Uniformly Consistent Estimation of Linear Regression Models with Strictly Exogenous Instruments*

Juan Carlos Escanciano‡
Indiana University
March 22, 2016

Abstract

This paper investigates estimation of linear regression models with strictly exogenous instruments under minimal identifying assumptions. Commonly used Instrumental Variables (IV) estimators are not uniformly consistent in this setting (uniformity is in the underlying data generating process). This negative result is just one way to formalize the well-documented fact of high sensitivity of IV to the presence of weak instruments. This paper introduces a uniformly consistent estimator under nearly the minimal identifying assumption. The proposed estimator, called the Integrated Instrumental Variables (IIV) estimator, is a weighted least squares estimator with trivial implementation. Monte Carlo evidence supports the theoretical claims and suggests that the IIV estimator is a robust alternative to IV and optimal IV in finite samples under weak identification and strictly exogenous instruments. 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 contrast to IV methods that fail to identify these parameters.

Keywords: 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 and Ed Vytlacil for useful comments. A previous 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, and Dufour, 1997). 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 uniform estimation with strictly exogenous instruments.

First, this paper discusses identification under a minimal assumption (i.e. a necessary and sufficient condition for identification). This identification analysis uncovers a tradeoff between the strength of the exogeneity and relevance conditions. Classical IV requires in general a weaker exogeneity condition than strict exogeneity, but this strict relaxation is at the cost of a stronger relevance condition than the minimal assumption used in this paper. This tradeoff may be most relevant for weak instruments, for which the classical rank condition fails, or it is close to being violated, but the minimal assumption may hold. An application to estimating the elasticity of intertemporal substitution with quarterly UK and USA data provides two of such cases in an empirically relevant setting. The paper then proposes an estimator, called the Integrated Instrumental Variables (IIV) estimator, that is uniformly consistent under nearly the minimal identifying assumption. The IIV estimator is a weighted least squares estimator with trivial implementation. Monte Carlo simulations show that the uniform consistency property translates into stable finite sample behaviour for the IIV estimator, comparing favorably with classical IV and optimal IV estimators. The new IIV estimator estimates a positive and significant elasticity of intertemporal substitution of 0.5 with quarterly UK data and an equally sensible estimate for its reciprocal of 1.94, in contrast to IV methods that fail to identify these parameters and provide puzzling low estimates for both parameters.

The observed data is a random sample from $W \equiv (Y, X', Z')'$, which is defined on the probability space $(\Omega, \mathcal{F}, P)$ 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 I 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.

I now introduce the minimal identifying assumption for \( \theta_0 \) in (1)-(2). I 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 discussion of completeness in a fully non-parametric 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 \( \tilde{Z} := E [X | Z] \). Then, with this notation linear completeness means \( \tilde{Z}' \lambda = 0 \) implies \( \lambda = 0 \), or equivalently

**Assumption LC**: \( E[\tilde{Z} \tilde{Z}'] \) is positive definite.

An example may help to fix ideas. Suppose \( X \) is a binary endogenous variable, \( X \in \{0, 1\} \), and \( Z \) is the instrument used in IV. Then, \( \tilde{Z} = P(X = 1 | Z) \). Assumption LC is then equivalent to \( \tilde{Z} \neq 0 \). The generally stronger IV relevance condition requires in addition that \( E [Z \tilde{Z}] \neq 0 \). Optimal IV uses as instrument a positive multiple of \( \tilde{Z} \), and as such, also identifies under the weak condition \( \tilde{Z} \neq 0 \).

3
Assumption LC is a nonparametric generalization of the classical rank condition in the sense that the nonparametric regression \( \tilde{Z} = E[Z|X] \) 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 minor addition relative to Donald and Newey (2001). A more important contribution of this paper is uncovering a tradeoff between the exogeneity and relevance conditions in the context of strict exogenous instruments, which to my knowledge has not been discussed before. The first result establishes the minimality of Assumption LC for identification.\(^1\) Its proof follows from the arguments above and therefore is omitted.

**Proposition 1:** In (1)-(2), \( \theta_0 \) is identified if and only if Assumption LC holds.

It follows from Proposition 1 that Assumption LC is implied by the traditional rank condition \( \text{rank}(E[ZX']) = p \) (and \( \text{rank}(E[ZZ']) = k \)), which is assumed for identification with classical IV methods. Assumption LC allows for the possibility of having less instruments than endogenous variables, \( k < p \), by exploiting (nonparametric) nonlinearities in the dependence between \( X \) and \( Z \). In contrast, in the classical linear identification assumption the order condition \( k \geq p \) is necessary. Thus, from this discussion one concludes that Classical IV requires the weaker exogeneity condition \( E[\varepsilon Z] = 0 \), but this relaxation is at the cost of a stronger relevance condition than Assumption LC. As discussed above, this tradeoff is most relevant for weak instruments, for which the classical rank condition fails, or it is close to being violated, but Assumption LC may hold.

Uniformity of inferences in the underlying probability measure \( P \) over a large class of probabilities is considered a desirable property.\(^2\) Henceforth, the dependence on \( P \) will be emphasized, as in e.g. \( m_P(Z) := E_P[X|Z] \). Assume \( P \in \mathcal{P} \), where \( \mathcal{P} \) is the class of probabilities satisfying the assumptions above in a uniform sense, i.e.

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

where \( \lambda_{\text{min}}(A) \) denotes the minimum eigenvalue of a symmetric matrix \( A \), and \( C \) is a generic

---

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

\(^2\)Of course, as noted by Bahadur and Savage (1956), the class cannot be too large.
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 would allow for $C$ to depend on $P$). Henceforth, I refer to this stronger condition as the “nearly minimal identifying assumption”.

Let $\hat{\theta}_{IV}$ denote the Two-Stage Least Squares (TSLS) estimator (assuming it exists). The following result shows that this commonly used estimator is not uniformly consistent over $P$. Its proof is simple: the TSLS estimator fails to identify $\theta_0$ over $P$; i.e. the mapping

$$P \to \theta_{IV}(P) := \left( E_P[XZ'] (E_P[ZZ'])^{-1} E_P[ZX'] \right)^{-1} E_P[XZ'] (E_P[ZZ'])^{-1} E_P[ZY]$$

is not defined for all $P \in \mathcal{P}$, as $\mathcal{P}$ contains joint distributions violating the IV rank condition.

**Proposition 2:** TSLS is not uniformly consistent over $\mathcal{P}$, i.e. there exists $\varepsilon > 0$ such that

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

This result holds even if I include in $\mathcal{P}$ the classical rank condition $\text{rank}(E_P[ZX']) = p$, as there are converging sequences of probabilities in $\mathcal{P}$ (with respect to total variational distance $d(P, Q) = 2 \sup_A |P(A) - Q(A)|$) satisfying the rank condition for all elements of the sequence but for its limit (see, e.g., Staiger and Stock, 1997). The proof of this result follows from an application of Proposition 1.A in Bickel et al. (1993). That is, IV fails to identify $\theta_0$ “uniformly over $\mathcal{P}$”. See Dufour (1997) for related discussion.

This paper introduces an estimator that is uniformly consistent over $\mathcal{P}$. The new Integrated Instrumental Variables (IIV) estimator of $\theta_0$ is a weighted least squares estimator, computed for a sample $\{W_i\}_{i=1}^n$ of size $n$ as

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

where $X$ is the $n \times p$ design matrix with rows $X'_i$, $Y = (Y_1, ..., Y_n)'$, $\Omega$ 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 $\sqrt{n}(\hat{\theta}_{IIV} - \theta_0)$ is (uniformly) asymptotically normal with an asymptotic variance that is consistently estimated by $\hat{\Gamma} = \hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Sigma}^{-1}$, where $\hat{\Sigma} = (X'\Omega X)/n^2$, $\hat{\Lambda} = (X'\Omega D\Omega X)/n^3$ and $D$ is a diagonal

---

3Identification in $\mathcal{P}$ is referred to as “uniform identification” with respect to $\mathcal{P}$ in Manski (1988, p.4). In this paper, we refer to “uniform identification over $\mathcal{P}$” as the following stronger condition: identification holds on the closure of $\mathcal{P}$ (with respect to the topology of weak convergence). Since $\mathcal{P}$ is closed both concepts coincide in the setting of this paper.
matrix with components \( \{\varepsilon_i^2\}_{i=1}^n \), where \( \varepsilon_i = Y_i - X_i'\hat{\theta}_{IIV} \) are IIV residuals.

An estimator that provides identification of \( \theta_0 \) over the whole \( \mathcal{P} \) is the optimal IV estimator (cf. Amemiya, 1974, 1977, Robinson, 1976, Chamberlain 1982, Newey 1990). This IV estimator uses the optimal instruments \( Z^*: = \sigma^{-2}_P(Z)m_P(Z) \), where \( \sigma^2_P(Z): = E_P[\varepsilon^2|Z] \). However, this estimator also fails to be uniformly consistent over \( \mathcal{P} \). This failure follows from \( \sigma^2_P(Z) \) being either too close to zero or to infinity. See also the general results in Pötscher (2002), who showed the lack of uniformly consistent estimators of \( m_P(Z) \) and \( \sigma^2_P(Z) \) over \( \mathcal{P} \). These results are independent of the estimators used for \( m_P(Z) \) and \( \sigma^2_P(Z) \), and apply equally to kernel, series or other estimators that solve ill-posed problems; see e.g. Donald and Newey (2001) and Carrasco (2012). The Monte Carlo simulations below provide empirical evidence of the lack of uniform consistency for optimal IV estimators, and suggest that the new estimator significantly outperforms optimal IV estimates when the minimal identifying Assumption LC is close to being violated.

The new estimator belongs to the class of minimum distance estimators, and is similar in spirit to the estimator proposed by Dominguez and Lobato (2004). These authors pointed out that in nonlinear models, IV, or the more general Generalized Method of Moments, may fail to identify. This problem does not arise in the linear setting of this paper, as optimal IV identifies \( \theta_0 \) over \( \mathcal{P} \). The distinctive feature of the present paper is the emphasis on uniformity (in identification, consistency and asymptotic normality). There is previous research in econometrics on uniform inference, but this work has mainly focused on uniformly valid confidence sets and tests, and different models. Kasy (2015) has recently investigated uniformity in the Delta Method, and has shown that the IV estimator fails to be uniformly asymptotically normal in the classical IV setting. The present paper uncovers a tradeoff between the exogeneity and relevance conditions with strictly exogenous instruments and shows its potential utility in the presence of weak instruments through a new uniformly consistent and asymptotically normal estimator.

The rest of the paper is organized as follows. Section 2 introduces formally the IIV estimator, and shows that this estimator identifies \( \theta_0 \) uniformly over \( \mathcal{P} \). 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 IV estimators, including optimal ones. 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

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

where \( i := \sqrt{-1} \) and \( V := E_P[ZZ'] \). Likewise, define \( h_{x,P}(u) := E_P[X \exp(iu'V^{-1/2}Z)] \) 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 \( |A| = (\text{tr}(A^cA))^{1/2} \), 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 (4) yields

\[ h_{y,P}^c(u) = h_{x,P}^c(u)\theta_0, \]  

Multiplying both sides of (5) by \( h_{x,P}(u) \), evaluating \( u = U \), with \( U \) a standard Gaussian random vector in \( \mathbb{R}^k \), and taking expectations with respect to \( U \) yield

\[ E_U[h_{x,P}(U)h_{y,P}^c(U)] = E_U[h_{x,P}(U)h_{x,P}^c(U)]\theta_0. \]  

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 estimator provided above). Moreover, identification is preserved as shown in the following result proved in the Appendix.

**Proposition 3:** Assumption LC is equivalent to

\[ E_U[h_{x,P}(U)h_{x,P}^c(U)] \text{ is positive definite.} \]  

In view of Proposition 3, under Assumption LC \( \theta_0 \) is identified over the whole \( \mathcal{P} \) as

\[ \theta_0 \equiv \theta_0(P) = (E_U[h_{x,P}(U)h_{x,P}^c(U)])^{-1} E_U[h_{x,P}(U)h_{y,P}^c(U)]. \]
This identification result and the analog principle suggest the IIV estimator

\[ \hat{\theta}_{IIV} \equiv \theta_0(P_n) = \left( E_U[h_{x,P_n}(U)h_{x,P_n}(U)] \right)^{-1} E_U[h_{x,P_n}(U)h_{y,P_n}(U)]; \]

(9)

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) and Stinchcombe and White (1998), other 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 \( E_{P}[Y w(Z, u)] \), 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 e.g. Escanciano (2006) for further discussion and 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

This section establishes uniform asymptotic theory for \( \hat{\theta}_{IIV} \).\(^5\) 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 \(-\)integrable functions \( \phi \) of the \( k \)-dimensional multivariate standard Gaussian density. In \( L_2(\phi) \) define the inner product

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

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

\(^5\)Uniform convergence in distribution means the following: \( T_n \xrightarrow{d} T \) uniformly over \( \mathcal{P} \) iff \( \sup_{P \in \mathcal{P}} |E_P[f(T_n)] - E_P[f(T)]| \to 0 \) for all bounded and continuous functions \( f(\cdot) \).
Using the introduced notation, the IIV estimator can be written as

$$\hat{\theta}_{IIV} = \langle h_{x,P_n}, h_{x,P_n} \rangle^{-1} \langle h_{x,P_n}, h_{y,P_n} \rangle,$$  \hspace{1cm} (10)

where, with some abuse of notation, I 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.

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

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

An implication of uniform consistency is that the mapping that IIV is identifying, i.e $P \rightarrow \theta_0(P)$ in (8), is continuous over $\mathcal{P}$ (with the variational distance); see Proposition 1.A in Bickel et al. (1993).

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

$$\sqrt{n}(\hat{\theta}_{IIV} - \theta_0) = \langle h_{x,P}, h_{x,P} \rangle^{-1} \sqrt{n} \langle h_{x,P}, h_{\varepsilon,P_n} \rangle + o_P(1).$$  \hspace{1cm} (11)

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

$$\sqrt{n} \langle h_{x,P}, h_{\varepsilon,P_n} \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i H_{x,P}(Z_i) + o_P(1),$$  \hspace{1cm} (12)

where $H_{x,P}(u) := E_P[X \exp(-0.5(u - Z)' V^{-1} (u - Z))]$. Equations (11) and (12) show that the IIV estimator has a uniformly asymptotic linear representation.

For uniform asymptotic normality I 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 restrictions 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.

$$\sqrt{n}(\hat{\theta}_{IIV} - \theta_0) \rightarrow_d N(0, \Gamma_P)$$

uniformly in $\mathcal{P}^{AN}$.

Note that $\Sigma_P$ is non-singular uniformly in $\mathcal{P}^{AN}$ (cf. Proposition 1). The estimation of the asymptotic variance $\Gamma_P$ was discussed in the Introduction. Confidence regions and tests based on the variance estimator and the asymptotic normality in Theorem 2 inherit the uniformity of the IIV estimator, and hence are asymptotically (uniformly) valid.

4 Monte Carlo Simulations

This section investigates the finite sample performance of the proposed estimator in comparison with IV methods in two simulation experiments. Specifically, I compare the performance of the IIV estimator ($\hat{\theta}_{IIV}$), the standard IV estimator ($\hat{\theta}_{IV}$) and the optimal IV (OIV) estimator ($\hat{\theta}_{OIV}$).

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

$$DGP1: \begin{cases} Y = \theta_0 X + \varepsilon \\ X = \gamma Z^2 + \nu \end{cases} \quad \left( \begin{array}{c} \varepsilon \\ \nu \end{array} \right) \sim N\left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right) \right),$$

with $Z \sim N(\mu, 1)$ drawn independently of $\varepsilon$ and $\nu$. The true parameter is zero, i.e. $\theta_0 = 0$.

Several values for the parameters $(\mu, \gamma, \rho)$ are considered. Since $E_P[XZ] = \gamma(\mu^3 + 3\mu)$, $\mu$ and $\gamma$ measure jointly the level of identification of the IV estimator using the instrument $Z$. Also, $\gamma$ alone measures the level of nonparametric identification, since $E_P[m^2_P(Z)] = \gamma^2 E_P[Z^4]$. The correlation parameter $\rho$ measures the level of endogeneity. In the simulations I consider several combinations of the parameter values $\mu \in \{0, 0.1, 0.5, 1\}$, $\gamma \in \{0.1, 0.5, 1\}$ and $\rho \in \{0.3, 0.9\}$. In all the experiments the number of Monte Carlo replications is 1000. All estimators are computed using standardized instruments (i.e. using $Z_i/\hat{\sigma}$, where $\hat{\sigma} = \sqrt{V}$ is the sample standard deviation of $\{Z_i\}$).

The optimal estimator $\hat{\theta}_{OIV}$ requires nonparametric estimation of $m_P(Z) = E_P[X \mid Z]$ and $\sigma^2_P(Z) = E_P[\varepsilon^2 \mid Z]$. I 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 I prefer to keep the exposition simple. I have considered other bandwidth choices, but they led to qualitatively similar conclusions.

Table 1 reports the bias, standard error (SE), and root mean squared (RMS) error for
the IV, OIV and IIV estimates for several values of \((\mu, \gamma, \rho)\) in DGP1 with \(n = 50\). It is evident from Table 1 that classical IV methods are very sensitive to the values of the parameters \((\mu, \gamma, \rho)\). In contrast, the proposed estimator has a satisfactory performance across the different parameter values (i.e. uniformly in the data generating processes). In particular, the results for the IIV estimator are not sensitive to the level of endogeneity \(\rho\). In contrast, IV estimators are very sensitive to \(\rho\), specially so under weak identification. For small and moderate values of \(\mu\) and \(\gamma\) the SE of IV and optimal IV is much higher than that of the proposed estimator (in some cases by two orders of magnitude). The new estimator has in most cases lower biases than optimal IV and IV estimators.

Table 1 also illustrates the difference between the classical identification assumption and the nonparametric identification condition in Assumption LC. The IV estimator performs badly when there is weak linear identification (\(\mu\) is small) but moderate or strong nonparametric identification (\(\gamma\) is large), as evidenced in the case \(\mu = 0.1\), \(\gamma = 1\) and \(\rho = 0.9\). Note also how the optimal IV estimator is highly affected by the weak nonparametric identification, to the extent that the non-optimal IV leads in several cases to smaller RMS (see e.g. \(\mu = 0.5\), \(\gamma = 0.1\) and \(\rho = 0.9\)). For completeness, I also report at the bottom of Table 1 the case with no nonparametric identification (\(\gamma = 0\)). Again, the proposed estimator has the lowest RMS. Unreported results with \(n = 200\) show a reduction in biases and standard errors when nonparametric identification holds (\(\gamma > 0\)). But somewhat unexpectedly, there are cases where the SE of optimal IV increases as the sample size increases, even when \(\gamma > 0\).

In sum, these results show an omnibus good performance of the IIV estimator, comparing favorably with more traditional estimators, specially when identification is weak or moderate.
TABLE 1
BIAS, STANDARD ERROR AND ROOT MEAN SQUARE ERROR. N=50

<table>
<thead>
<tr>
<th>μ</th>
<th>γ</th>
<th>ρ</th>
<th>BIAS</th>
<th>SE</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>IV</td>
<td>OIV</td>
<td>IIIV</td>
</tr>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>-6.37</td>
<td>0.12</td>
<td>0.27</td>
</tr>
<tr>
<td>0.9</td>
<td>10.03</td>
<td>0.65</td>
<td>355.40</td>
<td>10.69</td>
<td>0.43</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.49</td>
<td>0.01</td>
<td>20.20</td>
<td>0.57</td>
</tr>
<tr>
<td>0.9</td>
<td>0.66</td>
<td>0.14</td>
<td>9.59</td>
<td>10.62</td>
<td>0.23</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>-0.01</td>
<td>0.01</td>
<td>4.53</td>
<td>0.12</td>
</tr>
<tr>
<td>0.9</td>
<td>0.19</td>
<td>-0.03</td>
<td>0.01</td>
<td>20.41</td>
<td>0.20</td>
</tr>
<tr>
<td>0.1</td>
<td>0.3</td>
<td>-6.83</td>
<td>0.24</td>
<td>0.23</td>
<td>0.72</td>
</tr>
<tr>
<td>0.9</td>
<td>0.54</td>
<td>-2.52</td>
<td>0.71</td>
<td>19.37</td>
<td>0.41</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-1.02</td>
<td>0.09</td>
<td>37.67</td>
<td>3.21</td>
</tr>
<tr>
<td>0.9</td>
<td>0.26</td>
<td>-0.14</td>
<td>0.06</td>
<td>21.29</td>
<td>0.24</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.57</td>
<td>-0.01</td>
<td>0.14</td>
<td>6.86</td>
</tr>
<tr>
<td>0.9</td>
<td>11.74</td>
<td>-0.04</td>
<td>0.01</td>
<td>41.65</td>
<td>0.24</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.38</td>
<td>0.25</td>
<td>0.21</td>
<td>1.13</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.61</td>
<td>-3.20</td>
<td>0.60</td>
<td>13.20</td>
<td>0.44</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.08</td>
<td>0.01</td>
<td>2.81</td>
<td>0.74</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.09</td>
<td>-0.07</td>
<td>0.03</td>
<td>1.44</td>
<td>0.36</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.10</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.33</td>
<td>0.11</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.55</td>
<td>-3.27</td>
<td>0.11</td>
</tr>
<tr>
<td>0.9</td>
<td>0.17</td>
<td>2.21</td>
<td>0.32</td>
<td>13.43</td>
<td>0.47</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.00</td>
<td>-0.01</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>0</td>
<td>0.3</td>
<td>0.21</td>
<td>-0.05</td>
<td>0.28</td>
<td>25.55</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.62</td>
<td>1.67</td>
<td>0.91</td>
<td>57.40</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Note: μ and γ measure the level of IV and Nonparametric identification, respectively; ρ measures the endogeneity level.
In the second experiment I consider a regression model with an endogenous binary variable. The same model was investigated in Newey (1990). I generate data $Y$, $X$ and $Z$ according to the model

\[
DGP_2: \begin{cases} 
Y = \theta_{01} + X\theta_{02} + \varepsilon \\
X = 1(\alpha_{01} + \alpha_{02}Z + \nu > 0) 
\end{cases}, \quad \begin{pmatrix} \varepsilon \\ \nu \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),
\]

with $Z \sim N(0,1)$ independent of $\varepsilon$ and $\nu$. The true parameters are $\theta_{01} = \theta_{02} = 1$. The parameter $\alpha_{01}$ is fixed at $\alpha_{01} = 1$, but different values for the parameters $(\alpha_{02}, \rho)$ 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 $\rho$ measures the level of endogeneity. The parameter values considered are $\alpha_{02} \in \{1, 0.5, 0.1\}$ and $\rho \in \{0.2, 0.8\}$.

I compare the performance of the IIV estimator, the standard IV estimator using the instrument $Z$ for $X$, and the optimal IV estimator computed with a series estimator for the optimal instrument $p(Z)$. Note that no estimation of the conditional variance is considered here for simplicity. I follow Newey (1990) and estimate $p(z)$ by $\hat{p}(z) = A^q(z)'\gamma_n$, where $A^q(z) = (A_1(z), ..., A_k(z))$, $A_j(z) = z^{j-1}$, $j = 1, ..., q$, $\gamma_n = [S_n' S_n]^{-1} S_n' Y$, and

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

I 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 estimates of $\theta_{01}$ and $\theta_{02}$ for a sample size of $n = 100$. For all parameter values considered the IIV estimator for $\theta_{01}$ presents a smaller RMS than IV and optimal IV, specially for low values of $\alpha_{02}$. It is remarkable the robustness of the new estimator to weak nonparametric identification when $\alpha_{02} = 0.1$. For this case the RMS of the IV estimator is more than 50 times that of the IIV estimate. For $\theta_{02}$, the optimal IV has the smallest RMS for large values of $\alpha_{02}$, but its performance substantially deteriorates when $\alpha_{02}$ decreases, showing one more time that optimal IV is rather sensitive to weak nonparametric identification.
Table 2

Bias, Standard Error and Root Mean Square Error. DGP2. \( n = 100 \)

<table>
<thead>
<tr>
<th>( \alpha_{02} )</th>
<th>( \rho )</th>
<th>BIAS</th>
<th>SE</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_{01} )</td>
<td>1</td>
<td>0.2</td>
<td>1.01</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>1.02</td>
<td>1.76</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2</td>
<td>1.08</td>
<td>1.81</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.71</td>
<td>1.78</td>
<td>1.00</td>
</tr>
<tr>
<td>( 0.1 )</td>
<td>0.2</td>
<td>5.73</td>
<td>2.05</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.68</td>
<td>1.51</td>
<td>0.99</td>
</tr>
<tr>
<td>( \theta_{02} )</td>
<td>1</td>
<td>0.2</td>
<td>0.98</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.96</td>
<td>0.31</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2</td>
<td>0.90</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>1.35</td>
<td>0.79</td>
<td>1.01</td>
</tr>
<tr>
<td>( 0.1 )</td>
<td>0.2</td>
<td>-4.59</td>
<td>-0.07</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>2.99</td>
<td>4.00</td>
<td>2.07</td>
</tr>
</tbody>
</table>

Note: \( \alpha_{02} \) measures the level of Nonparametric identification; \( \rho \) 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, IV and optimal IV estimates are rather sensitive to weak identification, which empirically confirms its uniform inconsistency. These simulations suggest that the new estimator can be a sensible and simple alternative to standard IV procedures when the researcher is concerned about weak or moderate identification.

5 An Application to Estimating the Elasticity of Inter-temporal 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

\[
E_{\theta}[\Delta c_{t+1} - \alpha - \psi r_{t+1} | Z_t] = 0
\]

where \( \psi \) is the elasticity of intertemporal substitution (EIS), \( \Delta c_{t+1} \) is the growth rate of consumption, \( r_{t+1} \) is the log gross return on some asset, \( \alpha \) 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 (13) by several estimation strategies; see Hansen and Singleton (1983) and Hall (1988). One is a TSLS estimator with $\Delta c_{t+1}$ 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 $\psi$ and $1/\psi$ are simultaneously small and not statistically significant.

I revisit this issue applying the new IIV estimator to an international data set considered in Campbell (2003) and Yogo (2004). 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).

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 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 $\Delta c_t$, is the first difference in log real consumption per capita. Following Yogo (2004), I 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, $\pi_t$ 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.

---

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

7 This 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 $W_t$. 

15
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. I 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 (13) 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.

$$P (E_P[r_{t+1} | Z_t] \neq E_P[r_{t+1}]) > 0. \quad (14)$$

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

$$E_P[r_{t+1} - \beta - (1/\psi)\Delta c_{t+1} | Z_t] = 0 \quad (15)$$

provided

$$P (E_P[\Delta c_{t+1} | Z_t] \neq E_P[\Delta c_{t+1}]) > 0. \quad (16)$$

There is an extensive literature in econometrics on nonparametric tests for the hypothesis in (14) and (16). See, e.g., Bierens (1982) and Stinchcombe and White (1998). 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 $\Delta c_{t+1}$, $r_{f,t+1}$ or $r_{e,t+1}$, thus testing the linear completeness conditions in (14) and (16). The test statistics are quadratic forms in the residuals $\hat{u} = (\hat{u}_1, ..., \hat{u}_n)'$, 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

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

where $\Omega$ is defined after (3) and $\hat{\sigma}_n^2 = n^{-1} \sum_{t=1}^n \hat{u}_t^2$. I standardize the components of $Z_t$ by their sample standard deviation, so tests become scale invariant. Note that one may think
of \( C_{V,M_n} \) as a nonparametric extension of the classical first-stage \( F \)-test in IV regression.\(^8\)

The asymptotic distribution of \( C_{V,M_n} \) 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 \( C_{V,M_n} \) by that of

\[
C_{V,M_n}^* = \frac{\hat{u}^*\Omega\hat{u}^*}{n\hat{\sigma}_u^2},
\]

where \( \hat{u}^* = (\hat{u}_0^*, ..., \hat{u}_n^*)_t \), \( \hat{u}_t^* = (V_t\hat{u}_t - c_n) \), \( c_n = n^{-1}\sum_{t=1}^n V_t\hat{u}_t \), \( \hat{\sigma}_u^2 = n^{-1}\sum_{t=1}^n \hat{u}_t^2 \), 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

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

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 \( C_{V,M_n}^* \) 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 \( C_{V,M_n} \geq c_{n,\alpha}^* \), where \( c_{n,\alpha}^* \) is the \((1 - \alpha)\)th empirical quantile of \( B \) realizations of \( C_{V,M_n}^* \). Alternatively, I can use the bootstrap \( p \)-values, \( p_n^* \) say, rejecting \( H_0 \) when \( p_n^* < \alpha \), where

\[
p_n^* = \Pr \left( C_{V,M_n}^* \geq C_{V,M_n} \mid \{Y_t, X_t, Z_t\}_{t=1}^{n+1} \right).
\]

Table 3 reports the bootstrap p-values for the data sets considered and the three variables \( \Delta c_{t+1}, r_{f,t+1} \) and \( r_{e,t+1} \). The number of bootstrap replications is \( B = 5000 \). For completeness, I 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 predictability of interest rates. When looking for nonlinear (nonparametric) predictability, I 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 \( \psi \) and \( 1/\psi \) are nonparametrically identified for UK and USA data, but \( 1/\psi \) is weakly identified by IV methods. Thus, the UK and USA data highlight the difference between linear and nonlinear predictability, and hence the difference between lack

\(^8\)The \( F \)-test is computed as \( F = (\hat{u}'P_{Z,n}\hat{u})/n\hat{\sigma}_P^2 \), where \( P_{Z,n} = Z(Z'Z/n)^{-1}Z' \), \( \hat{\sigma}_P^2 = 4SSR_u/(n - 5) \) and \( SSR_u \) is the sum of unrestricted squared residuals.
of identification by linear methods and the linear completeness assumption. In view of this evidence, in what follows I 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>
<thead>
<tr>
<th>Country</th>
<th>Sample Period</th>
<th>$Y_{t+1} = \Delta c_{t+1}$</th>
<th>$Y_{t+1} = r_{f,t+1}$</th>
<th>$Y_{t+1} = r_{e,t+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUL</td>
<td>1970.3-1998.4</td>
<td>0.35 (1.79)</td>
<td>0.00 (21.81)</td>
<td>0.34 (1.82)</td>
</tr>
<tr>
<td>CAN</td>
<td>1970.3-1999.1</td>
<td>0.44 (3.03)</td>
<td>0.00 (15.37)</td>
<td>0.60 (2.51)</td>
</tr>
<tr>
<td>FR</td>
<td>1970.3-1998.3</td>
<td>0.90 (0.17)</td>
<td>0.00 (38.43)</td>
<td>0.61 (3.09)</td>
</tr>
<tr>
<td>GER</td>
<td>1979.1-1998.3</td>
<td>0.54 (0.83)</td>
<td>0.00 (17.66)</td>
<td>0.82 (0.69)</td>
</tr>
<tr>
<td>ITA</td>
<td>1971.4-1998.1</td>
<td>0.67 (0.73)</td>
<td>0.00 (19.01)</td>
<td>0.81 (1.10)</td>
</tr>
<tr>
<td>JAP</td>
<td>1970.3-1998.4</td>
<td>0.50 (1.18)</td>
<td>0.00 (8.64)</td>
<td>0.34 (3.49)</td>
</tr>
<tr>
<td>NTH</td>
<td>1977.3-1998.2</td>
<td>0.59 (0.89)</td>
<td>0.00 (12.05)</td>
<td>0.94 (0.73)</td>
</tr>
<tr>
<td>SWD</td>
<td>1970.3-1999.2</td>
<td>0.62 (0.48)</td>
<td>0.00 (17.08)</td>
<td>0.15 (2.24)</td>
</tr>
<tr>
<td>SWT</td>
<td>1976.2-1998.4</td>
<td>0.83 (0.97)</td>
<td>0.00 (8.55)</td>
<td>0.97 (0.11)</td>
</tr>
<tr>
<td>UK</td>
<td>1970.3-1999.1</td>
<td><strong>0.07 (2.52)</strong></td>
<td><strong>0.00 (17.04)</strong></td>
<td>0.60 (2.62)</td>
</tr>
<tr>
<td>USA</td>
<td>1947.3-1998.4</td>
<td><strong>0.00 (2.93)</strong></td>
<td><strong>0.00 (15.53)</strong></td>
<td>0.19 (2.88)</td>
</tr>
</tbody>
</table>

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

### 5.2 Uniformly Consistent Estimation and Inference

I now proceed to estimate $\psi$ in (13) and $1/\psi$ in (15) using the IIV estimator and interest rates and consumption growth for UK and USA data. I 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. I 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 $\psi$ 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, I do not get contradictory results when estimating $1/\psi$, with an estimate of 1.94 that is fairly consistent with the IIV estimate of $\psi$. For USA data, I 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**

<table>
<thead>
<tr>
<th>Country</th>
<th>Sample Period</th>
<th>$\psi$</th>
<th>IIV</th>
<th>TSLS</th>
<th>LIML</th>
<th>$1/\psi$</th>
<th>IIV</th>
<th>TSLS</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>1970.3-1999.1</td>
<td>0.50</td>
<td>0.17</td>
<td>0.16</td>
<td>1.94</td>
<td>1.06</td>
<td>6.21</td>
<td>(0.20)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>USA</td>
<td>1947.3-1998.4</td>
<td>0.66</td>
<td>0.06</td>
<td>0.03</td>
<td>1.41</td>
<td>0.68</td>
<td>34.11</td>
<td>(0.49)</td>
<td>(0.09)</td>
</tr>
</tbody>
</table>

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, I 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). I 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**

<table>
<thead>
<tr>
<th>Country</th>
<th>Sample Period</th>
<th>$\psi$</th>
<th>IIV</th>
<th>AR</th>
<th>CLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>1970.3-1999.1</td>
<td>[0.10, 0.90]</td>
<td>[0.04, 0.28]</td>
<td>[-0.12, 0.43]</td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>1947.3-1998.4</td>
<td>[-0.31, 1.63]</td>
<td>$\emptyset$</td>
<td>[-0.19, 0.22]</td>
<td></td>
</tr>
</tbody>
</table>
5.3 Nonparametric Tests for Overidentifying Restrictions

The proposed IIV estimator is based on the strict exogeneity assumption in (13). 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, I 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 $H_0: P (E_P[\Delta c_{t+1} - \alpha - \psi r_{t+1} | Z_t] = 0) = 1$, against the general nonparametric alternative $H_1: P (E_P[\Delta c_{t+1} - \alpha - \psi r_{t+1} | Z_t] \neq 0) > 0$. The CvM test is computed as a quadratic form in the residuals $\hat{u} = (\hat{u}_1, ..., \hat{u}_n)'$, with $\hat{u}_t = \Delta c_{t+1} - \hat{\alpha} - \hat{\psi} r_{t+1}$, $\hat{\theta}_{IIV} = (\hat{\alpha}, \hat{\psi})'$, simply as

$$CvM_n = \frac{1}{\sigma^2_n} \hat{u}' \Omega_{ind}^{'} \Omega_{ind} \hat{u},$$

where $\Omega_{ind} = H \Psi$, $H = I_n - X(X'X)^{-1}X'$, $I_n$ is the $n \times n$ identity matrix, and $\Psi$ 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 $\hat{u}^* = (V_1 \hat{u}_1, ..., V_n \hat{u}_n)'$, and $\{V_i\}_{i=1}^n$ are generated from (17). This bootstrap can be justified along the lines of Dominguez and Lobato (2003). These tests require consistent estimation of the parameter $\theta_0$, 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 under weak instruments in applications of economic interest. In addition, I have shown that it 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 I have emphasized minimal conditions for identification of linear regression models when a strict exogeneity assumption holds. Under the minimal identifying assumption the commonly used IV estimators are not uniformly consistent, given the plausibility of weak instruments. I 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 they are very sensitive to the level of nonparametric identification. This empirical evidence is theoretically supported by the lack of uniform consistency of optimal IV. 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.

It remains a topic of future research 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 do 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.
Appendix: Mathematical Proofs

Proof of Proposition 3: I shall prove that the negation of Assumption LC is equivalent to the negation of (7). That is, the negation of Assumption LC implies that there exists a non-zero vector, say \( \lambda \in \mathbb{R}^p \), such that

\[ E_P[\lambda'X \mid Z] = 0 \text{ a.s.} \quad (18) \]

Now, by the law of iterated expectations and Theorem 1 in Bierens (1982), (18) is equivalent to

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

Hence, \( \lambda'E_U[h_{x,P}(U)h_{x,P}(U)]\lambda = 0. \)

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 \( \Pi \) 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 \( \mathcal{F}_X \). I shall 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

\[ \left| \int_{\mathbb{R}^k \setminus \Pi} \phi(u)du \right| \leq \delta. \quad (19) \]

Then, write

\[
\langle h_{x,P_n}, h_{x,P_n} \rangle = \int_{\mathbb{R}^k} h_{x,P_n}(u)h_{x,P_n}^c(u)\phi(u)du.
\]

\[
= \int_{\Pi} h_{x,P_n}(u)h_{x,P_n}^c(u)\phi(u)du + \int_{\mathbb{R}^k \setminus \Pi} h_{x,P_n}(u)h_{x,P_n}^c(u)\phi(u)du.
\]

\[
= : I_{1n,P} + I_{2n,P}. \quad (20)
\]

I 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, I shall apply Theorem 2.8.1 of that reference. First,
note that the envelop satisfies

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

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), I 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 \( 1/n \), \( N(\varepsilon, \mathcal{F}_X, \|\cdot\|_{Q,1}) \) 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 \( \Pi \), in particular to \( \Pi_V = \{uV^{1/2} : u \in \Pi \} \). 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

\[ I_{1n,P} = \int_{\Pi} h_{x_j,P}(u)h_{x_k,P}(u)\phi(u)du + o_P(1) \text{ uniformly in } P \in \mathcal{P}. \]

On the other hand, it is straightforward to show that

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

by (19) and the inequality \( |h_{x_j,P_n}(u)h_{x_k,P_n}(u)| \leq |E_{P_n}[F_X]|^2 = O_P(1) \) uniformly in \( P \in \mathcal{P} \). Since \( \delta > 0 \) was arbitrary, conclude from (20) that for all \( j, k = 1, \ldots, p \),

\[ \langle h_{x_j,P_n}, h_{x_k,P_n} \rangle = \langle h_{x_j,P}, h_{x_k,P} \rangle + o_P(1) \text{ uniformly in } P \in \mathcal{P}. \]

The proof of

\[ \langle h_{x_j,P_n}, h_{y,P_n} \rangle = \langle h_{x_j,P}, h_{y,P} \rangle + o_P(1) \text{ uniformly in } P \in \mathcal{P}, \]

23
follows the same arguments, and hence is omitted. □

**Proof of Theorem 2:** The arguments of the proof of Theorem 1 yield that $h_{x,P_n} \xrightarrow{L_2} h_x$ 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 $M > 0$,

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

$$
\leq M^{-2} \sup_{P \in \mathcal{P}^{AN}} E_P \left[ \varepsilon^2 \right] 
$$

$$
\leq M^{-2} C,
$$

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

$$
\sqrt{n}(\hat{\theta}_{IV} - \theta_0) = \langle h_{x,P}, h_{x,P} \rangle^{-1} \sqrt{n} \langle h_{x,P}, h_{\varepsilon,P_n} \rangle + o_P(1) 
$$

$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \langle h_{x,P}, h_{x,P} \rangle^{-1} H_{x,P}(Z_i) \varepsilon_i + o_P(1). \tag{21}
$$

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

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

Then, the $2 + \delta$ finite moment of $\varepsilon$ implies that the uniform CLT of Petrov (1975, p. 118) is applicable, and the uniform asymptotic normality of $\sqrt{n}(\hat{\theta}_{IV} - \theta_0)$ follows from (21). □
References


