Food Marketing Policy Center

Nonparametric Instrumental Variable Estimation in Practice

by Michael Cohen, Philip Shaw,
and Tao Chen

Food Marketing Policy Center
Research Report No. 111
November 2008

Research Report Series

http://www.fmpc.uconn.edu

University of Connecticut
Department of Agricultural and Resource Economics

Nonparametric Instrumental Variable Estimation in Practice†

Michael Cohen
University of Connecticut
Philip Shaw
Fairfield University
Tao Chen
University of Connecticut

October 20, 2008

†We thank: Joel Horowitz and Gautam Tripathi for valuable discussions about this research agenda. We have also benefitted from discussions and correspondence with Xiaohong Chen, Patrick Gagliardini, Whitney Newey and Oliver Scaillet. Michael Cohen also thanks the Food Marketing Policy Center at the University of Connecticut for financial support for this research and the Booth School of Business at the University of Chicago for providing a great place to conduct the research.

Abstract

In this paper we examine the finite sample performance of two estimators one developed by Blundell, Chen, and Kristensen (2007) (BCK) and the other by Gagliardini and Scaillet (2007) (TIR). This paper focuses on the generalization and expansion of these estimators to a full nonparametric specification with multiple regressors. In relation to the classic weak instruments literature, we provide intuition on the examination of instruments relevance when the structural function is assumed to be unknown.

Simulations indicate that both estimators perform quite well in higher dimensions. This research also provides insights on the performance of bootstrapped confidence intervals for both estimators. We document that the BCK estimator's coverage probabilities are near their nominal levels even in small samples as long as the sieve order of expansion is restricted. The coverage probability for the TIR estimator's bootstrapped confidence intervals are near their nominal levels even when the order of sieve approximation is large. These results suggest that in small samples the TIR estimator has a much smaller bias then the BCK estimator but its variance is much larger. We provide two empirical examples. One is the classic wage returns to education example and the other looks at the relationship of corruption and GDP to economic growth. Results here suggests that the impact of corruption on growth depends nonlinearly on a countries level of development.

JEL codes: **C13**, **C14**, **C15**

Keywords: Nonparametric, Instrumental Variables, Information Regularized Estimators

1 Introduction

Economists are often interested in estimating functional relationships between a dependant variable y and a set of explanatory variables $x = [y_2 \;\; x_1]$. If the functional relationship is known, for example the relationship is linear in the parameters, then it is straightforward to estimate the functional relationship between variables using well know parametric techniques. In most cases the functional relationship is not known which can result in model mis-specification. One way to overcome this problem is to estimate the functional relationship nonparametrically. In this case, the economist no longer estimates a set of parameters. Instead she estimates points on the unknown function. To illustrate this consider the following model which assumes the error enters additively.

$$y = \phi(x) + v \tag{1}$$

If v is conditionally independent of x, $E(v|x) = 0$, ϕ can be estimated using standard techniques. However if $E(v|x) \neq 0$ more information will be required to identify ϕ. The parametric solution is to select instruments conditionally independent of v and correlated with x. This works fine in the parametric case but using instruments for non-parametric identification is non-trivial. The difficulty arises because the mapping of the distribution of the data into the regression function is not continuous(Kress, 1999).

Newey and Powell (2003) overcome ill-posedness by imposing bounds on the higher order derivatives of ϕ effectively making the mapping continuous. A similar approach is followed by Ai and Chen (2003) for a semi-parametric estimation problem of the same variety. Both implement a two stage estimation procedure wherein the estimate of ϕ is given by series approximation.

$$\phi(x) \cong \sum_{j=1}^{J} \gamma_j \mathbb{E}[p_j(x)|z] \tag{2}$$

Where $p_j()$ is a sequence of basis functions. Therefore the first stage is to estimate $\mathbb{E}[p_j(x)|z]$ non-parametrically and estimate γ's in the second stage by some minimum distance criterion.

Blundell et al. (2007) is the only empirical application that has explored this type of estimator. They estimate a shape invariant Engle curve system which admits a semi-parametric form. In their model demographic scaling parameters enter parametrically and total expenditure is treated as endogenous. Their semiparametric model and estimation technique is adapted from Ai and Chen (2003). Hall and Horowitz (2005) present two methods for estimating ϕ based on kernel and series approximating regressions and derive optimal convergence rates.

Severini and Tripathi (2006) explore identification issues with these models and note that point wise identification can easily fail. Their work also provides intuition on how to determine the identified part of the structural function ϕ. They examine the relationship between identification of the structural function and identification of linear functionals and uncover a connection between them.

Newey and Powell (2003) show that in their framework a condition needed for identification is the completeness of the conditional distribution $f(x|z)$. Severini and Tripathi (2006) shows that completeness of the conditional distribution is equivalent to a correlation between the model space and the instrument space. As we will show later, ill-posedness can be caused by weak or irrelevant instruments in the sense that there is a very low correlation between the model space and the instrument space.

Gagliardini and Scaillet (2008) and Darolles, Florens, and Renault (2006) propose estimators which exploit information in the L_2 norms of ϕ and its first derivative by regularizing the second stage estimates with their Sobelev norm. This regularization penalizes the highly oscillating components to achieve a continuous mapping. They call their estimators Tikhonov Regularized estimators (TIR) after seminal work by Tikhonov (1963) that proposes this type of regularization for ill-posed inverse problems such as the one studied in this literature.

A related strand of literature studies specification testing. In this strand two varieties of specification questions are addressed. First Gagliardini and Scaillet (2007) use the non-parametric estimation techniques described above to construct statistics that test the validity of parametric specifications of econometric models. The test of Gagliardini and Scaillet (2007) is carried out in the spirit of Härdle and Mammen (1993). The test statistic uses a integrated square error distance metric that measures the distance between the parametric and non-parametric specifications. Gagliardini and Scaillet (2007) construct this test statistic using their Tikhonov regularized non-parametric estimate of ϕ. Horowitz (2006) tests the hypothesis that ϕ belongs to some finite dimensional parametric family against a non-parametric alternative. His test does not require the estimation of ϕ explicitly. Therefore rendering a solution to the ill-posed problem is avoided.

Because the test mentioned above is only reasonable when $\mathbb{E}[v|z] = 0$ Horowitz (2008) presents a test that examines whether the solution is smooth. This is reasonable because non-existence of a solution is an extreme form of non-smoothness. Horowitz also points out that even if a solution exists and is not smooth rejecting the model is still reasonable. This argument calls upon economic theories that dictate many models such as demand or Engle curves be smooth. Therefore rejection of non-smooth models is justified on two grounds, non existence and economic mis-specification. Horowitz shows that it is impossible to construct a uniformly consistent test on a unrestricted set of alternatives thus he restricts his test to smooth or nonsmooth functions over the null and alternative hypothesis. This is a reasonable restriction because in most economic applications the function, as guided by theory, is often smooth is the sense of having s many square-integrable derivatives.

The purpose of this paper is to examine the use of the non-parametric estimators briefly described above and provide evidence of finite sample performance while offering techniques to estimate multidimensional ϕ's for vector valued applications. We focus on the Blundell et al. (2007) (BCK) and Gagliardini and Scaillet (2008) (TIR) estimators and verify their performance in small samples. We explore the performance of bootstrap methods in the construction of confidence intervals, as well as discuss instrument relevance and speak about issues surrounding the estimation of partial effects all specifically in the context of these estimation techniques.

The paper is organized as follows. In the first section we will present the estimators and discuss practical estimation issues including the extension to multiple dimensions. The second section will discuss ill-posedness and how it relates to the literature on weak identification. Thirdly we conduct monte carlo experiments and present results from these simulations. The fourth section looks at two empirical applications. Finally concluding remarks are made along with suggestions for future work.

2 The Estimators

In this section we present the estimator constructed by Blundell et al. (2007) and Gagliardini and Scaillet (2007). Both methods used rely upon a minimum distance criteria that can be posed in a very general framework. Take the following model:

$$y = \phi(y_2) + v \tag{3}$$

Under the assumption that $E(v|y_2) \neq 0$ estimation of $\phi(y_2)$ by traditional nonparametric methods yields poor and meaningless results. Now assume we observe a variable z that satisfies the following condition $E(v|z) = 0$. Taking expectations over Equation (3) we obtain the following equation:

$$m(\phi(y_2), z) = E(\phi(y_2)|z) - E(y|z) = 0 \tag{4}$$

In operator notation let $T(\phi) = E(\phi(y_2)|z)$ and $r = E(y|z)$ so that we can write Equation (4) as:

$$T(\phi)(z) - r(z) = 0 \tag{5}$$

The solution to Equation (5) is said to be well-posed if the solution exists, is unique, and continuous in r. Ill-posedness occurs because T^{-1} need not be continuous. One approach to dealing with ill-posedness is through regularization which is generally represented as follows:

$$\phi_\lambda = \underset{\phi}{argmin} \|T(\phi) - r\|^2 + \lambda\|\phi\| \tag{6}$$

Both Blundell et al. (2007) and Gagliardini and Scaillet (2007) rely upon the method above to estimate ϕ_λ. $\|\phi\|$ is the norm of the function and its derivatives, commonly referred to as a Sobolev norm, used by both Blundell et al. (2007) and Gagliardini and Scaillet (2007). $\|\phi\|$ is a penalization matrix determined by the Sobolev norm of $\|\phi\|$, heretofore C, and a scaling parameter λ. An estimate of ϕ can be found by minimizing $\|T(\phi)-r\|^2$ subject to the constraint that $\|\phi\| \leq p$. In practice the constraint on the penalization matrix may be unknown therefore Equation (6) will be solved for a given value of λ as suggested by Blundell et al. (2007).

Gagliardini and Scaillet (2007) approximate the mean integrated squared error of their estimator and propose that in practice one should choose a penalization parameter λ such that the MISE is minimized. Although this practice is not theoretically justifiable data driven methods for selection of the penalization parameter improve the finite sample performance of the estimator as simulations conducted later show. Blundell et al. (2007) take a different approach suggesting that one might choose various values of penalization parameters and display the results for each choice.

The main difference between the TIR estimator and the BCK estimator is how each component of Equation (6) is estimated. For example, Blundell et al. (2007) estimate both $T(\phi)$ and r via sieve estimation while Gagliardini and Scaillet (2007) use a combination of sieve estimation and kernel methods to obtain a solution to the empirical counterpart to Equation (6). Both methods rely upon the calculation of the penalization matrix C. Blundell et al. (2007) provide the following equation characterizing their calculation of the $k_n^{p+1} \times k_n^{p+1}$ penalization matrix C:

$$C_r = \int [\nabla^r \Psi_k(y_2, x_1)]'[\nabla^r \Psi_k(y_2, x_1)]dy_2 \tag{7}$$

$\Psi_k(y_2, x_1)$ is the complete set of basis functions which is $N \times k_n^{p+1}$ if x_1 is a $N \times p$ matrix. As in Blundell et al. (2007) one might choose $r = 0$ and $r = 2$ so that C is constructed as:

$$C = \int [\Psi_k(y_2, x_1)]'[\Psi_k(y_2, x_1)]dy_2 + \int [\nabla^2 \Psi_k(y_2, x_1)]'[\nabla^2 \Psi_k(y_2, x_1)]dy_2 \tag{8}$$

The choice of the order of derivative should be chosen based upon the application at hand. In the case of Blundell et al. (2007), they choose the order of derivative based upon the type of underlying function they focus on estimating. In practice the derivative and integration is with respect to the endogenous variable only leaving this matrix as a function of the exogenous variables. One approach might be to calculate the value for the penalization matrix at the mean values of the remaining exogenous variables. In monte carlo simulations the choice of which point does not seem to make any quantitative difference. Blundell et al. (2007) show that their estimator $\hat{\phi}^{BCK}(y_2, x_1)$ has a closed form solution given by:

$$\hat{\Pi}_{\lambda}^{BCK} = (\Psi' B(B'B)^{-1} B'\Psi + \lambda C)^{-1} \Psi' B(B'B)^{-1} B'y \tag{9}$$

For a single instrumental variable, $B(z, x_1)$ is a $N \times k_n^{p+1}$ complete set of basis functions. We construct $\hat{\phi}^{BCK}(y_2, x_1)$ by multiplying our $N \times k_n^{p+1}$ basis functions $\Psi_k(y_2, x_1)$ by the k_n^{p+1} vector of coefficients $\hat{\Pi}_{\lambda}^{BCK}$. Similarly, Gagliardini and Scaillet (2007) show that their estimator also has a closed form solution expressed as:

$$\hat{\Pi}_{\lambda}^{TIR} = (\lambda_N C + \frac{1}{N} \hat{P}'\hat{P})^{-1} \frac{1}{N} \hat{P}'\hat{R} \tag{10}$$

$\hat{P} = \hat{E}(\Psi_k(y_2, x_1)|z, x_1)$ and $\hat{R} = \Omega(z, x_1)\hat{E}(\Psi_k(y_2, x_1)|z, x_1)$ where $\Omega(z, x_1)$ is an optimal weighting matrix. Gagliardini and Scaillet (2007) estimate both $\hat{P}$ and $\hat{R}$ by kernel methods.

2.1 Instrument Relevance

Classic foundational work by Nelson and Startz (1990) and Staiger and Stock (1997) documents the importance of paying close attention to the quality of instruments with respect to their relevance. When the function $\phi(x)$ is known to be linear in parameters, Staiger and Stock

(1997) show that an F-stat of 10 or greater on the instruments in the first stage regression is needed to have confidence in point estimation and thus inference. Under an unknown $\phi(x)$, not much is known regarding how one might characterize what Staiger and Stock (1997) refer to as "weak instruments". It is intuitive that estimation in the nonparametric case should require more from the data and thus it should be true that the rule of thumb proposed by Staiger and Stock (1997) should not uniformly apply to estimation methods under an unknown $\phi(x)$. With respect to the estimation method proposed by Blundell et al. (2007), they describe what they refer to as a sieve measure of ill-posedness:

$$\tau_n \equiv \sup_{\phi(y_2)\in\Upsilon_n:\phi\neq 0} \frac{\sqrt{E\{\phi(y_2)\}^2}}{\sqrt{E\{E[\phi(y_2)|z]\}^2}} \leq \frac{1}{\mu_{k_n}} \tag{11}$$

where μ_{k_n} are the singular values of the matrix T^*T. Defining $T(\phi)(\cdot) = E[\phi(y_2)|z = \cdot]$ and T^* is the adjoint operator of T. The singular values are particularly important as the rate of the decay serves as a measure of instrument relevance. The slower the rate of decay the faster the rate of convergence of the estimator. Ideally one estimates the values of μ_{k_n} and examines the rate of decay with respect to some benchmark. For example, if y_2 and z are jointly normal with correlation ρ then the rate of convergence of the singular values can be expressed as an exponential function. Unfortunately, the rate of convergence depends upon the underlying distribution of $f(y_2, z)$ which is unknown to the economist and the particular sieve being used. One approach is to assume an underlying distribution for $f(y_2, z)$ and then see if there is an optimal rate of decay for that distribution and particular basis function approximation. If the performance of these estimators is independent of the type of basis function being used, this might be a worthy approach. From this derivation a distance test can be derived for a given distribution of data. Notice that for Equation (11) when z is replaced by y_2, assuming exogeneity of y_2, then τ_n is identically 1. In practice we can estimate this value to see how close it is to 1. If the instruments are less than perfect then the value for Equation (11) will be away from 1. The problem is we don't necessarily know how far from 1 is too far. To get some idea of how this estimator is distributed Figure (1) shows the probability density of the empirical analog to Equation (11) across three levels of correlation between the instrument z and y_2 under joint normality for cosine basis functions of the order five. From the plot we can see that when the instruments are strong the estimator $\hat{\tau}_n$ is distributed tightly around a median value of 1.04. For $\gamma = .5$, $\hat{\tau}_n$ is centered around a median value of 1.14 as compared to a median value of 1.17 when the instruments are completely irrelevant ($\gamma = 0$). This is an interesting result because it suggest that we could observe values of $\hat{\tau}_n$ close to 1 even though our instruments are completely irrelevant. From this, it seems that using this as a measure of instrument relevance might not be terribly informative. However, it is not likely to observe values for $\hat{\tau}_n$ greater than 1.4 when $\gamma = .5$, so if large values of $\hat{\tau}_n$ are observed in practice it would certainly be an indicator of weak instruments.

3 Monte Carlo Simulations

3.1 Single Regressor Case

This section investigates the finite sample performance of the TIR and BCK estimators for a single endogenous regressor over various functional forms and orders of approximation. The data is generated by the following process, the joint density,

$$\begin{bmatrix} z \\ v \\ y_2 \end{bmatrix} \overset{\mathrm{d}}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \gamma & 0 \\ \gamma & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix} \right) \tag{12}$$

, and the functional relationship,

$$y = \beta_0 y_2^* + (1 - \beta_0)\sin(\alpha\pi y_2^*) + v \tag{13}$$

Notice that α controls the frequency of the function and β_0 the degree of linearity. Here $y_2^* = \Phi(\frac{y_2 - \bar{y_2}}{\sigma_{y_2}})$. We allow α to vary which controls the "waviness" of the function in which case the order of approximation should be higher in order to pick up on the higher frequency.

Tables (1) and (2) present coverage probabilities of the quantiles for the BCK and TIR estimators respectively. Table (1) displays coverage probabilities for the second through seventh order approximations of the regression function. It is clear that once the polynomial reach the fourth order the confidence intervals contain 96 percent of the bootstrapped estimates or better. On the other hand Table (2) displays the coverage probabilities for alternative forms of the data generation process. Here the polynomial's order of approximation is 6. Results here document that the estimator performs uniformly well over the different permutations of the function. In the next section we continue by investigating the case where our regression function has to independent variates.

3.2 Multiple Independent Variables

In this section we increase the dimension of the estimation problem and investigate the finite sample performance of the TIR and BCK estimators through Monte Carlo simulations. We generate the data as follows:

$$\begin{bmatrix} z \\ v \\ y_2 \end{bmatrix} \overset{\mathrm{d}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \gamma & 0 \\ \gamma & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}\right) \tag{14}$$

The additional exogenous variable is $x_1 \overset{\mathrm{d}}{\sim} N(0,1)$. We generate the dependent variable as follows:

$$y = \beta_0 y_2^* + (1-\beta_0)\sin(\alpha\pi y_2^*) + \beta_1 x_1^* + (1-\beta_1)\sin(\alpha\pi x_1^*) + v \tag{15}$$

Figures (1) and (2) portray a representative function in the top left, that has $\beta_0 = 0$,

$\beta_1 = 0$, and $\alpha = 2$. Figure (1) plots the TIR estimate of the function. A first glace reveals that the estimator picks up all the relevant curvature in the true function. Additionally, the estimator also picks up the relevant scale. Subplot c. For this figure shows that the measurement error is small particularly over the interior of the function, with larger error revealing itself on the perimeter of the grid where the true model generates observations with lower probability. Figure (2) plots the same set of diagrams for the BCK estimates of the true function. The BCK estimator does a relatively good job picking up the contours in the interior of the function's support, However in subplot b of Figure (2) the estimator misses the true function with greater regularity. A glance at subplot c. of this figure verifies that the estimator does well where it has more information on the data generation process but comes up short and to a comparatively greater extent than the TIR estimator.

Table (3) summarizes and supports the discussion of the preceding paragraph by providing some metrics on relative performance of the estimators as compared to traditional sieve estimates that do not account for the endogeneity. This table documents that the TIR estimates have and smaller point wise bias however the variance of the TIR estimates are more than twice that of the BCK estimates. The mean square error is also not surprisingly higher for the TIR than the BCK estimates. We have provided a results appendix to the paper that provides more such tables for small and larger sample sizes and various permutations of the parameters for the dependant variable generation function. The findings are consistent over all the results. In general we find that in smaller samples the TIR estimates have a significantly smaller bias than the BCK estimates. On the other hand the BCK estimates have smaller variance that the TIR estimates. The paper continues with two empirical examples.

4 Two Applications

In this section we present a classic text book example which is measuring the returns to education. It is clear that one may propose a relationship between wages and education as the following:

$$log(wage) = \phi(educ^*, X^*) + v \tag{16}$$

It is widely accepted that $E(v|educ) \neq 0$ thus education is endogenous in the above equation. If the function $\phi(educ, X)$ is known, then we can proceed with the usual two-stage least squares estimation. In this example, we do not proceed under the assumption that this function is known but instead rely upon the estimators above combined with bootstrapped confidence intervals to provide some guidance as to what the function $\phi(educ, X)$ looks like. For the first application, we assume only one additional exogenous variable is contained in X which we take to be age. Thus the structural function we are interested in is $\phi(educ, age)$. The methods outlined in section 2.1 for expanding the dimension of the estimator would allow us to include more exogenous variables, for expositional purposes and graphical parsimony we proceed with a bivariate example. We estimate the wage equation across a variety of values for k_n.

Figure (4) shows the TIR estimate of $\phi(educ^*, age^*)$. We transform values $educ$ and age where $educ = \Phi(\frac{educ^* - \bar{educ^*}}{\sigma_{educ^*}})$ and $age = \Phi(\frac{age^* - \bar{age^*}}{\sigma_{age^*}})$ to fall within the unit interval. This is done because the estimator utilizes shifted Chebyshev polynomials as basis functions on the unit interval. The null model is constructed as:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 age + v \tag{17}$$

We obtain the values for the above coefficients from OLS and then plot the function for those values. From this, we can then simply check to see if the model implied by the null model are within the 95% confidence intervals constructed via the bootstrap. The rejection plot in Figure (3) shows for which regions the null model is not within the confidence intervals. The results suggest that we are able to reject the null model over a range of age and education values. The TIR estimate reveals that the relationship between education, age, and wages are in fact nonlinear over the range for which the linear model is rejected. In Figure (3) we report the BCK estimates for $k_n = 3, 4$ and TIR estimates for $k_n = 4$. Although the BCK estimates show a larger degree of curvature, we fail to reject the null model across all grid points consequently we fail to reject the simple linear model.

This second economic relationship we examine that has been debated in the literature for some time is the relationship between economic growth and corruption. From very early work of Leff (1964), Huntington (1968), and Rose-Ackerman (1978) economists have been interested in the relationship between economic growth and corruption. It was argued that corruption could be good for growth or bad for growth depending upon various factors. Some economists believed that corruption could be beneficial to growth because it can act as "speed money" while other economists believe that it is detrimental to growth. Due to the nature of institutional development it also thought to be true that corruption and growth are jointly determined thus a measure of corruption is likely to be endogenous and thus classical assumptions as needed in OLS break down. Mauro (1995) was the first to recognize this fact and correct for it under the assumption that the relationship between corruption and growth is known. Shaw, Katsaiti, and Jurgilas (2006) later shows that his results are uninformative due to irrelevant instruments which actually result in unbounded confidence intervals. In the past, most empirical work while acknowledging the endogeneity, has thought linearly about the relationship between economic growth and corruption. Most models assume a standard linear framework and employ the classical 2SLS. In this empirical example we estimate the relationship between growth and corruption in a purely nonparametric framework allowing for initial conditions to enter the model nonparametrically. Thus we estimate the following model:

$$growth = \phi(cor, gdp70) + v \tag{18}$$

We look at the average growth rate over the period of 1960-2004 (growth) using logarithmic of gross domestic product in 1970 (gdp70) as initial conditions. We chose 1970 as the base year only because it significantly increased our sample size which ends up as 96. The sample size is quite small but both estimators can perform moderately well for a moderate choice of k_n. We also measure the average level of corruption (cor) over the period of 1980-1997. The data for corruption was taken from Dreher, Kotsogiannis, and McCorriston (2007). The instrumental variable we use is a measure of ethnolinguistic fractionalization in 1985.

Notice that practically speaking the BCK and TIR estimators produce very different results. For the null model we present the unconditional mean of the growth rate across countries. We reject the null over a certain range of corruption and initial conditions. For this region we construct a contour plot in Figure (7) to get a better idea of what is going on in this range. Notice that regardless of the initial conditions there is a negative and significant effect of corruption on economic growth. One thing worth noticing is that the rate at which corruption impacts growth differs depending upon the initial conditions of the country. For example consider a country with initial conditions of .5. If corruption in that country increases from .20 to .35 then the growth rate will fall by .5%. For a country with initial conditions of approximately .73 the same increase in corruption would yield a 1% decrease in long-run growth. Therefore the impact of corruption on economic growth depends upon a countries level of development.

5 Conclusion

This research provides an exploration of recently developed methods for estimating functions nonparametrically that are identified by instrumental variables. We provided a practical discussion of two similar estimators and introduced how they could be practically extended to consider models with multiple regressions that enter a function in some unknown fashion. A practical discussion of instrument relevance is provided that guides practitioners in their choice of instruments in the context of the nonparametric instrumental variable estimation problem. Monte Carlo simulations present a comparison of the finite sample properties of the two estimators. We find that the TIR estimator has a much smaller bias then the BCK estimator but its variance is much larger.

Two empirical applications are explored. The classic wage education example is presented as a familiar context in which readers may evaluate the estimators. The second empirical application showcases an example where the use of nonparametric instrumental variable estimators may be used to provide important insights on questions that have, until now, been largely left to the speculation of inflexible priors on model specification. Nonparametric methods prove to be a useful way to conduct economic research without imposing dogmatic structure on new economic problems. The importance of dealing with endogenously determined relationships is of utmost importance to empirical practitioners and it is our hope that this research provides a stepping stone to further adoption of these techniques by economists in all fields.

References

Ai, C., & Chen, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, *71*(6), 1795–1843.

Blundell, R., Chen, X., & Kristensen, D. (2007). Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves. *Econometrica*, *75*(6), 1613–1669.

Darolles, S., Florens, J., & Renault, E. (2006). Nonparametric Instrumental Regression. Tech. rep., GREMAQ, University of Social Science, Toulouse, France.

Dreher, A., Kotsogiannis, C., & McCorriston, S. (2007). Corruption around the world: Evidence from a structural model. *Journal of Comparative Economics*, *35*(3), 443–466.

Gagliardini, P., & Scaillet, O. (2007). A Specification Test for Nonparametric Instrumental Regresion. Tech. rep., University of Lugano and Swiss Finance Institute.

Gagliardini, P., & Scaillet, O. (2008). Tikhonov Regularization for Nonparametric Instrumental Variable Estimators. Tech. rep., University of Lugano and Swiss Finance Institute.

Hall, P., & Horowitz, J. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, *33*(6), 2904.

Härdle, W. K., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, *21*(4), 1926–1947.

Horowitz, J. (2006). Testing a Parametric Model Against a Nonparametric Alternative with Identification Through Instrumental Variables. *Econometrica*, *74*(2), 521–538.

Horowitz, J. (2008). Specification Testing in Nonparametric Instrumental Variables Estimation. Northwestern University, Department of Economics.

Huntington, S. P. (1968). *Political Order in Changing Societies*. Yale University Press, New Haven, CT.

Kress, R. (1999). *Linear Integral Equations*. Springer.

Leff, N. (1964). Economic Development Through Bureaucratic Corruption. *American Behavioral Scientist*, 8–14.

Mauro, P. (1995). Corruption and Growth. *Quarterly Journal of Economics*, *110*(3), 681–712.

Nelson, C. R., & Startz, R. (1990). Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, *58*(4), 967–76.

Newey, W., & Powell, J. (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, pp. 1565–1578.

Rose-Ackerman, S. (1978). *Corruption: A Study in Political Economy.* Academic Press, New York, NY.

Severini, T., & Tripathi, G. (2006). Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors. *Econometric Theory*, *22*(02), 258–278.

Shaw, P., Katsaiti, M.-S., & Jurgilas, M. (2006). Corruption and Growth Under Weak Identification. Working papers 2006-17, University of Connecticut, Department of Economics.

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, *65*, 557–86.

Tikhonov, A. (1963). On regularization of ill-posed problems. In *Dokl. Akad. Nauk SSSR*, Vol. 153, pp. 49–52.

10%	25%	50%	75%	90%	k_n	J_n
0.96	0.36	0.94	0.38	0.92	2	10
0.98	0.42	0.97	0.43	0.97	3	10
0.97	0.97	0.98	0.97	0.96	4	10
0.97	0.99	1.00	1.00	0.98	5	10
0.99	1.00	1.00	1.00	1.00	6	10
1.00	1.00	1.00	1.00	1.00	7	10

Table 1: Coverage probabilities for BCK bootstrapped confidence intervals by quantile: $N = 100$, $\rho = .50$, $\gamma = .85$, $\beta_0 = .5$, $\alpha = 2$, $\lambda = 0$

10%	25%	50%	75%	90%	β_0	α
0.94	0.96	0.92	0.94	0.91	0.00	1.00
0.94	0.97	0.92	0.95	0.94	0.50	1.00
0.94	0.95	0.96	0.94	0.92	1.00	1.00
0.97	0.84	0.95	0.85	0.96	0.00	2.00
0.96	0.86	0.96	0.88	0.96	0.50	2.00
0.94	0.95	0.96	0.94	0.94	1.00	2.00

Table 2: Coverage probabilities for TIR bootstrapped confidence intervals by quantile: $N = 100$, $\rho = .50$, $\gamma = .85$, and $k_n = 6$

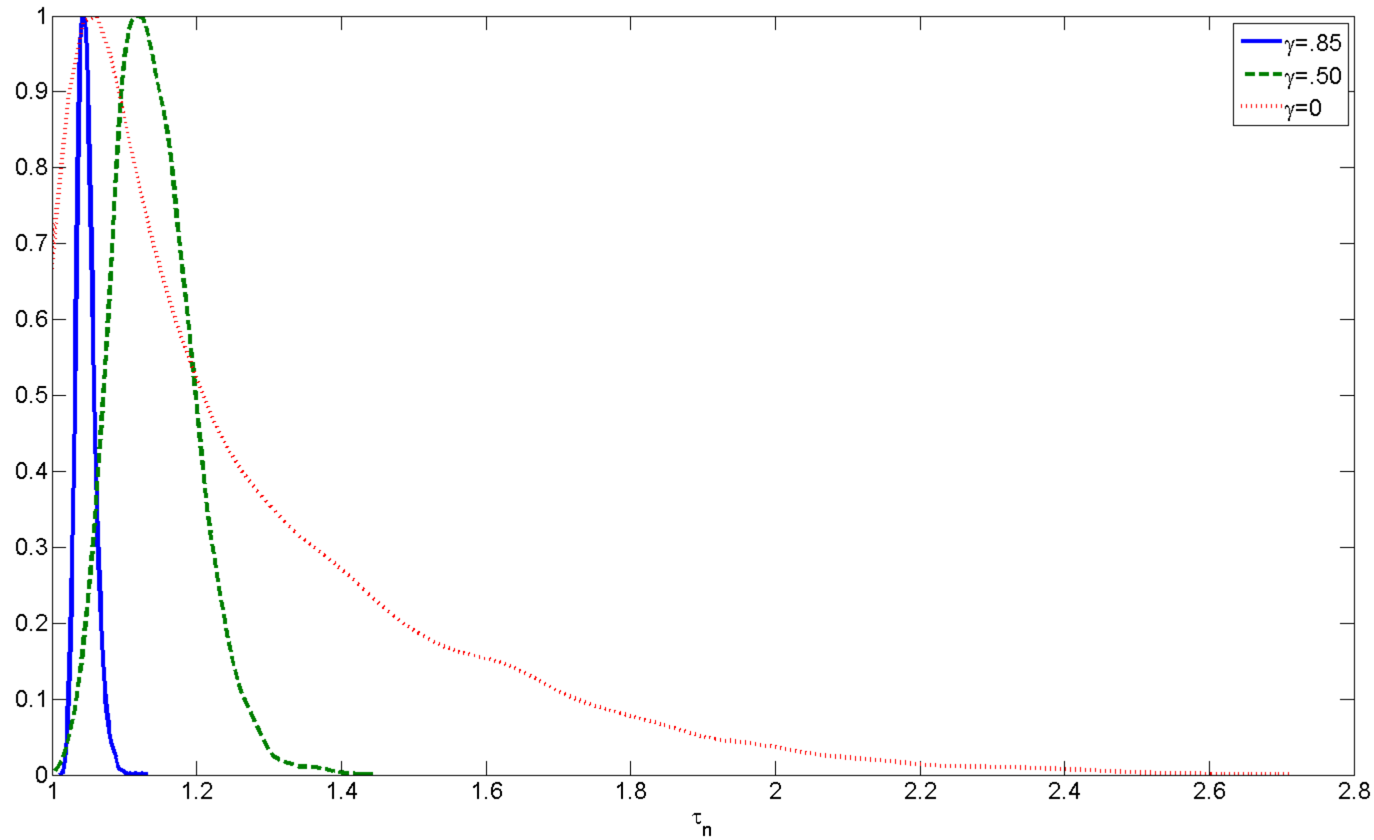
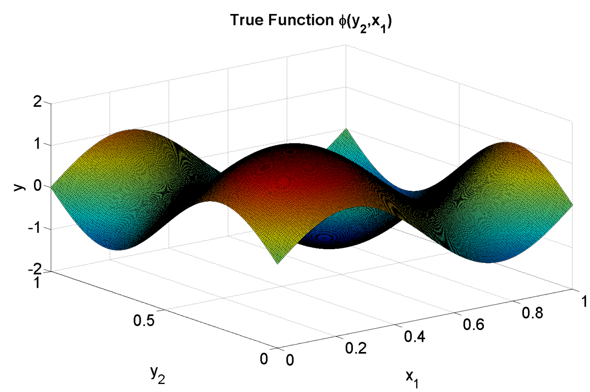
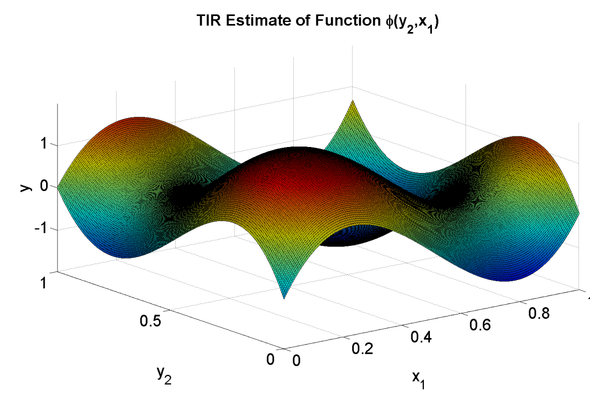


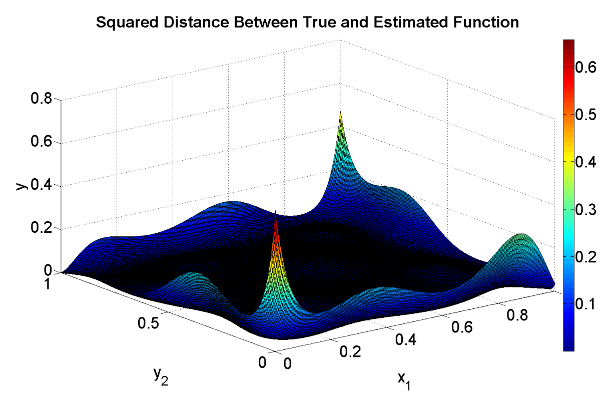
Figure 1: Probability distribution of sieve measure of ill-posedness τ_n for $k_n = 5$ cosine basis functions.



(a) True Function

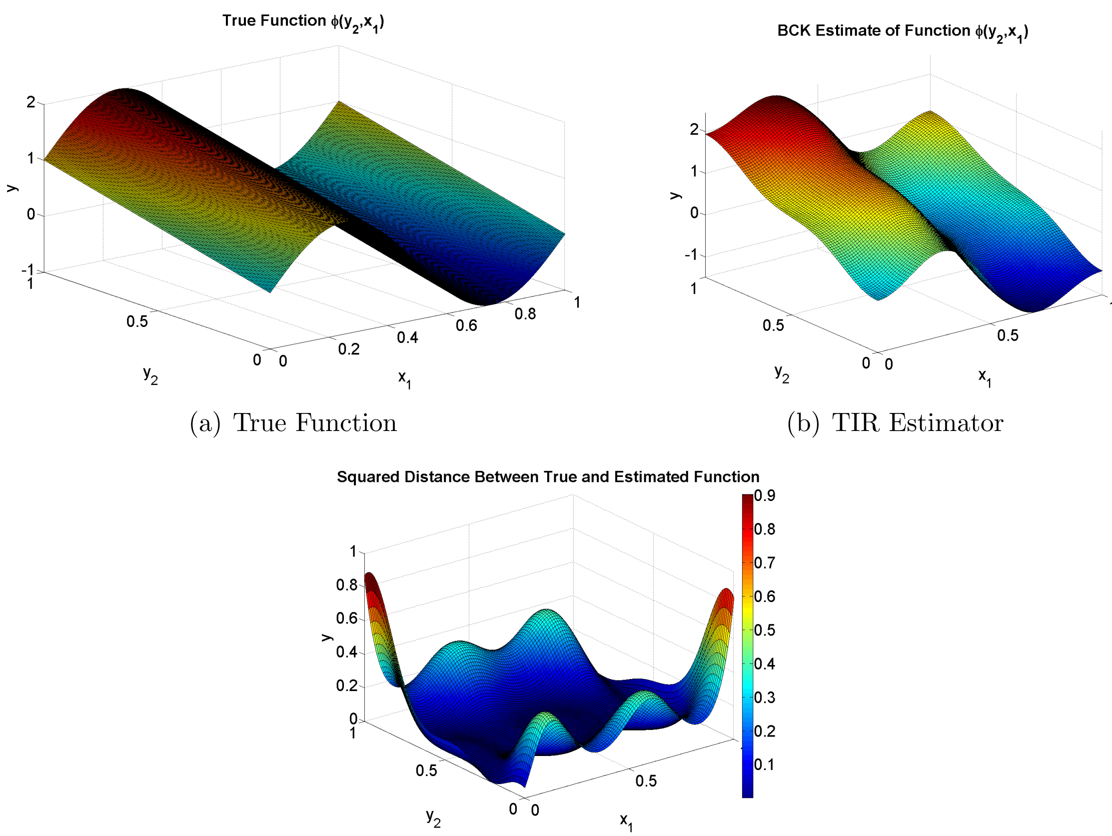


(b) TIR Estimator



(c) Squared Distance

Figure 2: Estimated function versus true function for $N = 300$, $\rho = .50$, $k_n = 5$, $\beta_0 = 0$, $\beta_1 = 0$, $\alpha = 2$, and $\gamma = .50$.



(a) True Function

(b) TIR Estimator

(c) Squared Distance

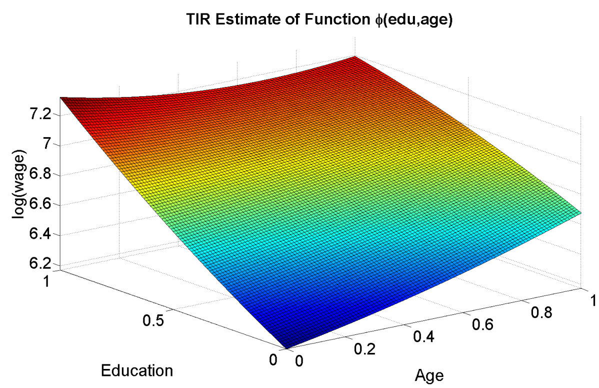
Figure 3: Estimated function versus true function for $N = 300$, $\rho = .50$, $k_n = 5$, $\beta_0 = 1$, $\beta_1 = 0$, $\alpha = 2$, and $\gamma = .50$.

Table 3: Performance of BCK ($\lambda = 0$) and TIR estimators by average pointwise $Bias^2$, Var, and MSE

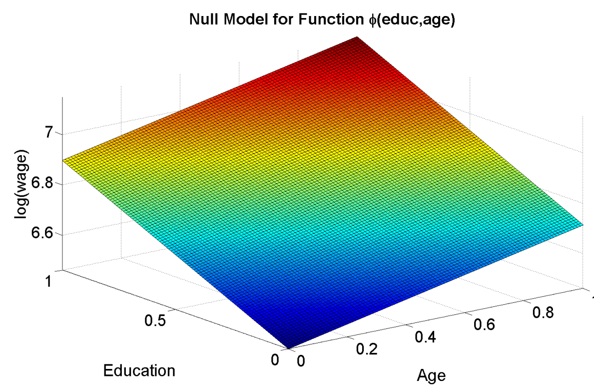
N=300, $k_n = 5$, $\gamma = .5$, $\rho = .5$			
	TIR	BCK	OLS
	$\beta_0 = 0$, $\beta_1 = 0$		
$Bias^2$	0.016	0.121	0.256
Var	0.588	0.239	0.077
MSE	0.604	0.360	0.333

Table 4: Performance of BCK ($\lambda = 0$) and TIR estimators by average pointwise $Bias^2$, Var, and MSE

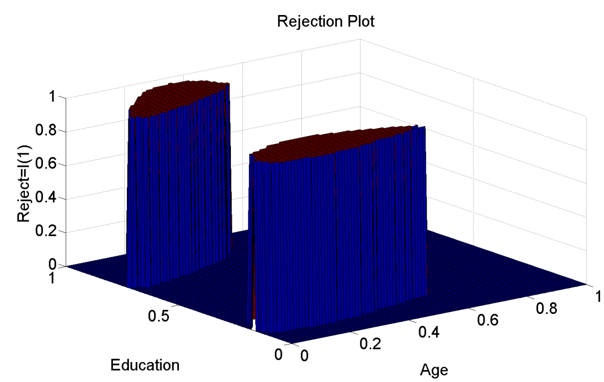
N=1000, $k_n = 5$, $\gamma = .5$, $\rho = .85$			
	TIR	BCK	OLS
	$\beta_0 = 0$, $\beta_1 = 0$		
$Bias^2$	0.049	0.069	0.715
Var	0.179	0.189	0.010
MSE	0.228	0.258	0.725



(a) TIR Estimate

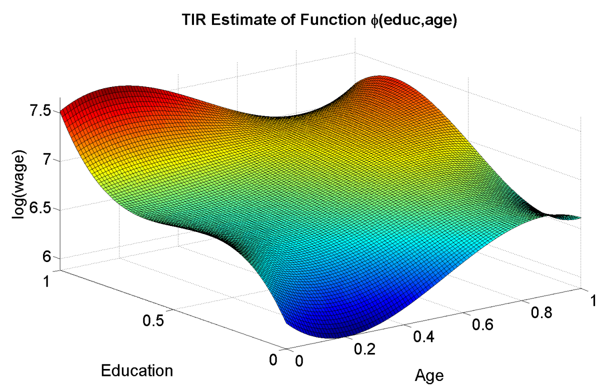


(b) Null Model

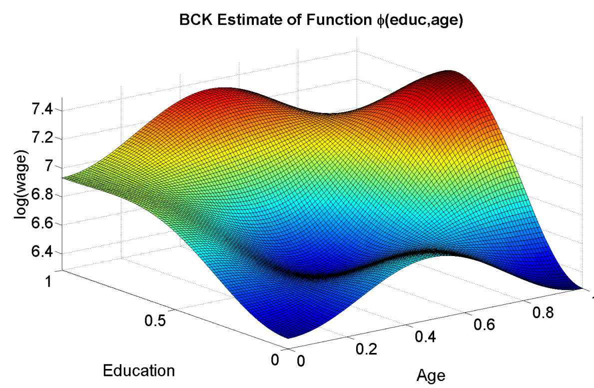


(c) Rejection Plot

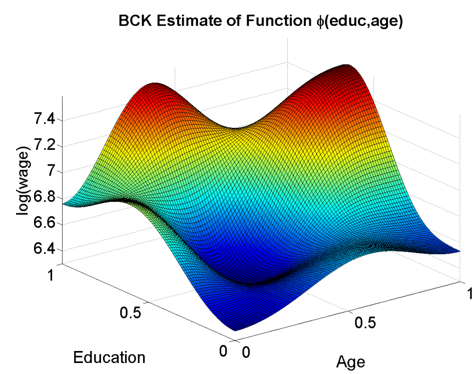
Figure 4: TIR estimate of $\phi(educ, age)$, null model, and rejection plot for bootstrapped 95% confidence intervals for $k_n = 3$.



(a) TIR Estimate $k_n = 4$

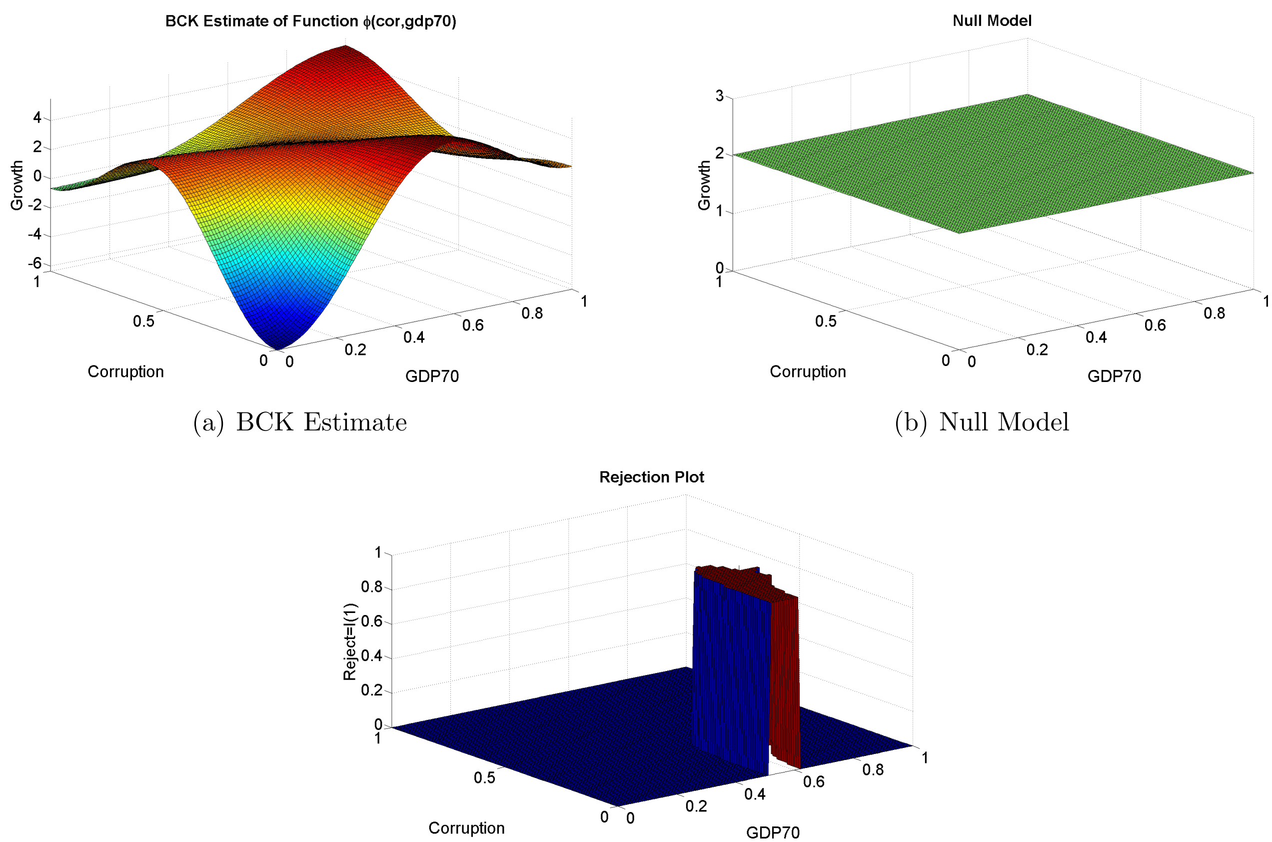


(b) BCK Estimate $k_n = 3$



(c) BCK Estimate $k_n = 4$

Figure 5: TIR and BCK estimates of $\phi(educ, age)$ across various k_n.



(a) BCK Estimate

(b) Null Model

(c) Rejection Plot

Figure 6: BCK ($\lambda = .01$) estimate of $\phi(cor, gdp70)$, null model, and rejection plot for bootstrapped 95% confidence intervals for $k_n = 3$.

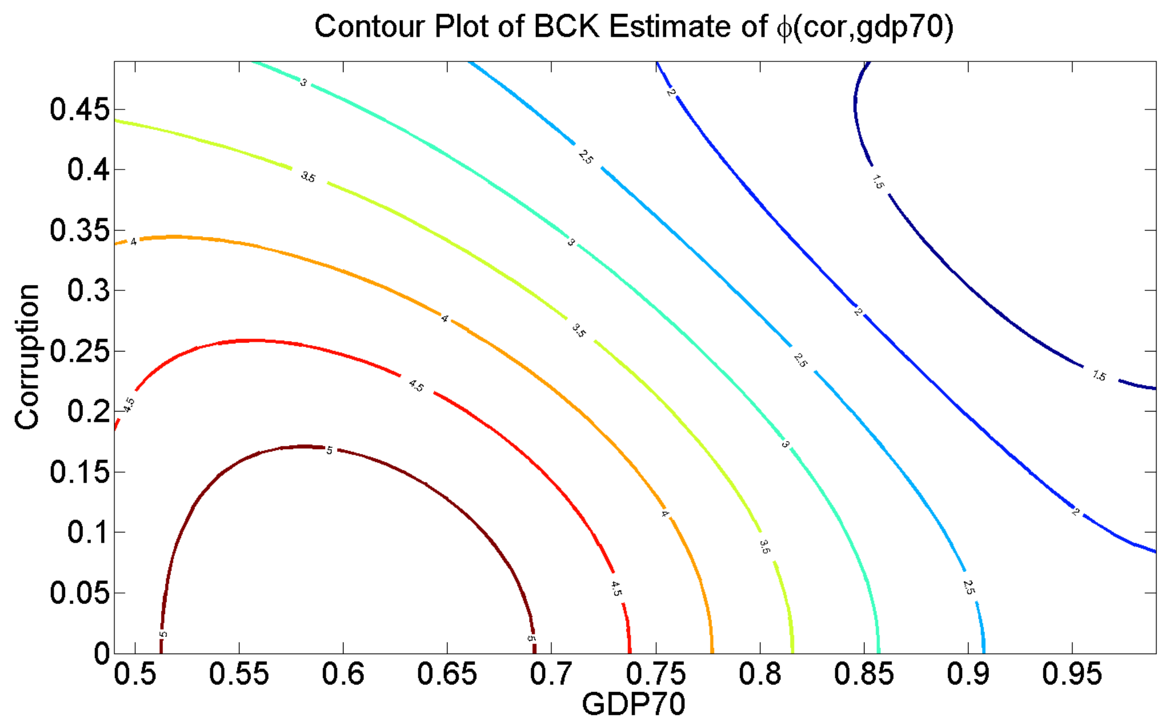
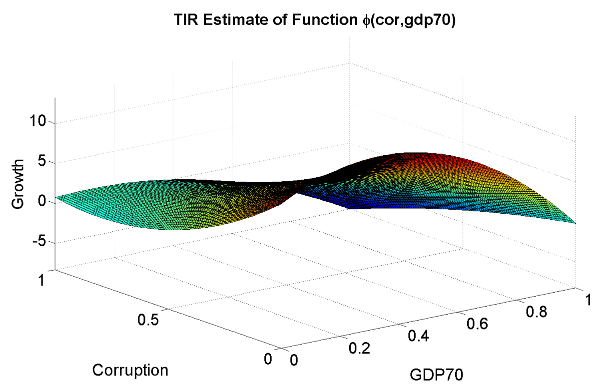
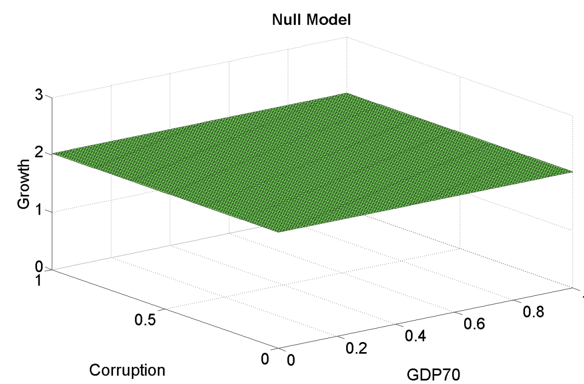


Figure 7: Contour Plot of Growth on Corruption and GDP70.



(a) TIR Estimate

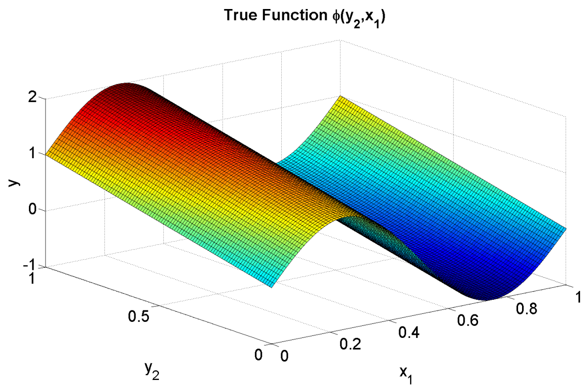


(b) Null Model

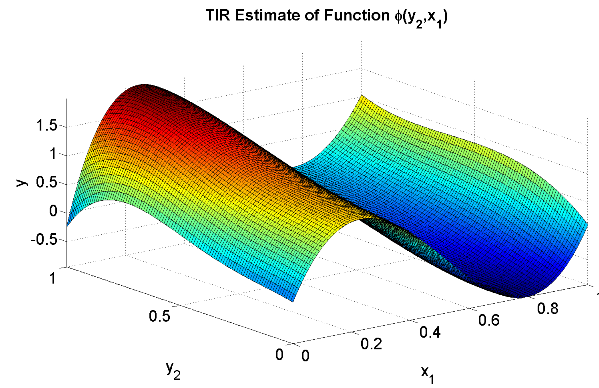
(c) Rejection Plot

Figure 8: TIR estimate of $\phi(cor, gdp70)$, null model, and rejection plot for bootstrapped 95% confidence intervals for $k_n = 3$.

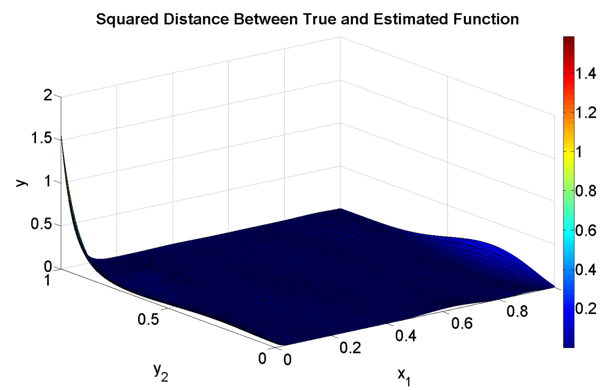
6 Results Appendix



(a) True Function

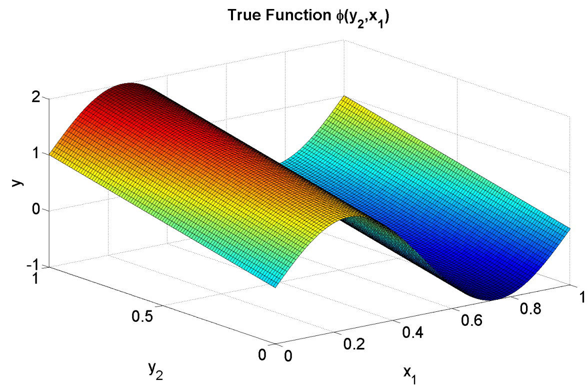


(b) TIR Estimator

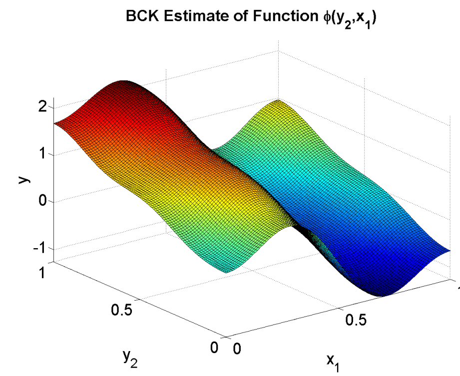


(c) Squared Distance

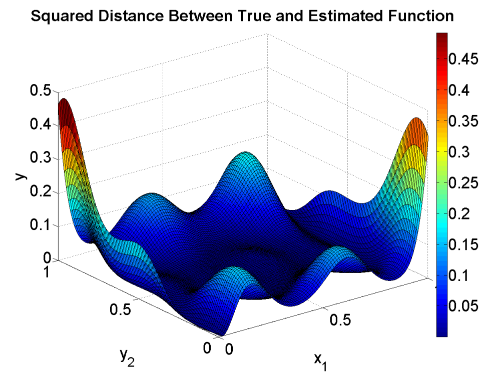
Figure 9: Estimated function versus true function for $N = 300$, $\rho = .50$, $k_n = 5$, $\beta_0 = 1$, $\beta_1 = 0$, $\alpha = 2$, and $\gamma = .85$.



(a) True Function



(b) BCK Estimator



(c) Squared Distance

Figure 10: Estimated function versus true function for $N = 300$, $\rho = .50$, $k_n = 5$, $\beta_0 = 1$, $\beta_1 = 0$, $\alpha = 2$, and $\gamma = .85$.

Table 5: Performance of BCK ($\lambda = 0$) and TIR estimators by average pointwise $Bias^2$, Var, and MSE

N=300, $k_n = 5$, $\gamma = .5$, $\rho = .5$			
	TIR	BCK	OLS
	$\beta_0 = 1$, $\beta_1 = 0$		
$Bias^2$	0.018	0.115	0.250
Var	0.571	0.232	0.074
MSE	0.588	0.347	0.324
	TIR	BCK	OLS
	$\beta_0 = 1$, $\beta_1 = 1$		
$Bias^2$	0.039	0.102	0.240
Var	0.628	0.228	0.073
MSE	0.667	0.330	0.312

Table 6: Performance of BCK ($\lambda = 0$) and TIR estimators by average pointwise $Bias^2$, Var, and MSE

N=1000, $k_n = 5$, $\gamma = .5$, $\rho = .85$			
	TIR	BCK	OLS
	$\beta_0 = 1$, $\beta_1 = 0$		
$Bias^2$	0.057	0.057	0.701
Var	0.178	0.165	0.009
MSE	0.235	0.222	0.710
	TIR	BCK	OLS
	$\beta_0 = 1$, $\beta_1 = 1$		
$Bias^2$	0.048	0.081	0.713
Var	0.185	0.192	0.011
MSE	0.233	0.273	0.725

10%	25%	50%	75%	90%	k_n	J_n
0.95	0.90	0.97	0.91	0.95	2	10
0.95	0.91	0.96	0.94	0.94	3	10
0.95	0.96	0.95	0.95	0.96	4	10
0.96	0.98	0.99	0.98	0.97	5	10
1.00	0.99	1.00	0.99	0.99	6	10
1.00	1.00	1.00	1.00	1.00	7	10

Table 7: Coverage probabilities for BCK bootstrapped confidence intervals by quantile: $N = 1000$, $\rho = .50$, $\gamma = .85$, $\beta_0 = 1$, $\alpha = 1$, $\lambda = 0$

10%	25%	50%	75%	90%	k_n	J_n
0.94	0.92	0.97	0.91	0.94	2	10
0.95	0.94	0.98	0.93	0.95	3	10
0.95	0.98	0.99	0.99	0.96	4	10
0.98	1.00	1.00	1.00	0.98	5	10
0.99	1.00	1.00	1.00	0.99	6	10
1.00	1.00	1.00	1.00	1.00	7	10

Table 8: Coverage probabilities for BCK bootstrapped confidence intervals by quantile: $N = 100$, $\rho = .50$, $\gamma = .85$, $\beta_0 = 1$,$\alpha = 1$, $\lambda = 0$

10%	25%	50%	75%	90%	k_n	J_n
0.88	0.86	0.06	0.92	0.28	2	10
0.96	0.94	0.94	0.89	0.89	3	10
0.97	0.99	0.94	0.97	0.92	4	10
0.99	1.00	1.00	0.98	0.96	5	10
0.99	1.00	1.00	0.99	0.99	6	10
1.00	1.00	1.00	1.00	1.00	7	10

Table 9: Coverage probabilities for BCK bootstrapped confidence intervals by quantile: $N = 100$, $\rho = .50$, $\gamma = .85$, $\beta_0 = 0$, $\alpha = 1$, $\lambda = 0$

10%	25%	50%	75%	90%	k_n	J_n
0.88	0.00	0.94	0.00	0.95	2	10
0.92	0.00	0.95	0.00	0.96	3	10
0.94	0.94	0.95	0.93	0.95	4	10
0.95	0.98	0.99	0.97	0.96	5	10
1.00	0.99	0.99	0.98	1.00	6	10
1.00	1.00	1.00	1.00	1.00	7	10

Table 10: Coverage probabilities for BCK bootstrapped confidence intervals by quantile: $N = 1000$, $\rho = .50$, $\gamma = .85$, $\beta_0 = .5$, $\alpha = 2$, $\lambda = 0$

FOOD MARKETING POLICY CENTER
RESEARCH REPORT SERIES

This series includes final reports for contract research conducted by Policy Center Staff. The series also contains research direction and policy analysis papers. Some of these reports have been commissioned by the Center and are authored by especially qualified individuals from other institutions. (A list of previous reports in the series is available on our web site.) Other publications distributed by the Policy Center are the Working Paper Series, Journal Reprint Series for Regional Research Project NE-165: *Private Strategies, Public Policies, and Food System Performance*, and the Food Marketing Issue Paper Series. Food Marketing Policy Center staff contribute to these series. Individuals may receive a list of publications in these series and paper copies of older Research Reports are available for $20.00 each, $5.00 for students. Call or mail your request at the number or address below. Please make all checks payable to the University of Connecticut. Research Reports can be downloaded free of charge from our web site given below.

Food Marketing Policy Center
1376 Storrs Road, Unit 4021
University of Connecticut
Storrs, CT 06269-4021

Tel: (860) 486-1927
FAX: (860) 486-2461
email: *fmpc@uconn.edu*
http://www.fmpc.uconn.edu