single instrument. *Labour Economics* **42**, 76-86. [Crossref]
10. Samuele Centorrino, Jeffrey Racine. 2016. Semiparametric Varying Coefficient Models with Endogenous Covariates. *SSRN Electronic Journal* . [Crossref]
11. Guy Tchuente. 2016. Estimation of Social Interaction Models Using Regularization. *SSRN Electronic Journal* . [Crossref]
12. Alina R. Buzanakova, Evgeniy M. Ozhegov. 2016. Demand for Performing Arts: The Effect of Unobserved Quality on Price Elasticity. *SSRN Electronic Journal* . [Crossref]
13. Xiaohong Chen, Timothy Christensen. 2015. Optimal Sup-Norm Rates, Adaptivity and Inference in Nonparametric Instrumental Variables Estimation. *SSRN Electronic Journal* . [Crossref]
14. Joel L. Horowitz. 2014. Ill-Posed Inverse Problems in Economics. *Annual Review of Economics* **6**:1, 21-51. [Crossref]
15. Patrick Kline, Enrico Moretti. 2014. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority *. *The Quarterly Journal of Economics* **129**:1, 275-331. [Crossref]
16. Melanie Meng Xue. 2014. Textiles and the Historical Emergence of Gender Equality in China. *SSRN Electronic Journal* . [Crossref]