

This is a postprint version of the following published document:

Escanciano, J. C. (2017). A simple and robust estimator for linear regression models with strictly exogenous instruments. *The Econometrics Journal*, 21 (1), pp. 36-54.

DOI: [10.1111/ectj.12087](https://doi.org/10.1111/ectj.12087)

A Simple and Robust Estimator for Linear Regression Models with Strictly Exogenous Instruments*

Juan Carlos Escanciano[†]

Indiana University

November 1, 2016

Abstract

This paper investigates estimation of linear regression models with strictly exogenous instruments under minimal identifying assumptions. The paper introduces a uniformly (in the data generating process) consistent estimator under nearly minimal identifying assumptions. The proposed estimator, called the Integrated Instrumental Variables (IIV) estimator, is a simple weighted least squares estimator and does not require the choice of a bandwidth or tuning parameter, or the selection of a finite set of instruments. Thus, the estimator is extremely simple to implement. Monte Carlo evidence supports the theoretical claims and suggests that the IIV estimator is a robust complement to optimal IV in finite samples. In an application with quarterly UK data, IIV estimates a positive and significant elasticity of intertemporal substitution and an equally sensible estimate for its reciprocal, in sharp contrast to IV methods that fail to identify these parameters.

Keywords: Uniform identification; Instrumental variables; Weak instruments; Uniform inference; Intertemporal elasticity of substitution.

JEL classification: C13, C26.

*I thank Lin Zhu for excellent research assistance, and Don Andrews, Carol Caetano, Marine Carrasco, Michael Jansson, Jeff Wooldridge, Ed Vytlacil and two anonymous referees for useful comments. A previous SSRN Working Paper version of this paper was circulated under the title: “The Integrated Instrumental Variables Estimator: Exploiting Nonlinearities for Identification of Linear Models”, dated February 2, 2010.

[†]Department of Economics, Indiana University, 105 Wylie Hall, 100 S. Woodlawn Avenue, Bloomington, IN 47405-7104, USA. E-mail: jescanci@indiana.edu.

1 Introduction

The linear regression model continues to be the workhorse model in empirical research. In the presence of endogenous regressors these models are often estimated by Instrumental Variables (IV) methods, where “exogeneity” and “relevance” conditions are required for the validity of the resulting inferences. The recent literature on weak instruments has emphasized the practical importance of uniform inferences (uniform in the underlying probability generating the data, see, e.g., Staiger and Stock, 1997, Dufour, 1997, Mikusheva, 2013). When estimators are uniformly consistent the sample size required to achieve good approximations does not depend on the (unknown) data generating process. The lack of uniformly consistent estimators in the standard IV setting, caused by the possibility of weak instruments, has motivated a focus on testing and confidence intervals (see Stock, Wright and Yogo, 2002, and Andrews and Stock, 2005, for recent surveys of the literature). This paper departs from this literature by investigating robust (i.e. uniform) *estimation* with strictly exogenous instruments.

This paper first discusses identification under a minimal assumption (i.e. a necessary and sufficient condition for identification) that we call *Linear Completeness* (in reference to the completeness assumption used in nonparametric IV, see Newey and Powell 2003). Classical IV methods may or may not identify under Linear Completeness, a point made in Dominguez and Lobato (2004). Interestingly, we show that in the present linear setting Optimal IV (OIV) identifies the structural parameters under the minimal identifying assumption. The OIV estimator, however, requires the choice of tuning parameters (e.g. bandwidths) for its implementation, and its root-mean square error seems to be very sensitive to the data generating process. This paper proposes an estimator, called the Integrated Instrumental Variables (IIV) estimator, that identifies under Linear Completeness, it is simple to implement and performs uniformly well in finite samples. We “explain” this excellent finite sample performance by showing that the IIV is uniformly consistent under nearly the minimal identifying assumption. The IIV estimator is a simple weighted least squares estimator and does not require the choice of a bandwidth or tuning parameter. Monte Carlo simulations show that the IIV estimator compares favorably with OIV estimators in terms of bias and root-mean square error. The Monte Carlo also uncovers a relative robustness of OIV confidence intervals to weak identification. The practical utility of the IIV estimator is shown in an application where it estimates a positive and significant elasticity of intertemporal substitution with quarterly UK data, and an equally sensible estimate for its reciprocal, in sharp contrast to IV methods that fail to identify these parameters and provide puzzling low estimates for both parameters (see, e.g., Yogo (2004) and references therein).

The observed data is a random sample from $W \equiv (Y, X', Z)'$, which is defined on the probability space (Ω, \mathcal{F}, P) and takes values in $\mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^k$, $p, k \in \mathbb{N}$. Henceforth, A' and $|A|$ denote the matrix transpose and the Euclidean norm of A , respectively. For simplicity of notation we drop the qualification “almost surely” in equalities between random variables. The components of $W \equiv (Y, X', Z)'$ satisfy $E[|Y|] < \infty$, $E[|X|^2] < \infty$,

$$Y = X'\theta_0 + \varepsilon, \tag{1}$$

where $\theta_0 \in \mathbb{R}^p$, and Z is a vector of strictly exogenous instruments, i.e.

$$E[\varepsilon|Z] = 0. \tag{2}$$

This paper deals with identification and estimation of the model defined in (1)-(2).

The linearity in (1) is often justified on the basis of simplicity and has been shown to be a useful approximation in many economic applications. Economic theory or randomization can be used to justify (2). For example, in dynamic Euler equations the strict exogeneity naturally holds provided Z is in the agent’s information set.

We now introduce the minimal identifying assumption for θ_0 in (1)-(2). We call this assumption *Linear Completeness*. Two distinct parameters $\theta_0 \neq \theta_1$ are observationally equivalent in the setting above if and only if

$$E[Y - X'\theta_0|Z] = E[Y - X'\theta_1|Z],$$

or equivalently

$$E[X'(\theta_0 - \theta_1)|Z] = 0.$$

Therefore, a necessary and sufficient condition for identification is a linear completeness assumption (see Newey and Powell (2003) for a related discussion of completeness in a fully nonparametric IV setting), i.e., for all $\lambda \in \mathbb{R}^p$,

$$E[X'\lambda|Z] = 0 \implies \lambda = 0.$$

Linear completeness can be characterized by a semiparametric rank condition as follows. Define the optimal predictor $\tilde{Z} := E[X|Z]$. Then, linear completeness means: $\tilde{Z}'\lambda = 0$ implies $\lambda = 0$, or equivalently

ASSUMPTION LC: $E[\tilde{Z}\tilde{Z}']$ is positive definite.

Assumption LC is a nonparametric generalization of the classical rank condition, in the sense that the nonparametric regression $\tilde{Z} = E[X|Z]$ is used rather than the parametric linear regression used in classical IV. Unlike fully nonparametric completeness conditions, see Canay, Santos and Shaikh (2013), Assumption LC is testable; see the empirical application below for nonparametric tests for Assumption LC in the context of linearized Euler equations. Assumption LC has been used before in the literature in the context of (1)-(2), see Donald and Newey (2001, Assumption 2(i)). Related assumptions in different models have been considered in e.g. Das (2005) and Cai et al. (2006). The necessity of Assumption LC is a contribution relative to Donald and Newey (2001). The first result, whose proof follows from the arguments above and therefore is omitted, establishes the minimality of Assumption LC for identification.¹

PROPOSITION 1: *In (1)-(2), θ_0 is identified if and only if Assumption LC holds.*

A popular method to estimate (1)-(2) is by IV, where the researcher chooses a finite set of instruments, i.e. a measurable transformation $f(Z) \in \mathbb{R}^m$, $m \geq p$, and uses the moment restrictions

$$E[(Y - X'\theta_0)f(Z)] = 0,$$

to identify and estimate θ_0 . From Proposition 1 it follows that Assumption LC is implied by the traditional rank condition $rank(E[f(Z)X']) = p$ (and $rank(E[f(Z)f(Z)']) = m$). However, the IV rank condition may not hold even when the model (1)-(2) is identified (see Dominguez and Lobato, 2004), depending on the choice of f and the data generating process. Thus, researchers must exercise some care when choosing a finite set of instruments f . Interestingly and in sharp contrast with the nonlinear setting of Dominguez and Lobato (2004), in the linear case the OIV estimator identifies θ_0 under the minimal Assumption LC (and an additional moment condition), as we now show. The OIV estimator uses the optimal instruments $Z^* := f^*(Z) = \sigma^{-2}(Z)m(Z)$, where $\sigma^2(Z) := E[\varepsilon^2|Z]$ and $m(Z) := E[X|Z]$; see Amemiya (1974, 1977), Robinson (1976), Chamberlain (1982), and Newey (1990). Then, identification of OIV means full rank of

$$E[\sigma^{-2}(Z)m(Z)X'] = E[\sigma^{-2}(Z)m(Z)m(Z)'],$$

which is equivalent to Assumption LC, provided $\sigma^2(Z) < \infty$. Thus, the OIV is not only appealing from an asymptotic efficiency point of view, but also from an identification point of view. Unfortunately, implementing the OIV requires nonparametric estimation of $m(Z)$ and

¹Given that Assumption LC is sufficient and necessary for identification, we refer to it as the minimal identifying assumption in the rest of the paper.

$\sigma^2(Z)$, with the associated choice of bandwidth or tuning parameters, in addition to further smoothness assumptions on suitable nonparametric curves; see e.g. Donald and Newey (2001) and Carrasco (2012). Nonparametric estimation of these quantities may severely affect the finite sample performance of the OIV for moderate and large values of k and common sample sizes. Monte Carlo simulations below provide evidence of the lack of robustness of OIV estimators, even when $k = 1$.

This paper introduces a simple estimator that identifies θ_0 under the minimal Assumption LC. Identification is achieved by using a continuum of instruments, in the same way the characteristic function uses a continuum of moments. The Integrated Instrumental Variables (IIV) estimator of θ_0 is a weighted least squares estimator, computed for a sample $\{W_i\}_{i=1}^n$ of size n as

$$\hat{\theta}_{IIV} := (\mathbf{X}'\Omega\mathbf{X})^{-1}(\mathbf{X}'\Omega\mathbf{Y}), \quad (3)$$

where \mathbf{X} is the $n \times p$ design matrix with rows X_i' , $\mathbf{Y} = (Y_1, \dots, Y_n)'$, Ω is the $n \times n$ matrix with elements $\exp(-0.5(Z_i - Z_s)'\hat{V}^{-1}(Z_i - Z_s))$, $1 \leq i, s \leq n$, and \hat{V} is the sample variance of $\{Z_i\}_{i=1}^n$. Section 2 provides motivation for this estimator, whereas Section 3 shows that $\sqrt{n}(\hat{\theta}_{IIV} - \theta_0)$ is (uniformly) asymptotically normal with an asymptotic variance that is consistently estimated by

$$\hat{\Gamma} = \hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Sigma}^{-1}, \quad (4)$$

where $\hat{\Sigma} = (\mathbf{X}'\Omega\mathbf{X})/n^2$, $\hat{\Lambda} = (\mathbf{X}'\Omega D\Omega\mathbf{X})/n^3$ and D is a diagonal matrix with components $\{\hat{\varepsilon}_i^2\}_{i=1}^n$, where $\hat{\varepsilon}_i = Y_i - X_i'\hat{\theta}_{IIV}$ are IIV residuals. The IIV estimator and related inferences can be implemented by standard OLS methods: multiply Y and X by $\Omega^{1/2}$ and run OLS inference on the transformed data.

The IIV estimator $\hat{\theta}_{IIV}$ is extremely simple to compute and it has an excellent finite sample performance, uniformly under different data generating processes. We formalize this finite sample behaviour by showing that it is uniformly consistent and uniformly asymptotically normal under a wide set of data generating processes. The IIV estimator is also more robust to high-dimensionality of Z than alternative minimum distance estimators proposed in Dominguez and Lobato (2004). The simplicity and robustness of the IIV estimator should be appealing to empirical researchers. This paper also contributes to the extensive literature on nonparametric testing based on a continuum of moment conditions, see e.g. Bierens (1982), Stinchcombe and White (1998) and Escanciano (2006), to the extent that the existing literature has been focussed on nonparametric testing, rather than on estimation. It also provides new applications to testing for nonparametric/semiparametric identification. Finally, the asymptotic theory of this paper, based on Hilbert space methods, leads to weaker conditions than the related asymptotic theory based on the supremum norm (see e.g. Dominguez and

Lobato, 2004 and references therein).

The rest of the paper is organized as follows. Section 2 introduces formally the IIV estimator, and shows that this estimator identifies θ_0 under the minimal identifying assumption. Section 3 establishes the uniform consistency and uniform asymptotic normality of the IIV estimator. Readers not interested in the technical underpinnings can skip Sections 2 and 3. Section 4 reports the results of some Monte Carlo experiments showing that the proposed estimator has an excellent performance in finite samples and compares favorably with OIV estimators. Section 5 shows the practical utility of the results with an application to estimating the elasticity of intertemporal substitution for international data. Section 6 concludes and discusses future research. An Appendix gathers the mathematical proofs of the main results.

2 The Integrated Instrumental Variables Estimator

This section introduces the IIV estimator. To that end, define

$$h_{Y,P}(u) := E_P[Y \exp(iu'V^{-1/2}Z)], \quad u \in \mathbb{R}^k, \quad (5)$$

where $i := \sqrt{-1}$, $V := E_P[ZZ']$ and where henceforth the dependence on P will be emphasized. Likewise, define the moment functions $h_{X,P}(u) := E_P[X \exp(iu'V^{-1/2}Z)]$ and $h_{\varepsilon,P}(u) := E_P[\varepsilon \exp(iu'V^{-1/2}Z)]$. Note that the strict exogeneity implies $h_{\varepsilon,P}(\cdot) \equiv 0$. Following Bierens (1982), the reciprocal is also true; $h_{\varepsilon,P}(\cdot) \equiv 0$ implies the strict exogeneity condition (2). The functions $h_{Y,P}$, $h_{X,P}$ and $h_{\varepsilon,P}$ are in general complex-valued. The definition of the Euclidean norm is extended to complex numbers as $|A| = (tr(A^c A))^{1/2}$, where A^c denotes the complex conjugate of the complex number A .

Since $h_{Y,P}(u)$ is linear in Y , substituting $Y = X'\theta_0 + \varepsilon$ in (5) yields

$$h_{Y,P}^c(u) = h_{X,P}^c(u)\theta_0, \quad (6)$$

Multiplying both sides of (6) by $h_{X,P}(u)$, evaluating u at $u = U$, with U a standard Gaussian random vector in \mathbb{R}^k , and taking expectations with respect to U yield the normal equations

$$E_U[h_{X,P}(U)h_{Y,P}^c(U)] = E_U[h_{X,P}(U)h_{X,P}^c(U)]\theta_0. \quad (7)$$

There are other distributions different from a standard Gaussian that could be used for U , but the standard Gaussian has the advantage of leading to simple closed form expressions for $E_U[h_{X,P}(U)h_{X,P}^c(U)]$ and $E_U[h_{X,P}(U)h_{Y,P}^c(U)]$ (and to the simple expression for the es-

timator provided above). Moreover, identification is achieved under the minimal identifying assumption as shown in the following result proved in the Appendix.

PROPOSITION 2: *Assumption LC is equivalent to*

$$E_U[h_{X,P}(U)h_{X,P}^c(U)] \text{ is positive definite.} \quad (8)$$

In view of Proposition 2, under Assumption LC, θ_0 is identified as

$$\theta_0 \equiv \theta_0(P) = (E_U[h_{X,P}(U)h_{X,P}^c(U)])^{-1} E_U[h_{X,P}(U)h_{Y,P}^c(U)]. \quad (9)$$

This identification result and the analog principle suggest the IIV estimator

$$\hat{\theta}_{IIV} \equiv \theta_0(P_n) = (E_U[h_{X,P_n}(U)h_{X,P_n}^c(U)])^{-1} E_U[h_{X,P_n}(U)h_{Y,P_n}^c(U)], \quad (10)$$

where P_n is the empirical probability measure pertaining to $\{W_i \equiv (Y_i, X'_i, Z'_i)'\}_{i=1}^n$. That is, $h_{Y,P_n}(u) := n^{-1} \sum_{s=1}^n Y_s \exp(iu'\hat{V}^{-1/2}Z_s)$ and $h_{X,P_n}(u) := n^{-1} \sum_{s=1}^n X_s \exp(iu'\hat{V}^{-1/2}Z_s)$. After some simple algebra it can be shown that $\hat{\theta}_{IIV}$ can be computed as in (3).

Following the logic of Bierens (1982), Stinchcombe and White (1998) and others, alternative estimators different from $\hat{\theta}_{IIV}$ can be constructed with similar uniform consistent properties if quantities such as $E_P[Y \exp(iu'V^{-1/2}Z)]$ are replaced by $E_P[Yw(Z, u)]$, for other suitable choices of $w(Z, u)$. For instance, for $w(Z, u) = 1(Z \leq u)$, where $Z \leq u$ is understood coordinate-wise, one gets an estimator in the spirit of that proposed by Dominguez and Lobato (2004) in a more general set-up. Note the indicator $1(Z \leq u)$ may lead to misleading inferences in applications where there is a large number of instruments, i.e. large k , as many $1(Z_i \leq u)$ could be zero when u is evaluated at the sample observations; see Escanciano (2006) for further discussion and substantial empirical evidence. The choice $\exp(iu'V^{-1/2}Z_t)$ leads to a computationally simple estimator that is robust to moderate values of the number of instruments and performs well in practice.

3 Uniform Asymptotic Theory for the IIV Estimator

Uniformity of inferences in the underlying probability measure P over a large class of probabilities is considered a desirable property.² Assume $P \in \mathcal{P}$, where \mathcal{P} is the class of proba-

²Of course, as noted by Bahadur and Savage (1956), the class cannot be too large.

bilities

$$\mathcal{P} := \left\{ \begin{array}{l} P : E_P [Y - X'\theta_0 | Z] = 0, E_P [|Y|] \leq C, E_P [|X|^2] \leq C, E_P [|Z|^2] \leq C \\ \lambda_{\min} (E_P [ZZ']) \geq C \text{ and } \lambda_{\min} (E_P [m_P(Z)m'_P(Z)]) \geq C \end{array} \right\},$$

where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of a symmetric matrix A , and C is a generic positive constant that may change from expression to expression. The independence of C on P will be useful in proving uniformity results. The condition $\lambda_{\min} (E_P [m_P(Z)m'_P(Z)]) \geq C$ is stronger than Assumption LC (which allows for C to depend on P). Henceforth, we refer to this stronger condition as the “nearly minimal identifying assumption”. Note that $\lambda_{\min} (E_P [m_P(Z)m'_P(Z)]) \approx 0$ corresponds to “weak identification” in the nonparametric sense of this paper. Thus, our uniformity results require strong identification of the conditional model.

This section establishes uniform asymptotic theory for $\hat{\theta}_{IIV}$.³ To that end, it is convenient to use asymptotic theory for random elements taking values in a suitable Hilbert space; see, e.g., van der Vaart and Wellner (1996, Chapter 1.8). Let $L_2(\phi)$ be the Hilbert space of all complex-valued and square ϕ -integrable functions, where ϕ is the k -dimensional multivariate standard Gaussian density. In $L_2(\phi)$ define the inner product

$$\langle f, g \rangle := \int_{\mathbb{R}^k} f(u)g^c(u)\phi(u)du.$$

$L_2(\phi)$ is endowed with the natural Borel σ -field induced by the norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$. We denote by $\xrightarrow{L_2}$ convergence in probability in $L_2(\phi)$, i.e., $h_n \xrightarrow{L_2} h \iff \|h_n - h\| \xrightarrow{P} 0$. We use uniform in \mathcal{P} versions of the $\xrightarrow{L_2}$ convergence to prove the uniform consistency of $\hat{\theta}_{IIV}$.

Using the introduced notation, the IIV estimator can be compactly written as

$$\hat{\theta}_{IIV} = \langle h_{X,P_n}, h_{X,P_n} \rangle^{-1} \langle h_{X,P_n}, h_{Y,P_n} \rangle, \quad (11)$$

where, with some abuse of notation, we have denoted by $\langle h_{X,P_n}, h_{X,P_n} \rangle$ the $p \times p$ matrix with elements $\langle h_{X_j,P_n}, h_{X_k,P_n} \rangle$, where h_{X_j,P_n} is the j -th component of h_{X,P_n} . Now, uniform consistency of $\hat{\theta}_{IIV}$ will follow from the uniform law of large numbers in $L_2(\phi)$ for h_{X,P_n} and h_{Y,P_n} , i.e. from $h_{X,P_n} \xrightarrow{L_2} h_{X,P}$ and $h_{Y,P_n} \xrightarrow{L_2} h_{Y,P}$, uniformly in $P \in \mathcal{P}$, and the continuity of the inner product.

³Uniform convergence in distribution means the following: $T_n \rightarrow_d T$ uniformly over \mathcal{P} iff $\sup_{P \in \mathcal{P}} |E_P[f(T_n)] - E_P[f(T)]| \rightarrow 0$ for all bounded and continuous functions $f(\cdot)$.

THEOREM 1: *The IIV estimator is uniformly consistent over \mathcal{P} , i.e. for all $\varepsilon > 0$,*

$$\sup_{P \in \mathcal{P}} P \left(\left| \hat{\theta}_{IIV} - \theta_0 \right| > \varepsilon \right) \longrightarrow 0.$$

Theorem 1 and some further results proved below show that, uniformly in $P \in \mathcal{P}$, the IIV estimator satisfies the following expansion

$$\sqrt{n}(\hat{\theta}_{IIV} - \theta_0) = \langle h_{X,P}, h_{X,P} \rangle^{-1} \sqrt{n} \langle h_{X,P}, h_{\varepsilon, P_n} \rangle + o_P(1). \quad (12)$$

The term $\sqrt{n} \langle h_{X,P}, h_{\varepsilon, P_n} \rangle$ has the uniform Bahadur expansion

$$\sqrt{n} \langle h_{X,P}, h_{\varepsilon, P_n} \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i H_{X,P}(Z_i) + o_P(1), \quad (13)$$

where $H_{X,P}(u) := E_P[X \exp(-0.5(u - Z)' V^{-1}(u - Z))]$. Equations (12) and (13) then show that the IIV estimator is uniformly asymptotically linear.

For uniform asymptotic normality, we need to restrict the class of probabilities to

$$\mathcal{P}^{AN} = \mathcal{P} \cap \left\{ P : E_P \left[|\varepsilon|^{2+\delta} \right] \leq C \right\},$$

where $\delta > 0$. This restriction allows the application of a uniform Central Limit Theorem (CLT) given in Petrov (1975, p. 118), which in turn yields the uniform asymptotic normality of $\sqrt{n}(\hat{\theta}_{IIV} - \theta_0)$ in the following result. Define $\Sigma_P := \langle h_{X,P}, h_{X,P} \rangle$, $\Lambda_P := E_P[\varepsilon^2 H_{X,P}(Z) H'_{X,P}(Z)]$ and $\Gamma_P := \Sigma_P^{-1} \Lambda_P \Sigma_P^{-1}$.

THEOREM 2: *The IIV estimator is uniformly asymptotically normal in \mathcal{P}^{AN} , i.e*

$$\sqrt{n}(\hat{\theta}_{IIV} - \theta_0) \rightarrow_d N(0, \Gamma_P) \text{ uniformly in } \mathcal{P}^{AN}.$$

Note that Σ_P is non-singular uniformly in \mathcal{P}^{AN} (cf. Proposition 1). The estimation of the asymptotic variance Γ_P was given in (4). Confidence regions and Wald tests based on the variance estimator and the asymptotic normality in Theorem 2 inherit the uniformity of the IIV estimator under additional assumptions that bound the eigenvalues of Λ_P away from zero, as shown in the following corollary. Define

$$\mathcal{P}^{INF} = \mathcal{P}^{AN} \cap \{ P : \lambda_{\min}(\Lambda_P) \geq C \}.$$

Let $\chi_{p,1-\alpha}^2$ denote the $(1 - \alpha)$ -th quantile of the chi-square distribution with p degrees of

freedom, and denote the asymptotic $(1 - \alpha)$ -th confidence region by

$$C_{n,\alpha} := \{\theta \in \mathbb{R}^p : n(\hat{\theta}_{IIV} - \theta)' \hat{\Gamma}^{-1} (\hat{\theta}_{IIV} - \theta) \leq \chi_{p,1-\alpha}^2\}.$$

COROLLARY 1: *IIV inference is uniformly valid on \mathcal{P}^{INF} , i.e*

$$\left| \sup_{P \in \mathcal{P}^{INF}} P(\theta_0 \in C_{n,\alpha}) - (1 - \alpha) \right| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

There is previous research in statistics and econometrics on uniform inference. This work has mainly focussed on uniformly valid confidence sets and tests.⁴ Abstract conditions for uniform consistency were given in Lecam and Schwartz (1960). More closely related to the current setting, Kasy (2015) investigates uniformity in the Delta Method, and shows that the IV estimator fails to be uniformly asymptotically normal in the classical IV setting. To the best of our knowledge, the uniformity of the OIV has not been studied. A simple observation shows that this estimator fails to be uniformly consistent over \mathcal{P} , as $\sigma_P^2(Z)$ can be made arbitrarily close to zero or to infinity. See also the general results in Pötscher (2002), who shows the lack of uniformly consistent estimators of $m_P(Z)$ and $\sigma_P^2(Z)$ over \mathcal{P} . These results are independent of the estimators used for $m_P(Z)$ and $\sigma_P^2(Z)$, and apply equally to kernel, series or other estimators that solve ill-posed problems. These results, however, do not directly imply the lack of uniformity for functionals of these nonparametric objects. Obtaining uniformity results for semiparametric estimators such as OIV is beyond the scope of this paper. Monte Carlo simulations below provide empirical evidence of the lack of uniform consistency for OIV estimators, and suggest that the new estimator significantly outperforms OIV estimates in terms of bias and mean square error when the minimal identifying Assumption LC is close to being violated.

4 Monte Carlo Simulations

This section investigates the finite sample performance of the proposed estimator ($\hat{\theta}_{IIV}$) in comparison with the OIV estimator ($\hat{\theta}_{OIV}$) in two simulation experiments.

In the first experiment we generate data Y , X and Z according to the model

$$DGP1 : \begin{cases} Y = \theta_{01} + \theta_{02}X + \varepsilon \\ X = \gamma Z + v \end{cases}, \quad \begin{pmatrix} \varepsilon \\ v \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

⁴See, e.g., Imbens and Manski (2004), Leeb and Pötscher (2005), Mikusheva (2007), Andrews and Guggenberger (2009) and Belloni, Chernozhukov, Fernandez-Val and Hansen (2015).

with Z a standard normal, drawn independently of ε and v . The true slope parameter is zero, i.e. $\theta_{02} = 0$, and $\theta_{01} = 1$. Several values for the parameters (γ, ρ) are considered. Since the determinant of $E_P[m_P(Z)m'_P(Z)]$ is γ^2 , γ measures the level of nonparametric identification. The correlation parameter ρ measures the level of endogeneity. In the simulations we consider several combinations of the parameter values $\gamma \in \{0.1, 0.25, 0.5\}$ and $\rho \in \{0.4, 0.8\}$. The number of Monte Carlo replications is 5000. All estimators are computed using standardized instruments (i.e. using $Z_i/\hat{\sigma}$, where $\hat{\sigma} = \sqrt{\hat{V}}$ is the sample standard deviation of $\{Z_i\}$).

The optimal estimator $\hat{\theta}_{OIV}$ requires nonparametric estimation of $m_P(Z)$ and $\sigma_P^2(Z)$. We estimate these quantities with Nadaraya-Watson estimators using a Gaussian kernel and a bandwidth $h_n = 1.06\hat{\sigma}n^{-1/3}$, which attains the optimal rate of convergence in the sense of minimizing the mean-square error of the resulting estimator, see Linton (2002). The term $1.06\hat{\sigma}$ is not necessarily the optimal one, but we prefer to keep the exposition simple. We have considered other bandwidth choices, but they led to qualitatively similar conclusions.

Table 1 reports the bias (BIAS), standard error (SE), root mean squared (RMS) error, and coverage probability (CP) at 95% for the slope OIV and IIV estimates for several values of (γ, ρ) in DGP1 with $n = 50, 200$ and 500 . It is evident from Table 1 that the bias and standard error of the OIV estimator is very sensitive to the values of the parameters (γ, ρ) . In contrast, the proposed estimator's RMS error has a satisfactory performance across the different parameter values (i.e. uniformly in the data generating processes). In particular, the RMS errors for the IIV estimator are not sensitive to the level of endogeneity ρ . For small and moderate values of γ the SE of OIV is much higher than that of the proposed estimator (in many cases by more than two orders of magnitude). That is, OIV is highly affected by the weak nonparametric identification. Somewhat unexpectedly, there are cases where the SE of OIV increases as the sample size increases. These cases correspond to weak identification with high endogeneity. In spite of the high sensitivity of OIV's root-mean square error, inference based on OIV is relatively stable. This robustness seems to be due to the simplified expression of OIV's asymptotic variance inverse, given by $E[\sigma^{-2}(Z)m(Z)m(Z)']$, which is bounded under weak identification. In contrast, IIV confidence intervals involve estimating an asymptotic variance that is less stable under weak identification and high levels of endogeneity. For other situations, including weak identification with moderate levels of endogeneity, IIV confidence intervals seem to perform quite well.

In sum, these results show an omnibus good performance of the IIV estimator, comparing favorably with the OIV estimator in terms of bias and root-mean square error, specially when identification is weak or moderate. These simulations also uncover a relative robustness of OIV's confidence intervals, which to the best of our knowledge has been unnoticed before.

TABLE 1

BIAS, STANDARD ERROR, ROOT MEAN SQUARE ERROR AND 95% COVERAGE.

n	γ	ρ	BIAS		SE		RMS		CP	
			<i>OIV</i>	<i>IIV</i>	<i>OIV</i>	<i>IIV</i>	<i>OIV</i>	<i>IIV</i>	<i>OIV</i>	<i>IIV</i>
50	0.1	0.4	14.63	0.31	1069.83	0.80	1069.93	0.86	0.89	0.95
		0.8	0.72	0.64	50.97	0.60	50.98	0.88	0.78	0.62
	0.25	0.4	0.45	0.11	26.03	0.61	26.03	0.63	0.91	0.95
		0.8	-0.64	0.21	68.84	0.57	68.85	0.61	0.82	0.80
	0.5	0.4	-0.89	-0.00	36.73	0.33	36.74	0.33	0.93	0.95
		0.8	-0.35	-0.01	14.27	0.36	14.27	0.36	0.89	0.90
200	0.1	0.4	-0.35	0.18	33.29	0.65	33.29	0.67	0.91	0.95
		0.8	3.27	0.36	216.58	0.61	216.61	0.72	0.83	0.74
	0.25	0.4	-0.21	-0.00	19.16	0.33	19.16	0.33	0.93	0.95
		0.8	-0.22	-0.00	18.61	0.35	18.61	0.35	0.89	0.90
	0.5	0.4	-0.02	0.00	0.30	0.15	0.30	0.15	0.94	0.95
		0.8	-0.06	-0.00	0.33	0.16	0.33	0.16	0.92	0.93
500	0.1	0.4	-1.47	0.05	67.30	0.50	67.31	0.51	0.92	0.96
		0.8	-3.12	0.10	343.61	0.50	343.63	0.51	0.87	0.86
	0.5	0.4	-0.08	-0.00	1.53	0.19	1.54	0.19	0.94	0.95
		0.8	0.35	-0.00	36.63	0.20	36.63	0.20	0.92	0.93
	1	0.4	-0.00	0.00	0.13	0.09	0.13	0.09	0.94	0.94
		0.8	-0.02	0.00	0.12	0.09	0.12	0.09	0.94	0.94

Note: γ measures the level of Nonparametric identification; ρ measures endogeneity.

In the second experiment we consider a regression model with an endogenous binary variable. The same model was investigated in Newey (1990). We generate data Y , X and Z according to the model

$$DGP2 : \begin{cases} Y = \theta_{01} + X\theta_{02} + \varepsilon \\ X = 1(\alpha_{01} + \alpha_{02}Z + v > 0) \end{cases}, \quad \begin{pmatrix} \varepsilon \\ v \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

with $Z \sim N(0, 1)$ independent of ε and v . The true parameters are $\theta_{01} = \theta_{02} = 1$. The parameter α_{01} is fixed at $\alpha_{01} = 1$, but different values for the parameters (α_{02}, ρ) are considered. Here, the parameters $\theta_0 = (\theta_{01}, \theta_{02})'$ are nonparametrically identified if the “propensity score” $p(z) = P(X = 1 | Z = z)$ depends on z . Since $p(z) = \Phi(\alpha_{01} + \alpha_{02}z)$, nonparametric identification (i.e. Assumption LC) holds provided $\alpha_{02} \neq 0$. Thus, $|\alpha_{02}|$ measures the level of nonparametric identification. The correlation parameter ρ measures the level of endogeneity. The parameter values considered are $\alpha_{02} \in \{1, 0.5, 0.1\}$ and $\rho \in \{0.2, 0.8\}$.

We compare the performance of the IIV estimator and the optimal IV estimator computed with a series estimator for the optimal instrument. I follow Newey (1990) and estimate $p(z)$ by $\hat{p}(z) = A^q(z)'\gamma_n$, where $A^q(z) = (A_1(z), \dots, A_k(z))$, $A_j(z) = z^{j-1}$, $j = 1, \dots, q$, $\gamma_n = [S_n'S_n]^{-1}S_n'\mathbf{Y}$, and $\sigma^2(z)$ by $\hat{\sigma}^2(z) = A^q(z)'\delta_n$, where $\delta_n = [S_n'S_n]^{-1}S_n'\mathbf{e}_2$, $\mathbf{e}_2 = (\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2)'$, $\hat{\varepsilon}_i = Y_i - X_i'\hat{\theta}_{OIV}$ and

$$S_n := \begin{bmatrix} A^q(Z_1)' \\ \vdots \\ A^q(Z_n)' \end{bmatrix}.$$

We report results with $q = 7$, but other values of q led to similar conclusions. Similarly, other choices of basis functions gave comparable results. These unreported simulations can be obtained from the author upon request.

Table 2 reports results for slope estimates of θ_{02} for sample sizes of $n = 50, 200$ and 500 . For all parameter values considered the IIV estimator for θ_{02} presents a smaller RMS than optimal IV. It is remarkable the robustness of the new estimator to weak nonparametric identification when $\alpha_{02} = 0.1$. The RMS of OIV is very sensitive to weak identification and the level of endogeneity. Inference based on OIV is robust and in general more accurate than that based on IIV.

TABLE 2

BIAS, STANDARD ERROR, ROOT MEAN SQUARE ERROR AND 95% COVERAGE.

n	α_{02}	ρ	BIAS		SE		RMS		CP	
			<i>OIV</i>	<i>IIV</i>	<i>OIV</i>	<i>IIV</i>	<i>OIV</i>	<i>IIV</i>	<i>OIV</i>	<i>IIV</i>
50	1.0	0.2	-2.03	0.98	134.98	0.71	135.00	1.22	0.96	0.96
		0.8	1.97	0.97	38.18	0.71	38.23	1.21	0.97	0.95
	0.5	0.2	0.90	0.95	109.10	1.48	109.11	1.76	0.97	0.99
		0.8	13.38	1.17	244.20	1.51	244.57	1.91	0.96	0.92
	0.1	0.2	5.47	1.38	116.21	2.44	116.34	2.81	0.95	0.99
		0.8	6.48	2.34	60.76	2.02	61.10	3.10	0.97	0.90
200	1.0	0.2	-0.07	1.00	58.55	0.33	58.55	1.06	0.96	0.94
		0.8	1.59	0.98	50.27	0.33	50.30	1.04	0.97	0.95
	0.5	0.2	16.02	0.97	400.52	0.68	400.84	1.19	0.96	0.98
		0.8	2.48	0.98	38.49	0.68	38.57	1.19	0.95	0.94
	0.1	0.2	-0.20	1.42	97.18	2.01	97.18	2.47	0.95	0.99
		0.8	-3.26	1.95	197.59	1.98	197.62	2.79	0.95	0.92
500	1.0	0.2	2.78	1.01	50.46	0.20	50.54	1.03	0.95	0.94
		0.8	-3.42	0.98	179.59	0.19	179.62	1.00	0.95	0.96
	0.5	0.2	0.11	0.97	42.17	0.43	42.17	1.06	0.93	0.95
		0.8	-1.20	0.99	189.95	0.40	189.95	1.06	0.95	0.94
	0.1	0.2	2.07	1.11	59.73	1.84	59.77	2.15	0.94	0.99
		0.8	0.32	1.54	59.55	1.88	59.55	2.43	0.95	0.92

Note: α_{02} measures the level of Nonparametric identification; ρ measures endogeneity.

To summarize, the new estimator performs quite well in finite samples of small and moderate size uniformly over different levels of nonparametric identification and endogeneity. In contrast, OIV has a bias and root-mean square error that is rather sensitive to weak identification and endogeneity. For confidence intervals based on asymptotic normality of estimates, we have uncovered a somewhat surprising robustness of OIV. Confidence intervals based on IIV may be sensitive to weak identification and high levels of endogeneity, but they behave well otherwise. These simulations suggest that the new estimator can be a sensible and simple complement to optimal IV procedures when the researcher is concerned about weak or moderate identification. Next section shows the utility of the IIV estimator in an economic application of interest. In this application Optimal IV is not feasible given the high dimensionality of Z and the small sample size.

5 An Application to Estimating the Elasticity of Intertemporal Substitution

In its log-linearized version, the Consumption-based Capital Asset Pricing Model (CCAPM) studied in Hansen and Singleton (1982) leads to the conditional moment restriction

$$E_P[\Delta c_{t+1} - \alpha - \psi r_{t+1} \mid Z_t] = 0 \quad (14)$$

where ψ is the elasticity of intertemporal substitution (EIS), Δc_{t+1} is the growth rate of consumption, r_{t+1} is the log gross return on some asset, α is a constant and Z_t is a vector of variables in the agent's information set at time t . The parameters $\theta_0 = (\alpha, \psi)'$ can be estimated from (14) by several estimation strategies; see Hansen and Singleton (1983) and Hall (1988). One is a TSLS estimator with Δc_{t+1} as the dependent variable; another is to apply TSLS with r_{t+1} as the dependent variable; a third one is to use a method that is invariant to the normalization, such as Limited-Information Maximum Likelihood (LIML). Under strong IV identification, these methods should be asymptotically equivalent, so it should not matter which method to use. In practice, it has been shown that it greatly matters, which provides indirect evidence of weak instruments; see Stock and Wright (2000) and Neely et al. (2001). This empirical evidence has been extended to international data in Campbell (2003). The weak instruments problem may explain the apparently contradictory results that estimates of ψ and $1/\psi$ are simultaneously small and not statistically significant.

We revisit this issue applying the new IIV estimator to an international data set considered in Campbell (2003) and Yogo (2004).⁵ The data consists of quarterly observations on equity markets at an aggregate level and macroeconomic variables for eleven countries: Australia (AUL), Canada (CAN), France (FR), Germany (GER), Italy (ITA), Japan (JAP), Netherlands (NTH), Sweden (SWD), Switzerland (SWR), the United Kingdom (UK) and the United States (USA). The primary sources of international data are Morgan Stanley Capital International and the International Financial Statistics of the International Monetary Fund. The sample periods vary by country, see Table 4. A full description of the data is given in Campbell (2003).⁶

The asset returns used are the real interest rate, denoted by $r_{f,t}$, and the real aggregate stock return, denoted by $r_{e,t}$. The real stock return is constructed as log of the gross stock

⁵The data set is available at Motohiro Yogo's web page. I thank Motohiro for making the data available.

⁶This application involves time series. It is straightforward to extend the uniform consistency results to time series—one needs to replace the law of large numbers for *iid* data by the Ergodic Theorem. Extensions are also possible for uniform asymptotic normality under martingale difference errors, which arise naturally in applications such as the present Euler equation. In these extensions P is the (stationary) marginal distribution of W_t .

return deflated by the consumer price index. The real interest rate is constructed in the same way, using an available proxy for the short-term interest rate. Real consumption growth, denoted by Δc_t , is the first difference in log real consumption per capita. Following Yogo (2004), we use as instruments $Z_t = (r_{t-1}, \pi_{t-1}, \Delta c_{t-1}, dp_{t-1})$, where r_t is the nominal interest rate, π_t is inflation, and dp_t is the log dividend-price ratio.

Yogo (2004) provided formal evidence of weak instruments for these data sets and instruments. In particular, he showed that quite often the first-stage F-test was less than 5 in these regressions, see Table 4 here, and he used inference methods (confidence intervals and test statistics) that were robust to identification failure. His main conclusion was that the EIS is small and not significant across the eleven countries considered.

The weak identification of IV comes here at not surprise, as is well-known that both consumption growth and asset returns are notoriously difficult to be linearly predicted. However, as this paper emphasizes, it is possible that these variables are linearly unpredictable but nonlinearly predictable using the set of instruments Z_t ; see Guidolin et al. (2009) and references therein for extensive empirical evidence of this nonlinearity. We first investigate empirically the possibility of nonparametric identification in the next section.

5.1 Testing for the Linear Completeness Condition

Proposition 1 implies that $\theta_0 = (\alpha, \psi)'$ is identified in the regression (14) iff

$$\lambda_1 E_P[r_{t+1} | Z_t] + \lambda_2 = 0 \implies \lambda_1 = 0, \lambda_2 = 0.$$

This is equivalent to Z_t being nonparametrically significant in $E_P[r_{t+1} | Z_t]$, i.e.

$$P(E_P[r_{t+1} | Z_t] \neq E_P[r_{t+1}]) > 0. \tag{15}$$

Likewise, if $\psi \neq 0$, then $1/\psi$ is identified from

$$E_P[r_{t+1} - \beta - (1/\psi)\Delta c_{t+1} | Z_t] = 0 \tag{16}$$

provided

$$P(E_P[\Delta c_{t+1} | Z_t] \neq E_P[\Delta c_{t+1}]) > 0. \tag{17}$$

There is an extensive literature in econometrics on nonparametric tests for hypothesis such as (15) and (17). See, e.g., Bierens (1982), Stinchcombe and White (1998) and Escanciano and Velasco (2006). Note that the validity of these nonparametric tests are pointwise in nature. Here, I follow Bierens (1982) and use a Cramer-von Mises (CvM) test for the null hypothesis

$H_0 : P(E_P[Y_{t+1} | Z_t] = E_P[Y_{t+1}]) = 1$ against $H_1 : P(E_P[Y_{t+1} | Z_t] \neq E_P[Y_{t+1}]) > 0$, where Y_{t+1} denotes either Δc_{t+1} , $r_{f,t+1}$ or $r_{e,t+1}$, thus testing the linear completeness conditions in (15) and (17). The test statistics are quadratic forms in the residuals $\hat{u} = (\hat{u}_1, \dots, \hat{u}_n)'$, with $\hat{u}_t = (Y_{t+1} - \bar{Y}_n)$ and $\bar{Y}_n = n^{-1} \sum_{t=1}^n Y_{t+1}$, computed as

$$CvM_n = \frac{\hat{u}'\Omega\hat{u}}{n\hat{\sigma}_u^2},$$

where Ω is defined after (3) and $\hat{\sigma}_u^2 = n^{-1} \sum_{t=1}^n \hat{u}_t^2$. We standardize the components of Z_t by their sample standard deviation, so tests become scale invariant. Note that one may interpret CvM_n as a nonparametric extension of the classical first-stage F -test in IV regression.⁷

The asymptotic distribution of CvM_n is not pivotal, but its critical values can be approximated by a wild-bootstrap procedure; see Dominguez and Lobato (2003). That is, one can approximate the asymptotic distribution of CvM_n by that of

$$CvM_n^* = \frac{\hat{u}^*\Omega\hat{u}^*}{n\hat{\sigma}_u^{*2}},$$

where $\hat{u}^* = (\hat{u}_1^*, \dots, \hat{u}_n^*)'$, $\hat{u}_t^* = (V_t\hat{u}_t - c_n)$, $c_n = n^{-1} \sum_{t=1}^n V_t\hat{u}_t$, $\hat{\sigma}_u^{*2} = n^{-1} \sum_{t=1}^n \hat{u}_t^{*2}$, and where $\{V_t\}_{t=1}^n$ is a sequence of independent and identically distributed (*iid*) random variables with zero mean, unit variance, bounded support and also independent of the sequence $\{Y_t, X_t, Z_t\}_{t=1}^{n+1}$. Commonly used examples of $\{V_t\}$ sequences are *iid* Bernoulli variables with

$$P(V_t = 0.5(1 - \sqrt{5})) = b \quad P(V_t = 0.5(1 + \sqrt{5})) = 1 - b, \quad (18)$$

where $b = (1 + \sqrt{5})/2\sqrt{5}$, or Rademacher sequences $P(V_t = 1) = 0.5$ and $P(V_t = -1) = 0.5$. The critical values of CvM_n^* are approximated by Monte Carlo simulations. Thus, the null hypothesis of lack of linear completeness will be rejected at the $100\alpha\%$ of significance when $CvM_n \geq c_{n,\alpha}^*$, where $c_{n,\alpha}^*$ is the $(1 - \alpha)$ -th empirical quantile of B realizations of CvM_n^* . Alternatively, I can use the bootstrap p -values, p_n^* say, rejecting H_0 when $p_n^* < \alpha$, where $p_n^* = \Pr(CvM_n^* \geq CvM_n | \{Y_t, X_t, Z_t\}_{t=1}^{n+1})$.

Table 3 reports the bootstrap p -values for the data sets considered and the three variables Δc_{t+1} , $r_{f,t+1}$ and $r_{e,t+1}$. The number of bootstrap replications is $B = 5000$. For completeness, we also report the values of the first-stage F test from Yogo (2004). This is useful in order to compare parametric and nonparametric identification failures. The F tests suggest low linear predictability for *all* countries in consumption growth and stock returns, but high

⁷The F -test is computed as $F = (\hat{u}'P_{Z,n}\hat{u})/n\hat{\sigma}_F^2$, where $P_{Z,n} = Z(Z'Z/n)^{-1}Z'$, $\hat{\sigma}_F^2 = 4SSR_u/(n - 5)$ and SSR_u is the sum of unrestricted squared residuals.

predictability of interest rates. When looking for nonlinear (nonparametric) predictability, we find that interest rates are highly predictable, stock returns are not predictable, and consumption growth shows low or no predictability for all countries but for UK and USA. Hence, two conclusions arise from this identification analysis: (i) the evidence of weak linear identification in international data found in Yogo (2004) and others can be extended to weak *nonparametric* identification when considering stock returns, and (ii) when using interest rates, it appears that ψ and $1/\psi$ are nonparametrically identified for UK and USA data, but $1/\psi$ is weakly identified by traditional IV methods. Thus, the UK and USA data highlight the difference between linear and nonlinear predictability, and hence the difference between lack of identification by linear methods and the linear completeness assumption. In view of this evidence, in what follows we restrict attention to these two countries and interest rates data. It is interesting to note that these findings agree with the extensive empirical evidence in Guidolin et al. (2009). These authors investigate nonlinear predictability in G7 countries and conclude that: “US and UK data return data appear to be “special,” in the sense that good predictive performance demands the estimation of non-linear models”.

TABLE 3

LINEAR COMPLETENESS TESTS. BOOTSTRAP P-VALUES				
Country	Sample Period	$Y_{t+1} = \Delta c_{t+1}$	$Y_{t+1} = r_{f,t+1}$	$Y_{t+1} = r_{e,t+1}$
AUL	1970.3-1998.4	0.35 (1.79)	0.00 (21.81)	0.34 (1.82)
CAN	1970.3-1999.1	0.44 (3.03)	0.00 (15.37)	0.60 (2.51)
FR	1970.3-1998.3	0.90 (0.17)	0.00 (38.43)	0.61 (3.09)
GER	1979.1-1998.3	0.54 (0.83)	0.00 (17.66)	0.82 (0.69)
ITA	1971.4-1998.1	0.67 (0.73)	0.00 (19.01)	0.81 (1.10)
JAP	1970.3-1998.4	0.50 (1.18)	0.00 (8.64)	0.34 (3.49)
NTH	1977.3-1998.2	0.59 (0.89)	0.00 (12.05)	0.94 (0.73)
SWD	1970.3-1999.2	0.62 (0.48)	0.00 (17.08)	0.15 (2.24)
SWT	1976.2-1998.4	0.83 (0.97)	0.00 (8.55)	0.97 (0.11)
UK	1970.3-1999.1	0.07 (2.52)	0.00 (17.04)	0.60 (2.62)
USA	1947.3-1998.4	0.00 (2.93)	0.00 (15.53)	0.19 (2.88)

Note: First stage F-statistics in parenthesis. Bootstrap replications $B = 5000$.

5.2 Uniformly Consistent Estimation and Inference

We now proceed to estimate ψ in (14) and $1/\psi$ in (16) using the IIV estimator and interest rates and consumption growth for UK and USA data. We compare in Table 4 the IIV

estimator with Yogo's TSLS and LIML estimates. Standard errors, computed under the assumption of martingale difference errors, are provided in parenthesis. We also computed HAC standard errors with automatic lag-length choice, but results were very similar, so they are not reported. The first fact to note is that the three estimators give very different results, which may be considered as evidence of identification problems. The IIV estimates for ψ are considerably larger than those of the TSLS and LIML estimates. The case of UK is particularly illuminating and highlights the benefits of the present approach. This is an example where there is weak identification by IV methods, but the parameters seem to be nonparametrically identified. The IIV estimate of the EIS for UK data is 0.5, and is significantly different from zero. The TSLS estimate is much smaller (0.17) and not significant at 5%. In contrast to the puzzling results with TSLS, IIV's estimation of $1/\psi$ is not contradictory, with an estimate of 1.94 that is fairly consistent with the IIV estimate of ψ . For USA data, we obtain an estimate of 0.66, that although much larger than the TSLS and LIML estimates, it is not significantly different from zero.

TABLE 4

ESTIMATES: INTEREST RATES AND CONSUMPTION REGRESSIONS							
Country	Sample Period	ψ			$1/\psi$		
		IIV	TSLS	LIML	IIV	TSLS	LIML
UK	1970.3-1999.1	0.50	0.17	0.16	1.94	1.06	6.21
		(0.20)	(0.13)	(0.13)	(0.8)	(0.45)	(5.17)
USA	1947.3-1998.4	0.66	0.06	0.03	1.41	0.68	34.11
		(0.49)	(0.09)	(0.10)	(1.02)	(0.48)	(112.50)

In Table 5, 95% asymptotic confidence intervals for the EIS based on the IIV estimates are compared with those from other methods that are robust to weak identification. More concretely, we compare with confidence intervals computed inverting the Anderson-Rubin's test (AR, see Anderson and Rubin, 1949) and the conditional likelihood ratio (CLR) test of Moreira (2003). An excellent description of these procedures can be found in Yogo (2004). We first note that the confidence intervals are different across the different methods. This can be partly accounted for by the information provided by the nonlinearity in the data. For UK data, IIV provides confidence intervals consistent with larger values for the EIS than those provided by IV robust methods. In particular, according to the CLR the EIS is not significantly different from zero. IIV confidence intervals suggest that for USA data the EIS is not significantly different from zero at 5%.

TABLE 5

CONFIDENCE INTERVALS. INTEREST RATES				
Country	Sample Period	ψ		
		IIV	AR	CLR
UK	1970.3-1999.1	[0.10, 0.90]	[0.04, 0.28]	[-0.12, 0.43]
USA	1947.3-1998.4	[-0.31, 1.63]	\emptyset	[-0.19, 0.22]

5.3 Nonparametric Tests for Overidentifying Restrictions

The proposed IIV estimator is based on the strict exogeneity assumption in (14). Much as in the classical linear case, one can test for (nonparametric) overidentifying restrictions. This robustness checks provide some empirical validity to the previous results. To that end, we extend the nonparametric tests of Dominguez and Lobato (2003) to a linear regression setting (these authors consider the case of no covariates in the regression). The test statistics are of CvM type for testing $H_0 : P(E_P[\Delta c_{t+1} - \alpha - \psi r_{f,t+1} | Z_t] = 0) = 1$, against the general nonparametric alternative $H_1 : P(E_P[\Delta c_{t+1} - \alpha - \psi r_{f,t+1} | Z_t] \neq 0) > 0$. The CvM test is computed as a quadratic form in the residuals $\hat{u} = (\hat{u}_1, \dots, \hat{u}_n)'$, with $\hat{u}_t = \Delta c_{t+1} - \hat{\alpha} - \hat{\psi} r_{f,t+1}$, $\hat{\theta}_{IIV} = (\hat{\alpha}, \hat{\psi})'$, simply as

$$CvM_n = \frac{1}{\hat{\sigma}_u^2 n^2} \hat{u}' \Omega'_{ind} \Omega_{ind} \hat{u},$$

where $\Omega_{ind} = H\Psi$, $H = I_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, I_n is the $n \times n$ identity matrix, and Ψ is the $n \times n$ matrix with elements $1(Z_t \leq Z_s)$. The asymptotic distribution of CvM_n is approximated by that of the bootstrap analogue $CvM_n^* = n^{-2} \hat{u}^* \Omega'_{ind} \Omega_{ind} \hat{u}^*$, where $\hat{u}^* = (V_1 \hat{u}_1, \dots, V_n \hat{u}_n)'$, and $\{V_t\}_{t=1}^n$ are generated from (18). This bootstrap can be justified along the lines of Dominguez and Lobato (2003). These tests require consistent estimation of the parameter θ_0 , which is supported by the empirical evidence above. The conclusion from applying these consistent strict exogeneity tests is that the strict exogeneity assumption is not rejected for both data sets. The bootstrap p-values are 0.252 and 0.282 for UK and USA, respectively. Hence, these tests results validate previous inferences.

To sum, this application has shown that the new IIV estimator can complement existing inferential procedures in applications of economic interest. In addition, we have shown that is possible to test for the linear completeness condition by means of consistent tests of conditional moment restrictions. The application to UK data provides an example where the EIS is weakly identified by linear methods, but strongly identified with the nonparametric methods of this paper.

6 Conclusions and Extensions

In this paper, we have emphasized minimal conditions for nonparametric identification of linear regression models when a strict exogeneity assumption holds. We have proposed an estimator that is uniformly consistent under nearly the minimal identifying condition. The new estimator should be appealing to practitioners given its simplicity and robustness in finite samples.

Optimal IV estimators, optimal in the sense of minimum asymptotic variance for regular estimators, also identify under the minimal identification assumption in linear models, but as shown in the simulations, their bias and root-mean square error are very sensitive to the level of nonparametric identification. The IIV estimator seems to outperform the optimal IV estimator for moderate and small sample sizes when nonparametric identification is weak. These results suggest that classical definitions of optimality in semiparametric models need to be challenged when optimal procedures are not uniformly (globally) valid. Confidence intervals based on OIV, however, seem to be quite robust. This interesting finding deserves further investigation.

It will be also interesting to compare our IIV estimator with other estimates such as the Fuller- k class; see Section 6 in Stock et al. (2002) for a survey. Note that, unlike the new estimator, existing robust procedures may not identify the parameters under nearly minimal conditions. Whether these theoretical results translate into finite sample performance will be investigated in future work.

Another interesting extension is to testing for the linear completeness condition in linear models with multivariate regressors. The negation of Assumption LC leads to a conditional moment restriction, which suggests that this assumption could be tested using existing consistent tests of conditional moment restrictions. There is, however, no guarantee that parameters in the resulting conditional moment restriction are identified, which invalidates the application of standard consistent tests. This issue is an interesting topic for future research.

More broadly, this paper has emphasized uniform consistent estimation under nearly minimal identifying conditions. Minimum distance estimators, as discussed by Wolfowitz (1957) and others, are good candidates for uniform consistent estimators. Developing such uniform procedures in general econometric models deserves further investigation.

7 Appendix: Mathematical Proofs

Proof of Proposition 2: We prove that the negation of Assumption LC is equivalent to the negation of (8). That is, the negation of Assumption LC implies that there exists a non-zero vector, say $\lambda \in \mathbb{R}^p$, such that

$$E_P[\lambda'X \mid Z] = 0 \text{ a.s.} \quad (19)$$

Now, by the law of iterated expectations and Theorem 1 in Bierens (1982), (19) is equivalent to

$$\lambda'h_{X,P}(u) = E_P[E_P[\lambda'X \mid Z] \exp(iu'V^{-1/2}Z)] = 0 \text{ almost everywhere (a.e.)}$$

Hence, $\lambda'E_U[h_{X,P}(U)h_{X,P}^c(U)]\lambda = 0$. \square

Proof of Theorem 1: Define the class of functions

$$\mathcal{F}_X = \{w = (y, x, z) \rightarrow x \exp(iu'z), u \in \Pi\},$$

where Π is a compact, convex subset of \mathbb{R}^k with nonempty interior. Note that $|x \exp(iu'z)| \leq |x|$ and therefore $F_X(w) = |x|$ is an envelop for the class \mathcal{F}_X . We use a number of results from van der Vaart and Wellner (1996), which is hereinafter denoted by VW.

Fix an arbitrary $\delta > 0$, and choose a compact and convex set $\Pi \equiv \Pi_\delta$ such that

$$\left| \int_{\mathbb{R}^k \setminus \Pi} \phi(u) du \right| \leq \delta. \quad (20)$$

Then, write

$$\begin{aligned} \langle h_{x_j, P_n}, h_{x_k, P_n} \rangle &= \int_{\mathbb{R}^k} h_{x_j, P_n}(u) h_{x_k, P_n}^c(u) \phi(u) du. \\ &= \int_{\Pi} h_{x_j, P_n}(u) h_{x_k, P_n}^c(u) \phi(u) du + \int_{\mathbb{R}^k \setminus \Pi} h_{x_j, P_n}(u) h_{x_k, P_n}^c(u) \phi(u) du. \\ &=: I_{1n, P} + I_{2n, P}. \end{aligned} \quad (21)$$

We deal with $I_{1n, P}$ by showing that the class \mathcal{F}_X is Glivenko-Cantelli uniformly in $P \in \mathcal{P}$, as defined in VW (p. 167). To that end, we apply Theorem 2.8.1 of that reference. First,

note that the envelop satisfies

$$\lim_{M \rightarrow \infty} \sup_{P \in \mathcal{P}} E_P [|X| 1(|X| > M)] = 0,$$

by the uniformly bounded second moment of X on \mathcal{P} .

Using standard empirical processes results, see e.g. Lemma 2.13 in Pakes and Pollard (1989), we conclude

$$\sup_Q \log N \left(\varepsilon \|F_X\|_{Q,1}, \mathcal{F}_X, \|\cdot\|_{Q,1} \right) = o(n),$$

where the supremum in Q is over the set of all discrete probability measures with atoms of size integer multiples of $1/n$, $N \left(\varepsilon, \mathcal{F}_X, \|\cdot\|_{Q,1} \right)$ is the covering number defined in VW (p. 83), and $\|\cdot\|_{Q,1}$ is the L_1 norm $\|F\|_{Q,1} := \int |F| dQ$. Then, by VW (Theorem 2.8.1), \mathcal{F}_X is Glivenko-Cantelli uniformly in $P \in \mathcal{P}$. Note that these arguments can be applied to any compact Π , in particular to $\Pi_V = \{u'V^{-1/2} : u \in \Pi\}$. Therefore, uniformly in $u \in \Pi$ and $P \in \mathcal{P}$,

$$h_{X,P_n}(u) = n^{-1} \sum_{s=1}^n X_s \exp(iu'\hat{V}^{-1/2}Z_s) = h_{X,P}(u) + o_P(1),$$

where I have used that, uniformly in $P \in \mathcal{P}$, $\hat{V} = V + o_P(1)$ and hence, $u'\hat{V}^{-1/2} \in \Pi_V$ with probability approaching one, and by continuity

$$E_P[Y \exp(iu'\hat{V}^{-1/2}Z)] = E_P[Y \exp(iu'V^{-1/2}Z)] + o_P(1).$$

Thus, by the continuous mapping theorem

$$I_{1n,P} = \int_{\Pi} h_{X_j,P}(u) h_{X_k,P}^c(u) \phi(u) du + o_P(1) \text{ uniformly in } P \in \mathcal{P}.$$

On the other hand, it is straightforward to show that

$$I_{2n,P} = O_P(\delta) \text{ uniformly in } P \in \mathcal{P},$$

by (20) and the inequality $|h_{X_j,P_n}(u) h_{X_k,P_n}^c(u)| \leq |E_{P_n}[F_X]|^2 = O_P(1)$ uniformly in $P \in \mathcal{P}$. Since $\delta > 0$ was arbitrary, conclude from (21) that for all $j, k = 1, \dots, p$,

$$\langle h_{X_j,P_n}, h_{X_k,P_n} \rangle = \langle h_{X_j,P}, h_{X_k,P} \rangle + o_P(1) \text{ uniformly in } P \in \mathcal{P}. \quad (22)$$

The proof of

$$\langle h_{X_j, P_n}, h_{Y, P_n} \rangle = \langle h_{X_j, P}, h_{Y, P} \rangle + o_P(1) \text{ uniformly in } P \in \mathcal{P},$$

follows the same arguments, and hence is omitted. \square

Proof of Theorem 2: The arguments of the proof of Theorem 1 yield that $h_{X, P_n} \xrightarrow{L_2} h_{X, P}$ uniformly in $P \in \mathcal{P}$, and therefore, uniformly in $P \in \mathcal{P}^{AN} \subset \mathcal{P}$. On the other hand, by Chebyshev's inequality, for any $M > 0$,

$$\begin{aligned} \sup_{P \in \mathcal{P}^{AN}} P(\|\sqrt{n}h_{\varepsilon, P_n}\| > M) &\leq M^{-2} \sup_{P \in \mathcal{P}^{AN}} E_P \left[\|\sqrt{n}h_{\varepsilon, P_n}\|^2 \right] \\ &\leq M^{-2} \sup_{P \in \mathcal{P}^{AN}} E_P [\varepsilon^2] \\ &\leq M^{-2} C, \end{aligned}$$

where the second inequality uses the *iid* and strict exogeneity assumption. Thus, $\sqrt{n}h_{\varepsilon, P_n} = O_P(1)$ uniformly in $P \in \mathcal{P}^{AN}$ in $L_2(\phi)$. These asymptotic results, $\hat{V} = V + o_P(1)$, and the continuity of the inner product yield, uniformly in $P \in \mathcal{P}$, the expansion

$$\begin{aligned} \sqrt{n}(\hat{\theta}_{IIV} - \theta_0) &= \langle h_{X, P}, h_{X, P} \rangle^{-1} \sqrt{n} \langle h_{X, P}, h_{\varepsilon, P_n} \rangle + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \langle h_{X, P}, h_{X, P} \rangle^{-1} H_{X, P}(Z_i) \varepsilon_i + o_P(1). \end{aligned} \quad (23)$$

Note that by $\lambda_{\min}(E_P[m_P(Z)m'_P(Z)]) \geq C$, bounded moment of X and definition of $H_{X, P}$, there exists a constant C such that

$$|\langle h_{X, P}, h_{X, P} \rangle^{-1} H_{X, P}(Z)| \leq C.$$

Then, the $2 + \delta$ finite moment of ε implies that the uniform CLT of Petrov (1975, p. 118) is applicable, and the uniform asymptotic normality of $\sqrt{n}(\hat{\theta}_{IIV} - \theta_0)$ follows from (23). \square

Proof of Corollary 1: We show that

$$\hat{W}_P := n(\hat{\theta}_{IIV} - \theta)' \hat{\Gamma}^{-1} (\hat{\theta}_{IIV} - \theta)$$

converges uniformly in distribution to a χ_p^2 , uniformly in \mathcal{P}^{INF} . By Theorem 2 and standard

arguments, it suffices to show that

$$\widehat{\Gamma} = \Gamma + o_P(1)$$

uniformly in \mathcal{P}^{INF} and that Γ has its eigenvalues bounded away from zero (uniformly in \mathcal{P}^{INF}). Note by (22) we only need to show that

$$\widehat{\Lambda} = \Lambda + o_P(1),$$

uniformly in \mathcal{P}^{INF} . To that end, write

$$\begin{aligned} \widehat{\Lambda} &= \frac{1}{n} \sum_{i=1}^n \widehat{\varepsilon}_i^2 \widehat{H}_{X,P}(Z_i) \widehat{H}'_{X,P}(Z_i), \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \widehat{H}_{X,P}(Z_i) \widehat{H}'_{X,P}(Z_i) + \frac{1}{n} \sum_{i=1}^n (\widehat{\varepsilon}_i^2 - \varepsilon_i^2) \widehat{H}_{X,P}(Z_i) \widehat{H}'_{X,P}(Z_i) \\ &\equiv A_1 + A_2, \end{aligned}$$

where

$$\widehat{H}_{X,P}(u) := \frac{1}{n} \sum_{j=1}^n X_j \exp(-0.5 (u - Z_j)' \widehat{V}^{-1} (u - Z_j)).$$

Applying Theorem 2.8.1 in VW we conclude that for any compact set Π the class

$$\{(x, z) \rightarrow x \exp(-0.5 (u - z)' V^{-1} (u - z)), u \in \Pi\}$$

is Glivenko-Cantelli uniformly in $P \in \mathcal{P}^{INF}$. Thus, by simple algebra and the above uniform law of large numbers we can show, uniformly in $P \in \mathcal{P}^{INF}$,

$$\begin{aligned} A_1 &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 H_{X,P}(Z_i) H'_{X,P}(Z_i) + o_P(1) \\ &= \Lambda + o_P(1), \end{aligned}$$

where the last equality follows from a uniform law of large numbers after showing that

$$\begin{aligned} &\sup_{P \in \mathcal{P}^{INF}} E_P \left[|\varepsilon_i H_{X,P}(Z_i)|^{2+\delta} \right] \\ &\leq C \sup_{P \in \mathcal{P}^{INF}} (E_P[|X|])^{2+\delta} \sup_{P \in \mathcal{P}^{INF}} E_P \left[|\varepsilon_i|^{2+\delta} \right] \\ &< \infty, \end{aligned}$$

since $|H_{X,P}(u)| \leq E_P[|X|]$.

Similarly, straightforward algebra and the previous arguments show that $A_2 = o_P(1)$, uniformly in $P \in \mathcal{P}^{INF}$. The Corollary then follows from Lemma 2 in Kasy (2015). \square

References

- [1] AMEMIYA, T. (1974): “The Non-linear Two-Stage Least Squares Estimator,” *Journal of Econometrics*, 2, 105-110.
- [2] AMEMIYA, T. (1977): “The Maximum Likelihood and the Non-linear Three-Stage Least Squares Estimator in the General Non-linear Simultaneous Equation Model,” *Econometrica*, 45: 955-968.
- [3] ANDERSON, T.W. AND RUBIN, H. (1949): “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations,” *Annals of Mathematical Statistics*, 20, 46-63.
- [4] ANDREWS, D.W.K AND GUGGENBERGER, P. (2009): “Validity of Subsampling and ‘Plug-in Asymptotic’ Inference for Parameters Defined by Moment Inequalities,” *Econometric Theory*, 25(3): 669-709.
- [5] ANDREWS, D.W.K. AND STOCK, J. (2005): “Inference with Weak Instruments,” invited survey note for the 2005 World congress of the Econometric Society.
- [6] BAHADUR, R.R AND SAVAGE, L.J. (1956): “The Nonexistence of Certain Statistical Procedures in Nonparametric Problems,” *Annals of Mathematical Statistics*, 27, 1115-1122.
- [7] BELLONI, A., CHERNOZHUKOV, V., FERNANDEZ-VAL, I. AND C. HANSEN (2015): “Program Evaluation with High-Dimensional Data,” arXiv:1311.2645v5.
- [8] BICKEL, P.J., KLAASSEN, C.A., RITOV, Y. AND WELLNER, J.A. (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore.
- [9] BIERENS, H. J. (1982): “Consistent Model Specification Tests,” *Journal of Econometrics*, 20, 105-134.
- [10] CAI, Z., DAS, M., XIONG, H. AND WU, X. (2006): “Functional Coefficient Instrumental Variables Models,” *Journal of Econometrics*, 133, 207-241.
- [11] CAMPBELL, J.Y. (2003): “Consumption-Based Asset Pricing”, Chapter 13 in George Constantinides, Milton Harris, and Rene Stulz eds. *Handbook of the Economics of Finance* Vol. IB, North-Holland, Amsterdam, 803-887.

- [12] CANAY, I.A., SANTOS, A. AND A. M. SHAIKH (2013): “On the Testability of Identification in Some Nonparametric Models with Endogeneity”, *Econometrica*, 81, 2535-2559.
- [13] CARRASCO, M. (2012): “A Regularization Approach to the Many Instruments Problem,” *Journal of Econometrics*, 170, 383-398.
- [14] CHAMBERLAIN, G. (1987): “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions,” *Journal of Econometrics*, 34, 305-334.
- [15] DAS, M. (2005): “Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors,” *Journal of Econometrics*, 124, 335-361.
- [16] DOMÍNGUEZ, M. AND LOBATO, I. N. (2003): “A Consistent Test for the Martingale Difference Hypothesis,” *Econometric Reviews*, 22, 351-377.
- [17] DOMINGUEZ, M.A. AND LOBATO, I.N. (2004): “Consistent Estimation of Models Defined by Conditional Moment Restrictions,” *Econometrica*, 72, 1601-1615.
- [18] DONALD, S.G. AND W.K. NEWEY (2001): “Choosing the Number of Instruments,” *Econometrica*, 69, 1161-1191.
- [19] DUFOUR, J.M. (1997): “Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models,” *Econometrica*, 65, 1365-1387.
- [20] ESCANCIANO, J. C. (2006): “A Consistent Diagnostic Test for Regression Models Using Projections,” *Econometric Theory*, 22, 1030-1051.
- [21] ESCANCIANO, J.C. AND C. VELASCO (2006): “Generalized spectral tests for the martingale difference hypothesis,” *Journal of Econometrics*, 134, 151-185.
- [22] GUIDOLIN, M. HYDE, S. MCMILLAN, D. AND ONO, S. (2009): “Non-linear Predictability in Stock and Bond Returns: When and Where is it Exploitable?,” *International Journal of Forecasting*, 25, 373-399.
- [23] HALL, R. E. (1988): “Intertemporal Substitution in Consumption,” *Journal of Political Economy*, 96, 339-357.
- [24] HANSEN, L. P. AND SINGLETON, K. J. (1982): “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica*, 50, 1269-1286.
- [25] HANSEN, L. P. AND SINGLETON, K. J. (1983): “Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns,” *Journal of Political Economy*, 91, 249-265.

- [26] IMBENS, G. AND MANSKI, C.F (2004): “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72, 1845-1857.
- [27] KASY, M. (2015): “Uniformity and the Delta Method,” unpublished manuscript.
- [28] LECAM, L. AND L. SCHWARTZ (1960): “A Necessary and Sufficient Condition for the Existence of Consistent Estimates,” *Annals of Mathematical Statistics*, 31, 140-150.
- [29] LEEB, H. AND PÖTSCHER, B. M. (2005): “Model selection and inference: Facts and fiction,” *Econometric Theory*, 21, 21–59.
- [30] LINTON, O. (2002): “Edgeworth Approximations for Semiparametric Instrumental Variable Estimators and Test Statistics”, *Journal of Econometrics*, 106, 325-368.
- [31] MANSKI, C.F. (1988): *Analog Estimation Methods in Econometrics*, London: Chapman and Hall.
- [32] MIKUSHEVA, A. (2007): “Uniform Inference in Autoregressive Models,” *Econometrica*, 75,1411-1452.
- [33] MIKUSHEVA, A. (2013): “Survey on Statistical Inferences in Weakly-Identified Instrumental Variable Models,” *Applied Econometrics*, 29, 116-131.
- [34] MOREIRA, M.J. (2003): “A Conditional Likelihood Ratio Test for Structural Models,” *Econometrica*, 71, 1027-1048.
- [35] NEELY, C., ROY, A. AND WHITEMAN, C.H. (2001): “Risk Aversion versus Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption CAPM,” *Journal of Business and Economic Statistics*, 19 395-403.
- [36] NEWEY, W. (1990): “Efficient Instrumental Variables Estimation of Nonlinear Models,” *Econometrica* 58, 809-837.
- [37] NEWEY, W. K., AND J. POWELL (2003): “Instrumental Variables Estimation for Nonparametric Models,” *Econometrica*, 71, 1565–1578.
- [38] PAKES, A. AND POLLARD, D. (1989): “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57, 1027-58.
- [39] PETROV, V. V. (1975): *Sums of independent random variables*. Springer, New York.

- [40] PÖTSCHER, B. M. (2002): “Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters,” *Econometrica*, 70, 1035–1065.
- [41] ROBINSON, P. M. (1976): “Instrumental Variables Estimation of Differential Equations,” *Econometrica*, 44, 765–776.
- [42] STINCHCOMBE, M. and WHITE, H. (1998): “Consistent Specification Testing with Nuisance Parameters Present only Under the Alternative,” *Econometric Theory*, 14, 295–325.
- [43] STAIGER, D. AND STOCK, J.H. (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557–586.
- [44] STOCK, J.H. AND WRIGHT, J.H. (2000): “GMM with Weak Identification,” *Econometrica*, 68, 1055–1096.
- [45] STOCK, J. H., WRIGHT, J. H. AND YOGO, M. (2002): “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics*, 20, 518–529.
- [46] VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer Series in Statistics. Springer-Verlag, New York, 1 edn.
- [47] WOLFOWITZ, J. (1957): “The minimum distance method,” *The Annals of Mathematical Statistics*, 28, 75–88.
- [48] WOOLDRIDGE, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- [49] YOGO, M. (2004): “Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak,” *Review of Economics and Statistics*, 86, 797–810.