The Integrated Instrumental Variables Estimator: Exploiting Nonlinearities for Identification of Linear Models

Juan Carlos Escanciano

Indiana University
February 2, 2010

This paper can be downloaded without charge from the Social Science Research Network electronic library at: http://ssrn.com/abstract=1550332.

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: http://www.indiana.edu/~caepr. CAEPR can be reached via email at caep@indiana.edu or via phone at 812-855-4050.

©2008 by NAME. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.
The Integrated Instrumental Variables Estimator:
Exploiting Nonlinearities for Identification of Linear Models

Juan Carlos Escanciano

This version: February 2, 2010.

Abstract: A new estimator for linear models with endogenous regressors and
strictly exogenous instruments is proposed. The new estimator, called the In-
tegrated Instrumental Variables (IIV) estimator, only requires minimal assump-
tions to identify the true parameters, thereby providing a potential robust alter-
native to classical Instrumental Variables (IV) methods when instruments and
endogenous variables are partially uncorrelated (i.e. weak identification holds)
but are nonlinearly dependent. The IIV estimator is simple to compute, as it can
be written as a weighted least squares estimator and it does not require to solve
an ill-posed problem and the subsequent regularization. Monte Carlo evidence
suggests that the IIV estimator can be a valuable alternative to IV and opti-
mal IV in finite samples under weak identification. An application to estimating
the elasticity of intertemporal substitution highlights the merits of the proposed
approach over classical IV methods.

Key words and phrases: Identification; Instrumental variables; Weak instru-
ments; Efficient IV; Intertemporal elasticity of substitution.

JEL Classification Number: C13.

1Department of Economics, Indiana University, email: jescanci@indiana.edu. Research funded by the
Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908. I thank Lin Zhu for excellent research
assistance, and Marine Carrasco, David Jacho-Chávez, Pravin Trivedi and Ed Vytlacil for useful comments.
All errors are my own.
1 Introduction

The linear regression model continues to be the workhorse model in most empirical applications. When facing possible endogeneity, these models are often estimated by instrumental variables (IV) and related methods where some type of exogeneity and “relevancy” are required properties for the validity of the resulting inference. Recently, there has been a very active body of research devoted to the development of inference methods robust to failures of the relevancy property, leading to the so-called weak instruments literature (see Stock, Wright and Yogo, 2002, and Andrews and Stock, 2005, for recent surveys). A tenet of the traditional and the more recent literatures on instrumental variables has been the focus on identification by exploiting the linear variability among endogenous and exogenous variables. In contrast, this article proposes identification and inference methods that exploit nonlinear variability, thereby providing a potential solution to the problem of estimation under weak instruments in situations where endogenous and exogenous variables are partially uncorrelated but nonlinearly dependent.

To be more precise, let \( \{W_t \equiv (Y_t, X_t', Z_t')' : t \in \mathbb{Z}\} \) be a strictly stationary and ergodic time series vector defined on the probability space \((\Omega, \mathcal{F}, P)\), and taking values in \(\mathbb{R}^p \times \mathbb{R}^k\), \(p, k \in \mathbb{N}\). Henceforth, \(A^c, A'\) and \(|A|\) denote the complex conjugate, the matrix transpose and the Euclidean norm \(|A| = (tr(A^c A))^{1/2}\), respectively. We assume the components of \(W_t\) are related through the linear model\(^2\)

\[
Y_t = X_t' \theta_0 + \varepsilon_t, \tag{1}
\]

where \(\theta_0 \in \mathbb{R}^p\), and where the so-called strict exogeneity assumption holds, i.e.

\[
E[\varepsilon_t | Z_t] = 0 \text{ almost surely (a.s.)} \tag{2}
\]

This article deals with identification and estimation of the model defined in (1)-(2). Our methods rely on the strict exogeneity in (2) and will not be applicable in cases where (2) does not hold. Thus, condition (2) will be a maintained assumption throughout the article. In some economic applications, such as in dynamic Euler equations, the strict exogeneity naturally holds provided \(Z_t\) is in the agent’s information set at time \(t - 1\). In other applications, such as dynamic panel data models, the strict exogeneity requires serious thought and introspection, see e.g. Wooldridge (2002, Chapter 7) for discussion.

The following identification assumption plays a crucial role in our development. We

\(^2\)For the sake of exposition, we only consider scalar responses, but our methods can be easily extended to systems of equations.
call this assumption the *Conditional Identification* (CI) assumption, emphasizing the use of conditional moments, as opposed to more traditional rank conditions in IV that use unconditional moments. Define \( m(Z_t) := E[X_t \mid Z_t] \).

**Assumption CI**: \( E[m(Z_t)m'(Z_t)] \) is positive definite (p.d.).

Related conditions to Assumption CI have been used in the econometrics literature for nonparametric identification of certain models, see e.g. Das (2005) and Cai et al. (2006), but its use in linear models has not been emphasized. From standard arguments of linear regression it follows that Assumption CI is sufficient and necessary for identification of the parameters in the model defined by (1)-(2). In particular, it follows that under (2), CI is weaker than the traditional rank condition \( \text{rank}(E[Z_tX_t']) = p \), which is usually required for identification by classical IV methods. Throughout this article we shall refer to the classical rank identification condition as the *linear* or *unconditional* identification assumption. The difference between the rank condition and Assumption CI is illustrated in the following example.

**Example 1**: Assume that \( Y, X \) and \( Z \) are such that

\[
\begin{align*}
Y &= \alpha_0 + X\beta_0 + \varepsilon \\
X &= Z^2 + u
\end{align*}
\]

with \( Z \sim N(0,1) \) independent of \( \varepsilon \) and \( u \). In this example, IV with instrument \( Z \) does not identify \( \theta_0 = (\alpha_0, \beta_0)' \), since \( E[XZ] = \gamma E[Z^3] = 0 \), but \( \theta_0 \) is identified under CI, simply as

\[
\alpha_0 = E[Y] - E[X]\beta_0, \quad \beta_0 = \frac{E[Y \mid Z] - E[Y]}{E[X \mid Z] - E[X]}.
\]

Notice that Assumption CI in this example boils down to \( P(E[X \mid Z] \neq E[X]) > 0 \). \( \Box \)

We stress that Assumption CI allows for the possibility of having less instruments than endogenous variables, \( k < p \), by exploiting non-linearities in the dependence between \( X_t \) and \( Z_t \). This stands in sharp contrast with the classical linear identification assumption, where the order condition \( k \geq p \) is necessary. In this article, we are particularly interested in situations where CI holds but traditional rank conditions for IV do not or are close to not being satisfied (i.e. the so-called weak identification holds). This is likely to occur in some economic applications, see the examples discussed in Stock et al. (2002). An example where nonlinearity naturally arises is the case of regression with endogenous dummy variables, as in the models discussed in Heckman (1978). See also Das (2005) for a related identification result in more general nonparametric models with discrete endogenous variables.
Example 2: Assume that $Y$, $D$, and $X$ are related through the regression

$$Y = \beta_0 + \beta_1 D + \beta_2' X + \varepsilon,$$

where $X$ is exogenous but $D$ is an endogenous dummy variable. For instance, $D$ can be an indicator of treatment. Assumption C1 entails the propensity score $p(x) = \Pr(D = 1 \mid X = x)$ to be nonlinear in $x$. □

The main purpose of this article is to develop inferential procedures that rely just on the identification assumption CI, and not on the satisfaction of the classical linear identification conditions. Thus, these inferential methods have the potential to be robust to weak instruments, by exploiting nonlinear variation between endogenous and exogenous variables.

We propose a new estimator that achieves identification under the minimal identification assumption CI. The new Integrated Instrumental Variables (IIV) estimator for $\theta_0$ is simply a weighted least squares estimator, computed as

$$\hat{\theta}_{IIV} := (X'WX)^{-1}(X'WY), \tag{3}$$

where $X$ is the $n \times p$ design matrix with rows $X_t'$, $Y = (Y_1, \ldots, Y_n)'$, $W$ is the $n \times n$ matrix with elements $\exp(-0.5|Z_t - Z_s|^2)$, $1 \leq t, s \leq n$, and $n \geq 1$ is the sample size. Section 2 provides motivation for this choice of weighting matrix $W$. We prove below that unlike other popular choices of $W$, e.g. IV methods where $W = P_Z = Z(Z'Z)^{-1}Z'$, our choice of $W$ guarantees identification of $\theta_0$ even in cases where instruments provide weak or no linear identification but Assumption CI still holds.

Another estimator that provides identification under the minimal Assumption CI is the optimal IV estimator (cf. Amemiya, 1974, 1977, Robinson, 1976, Chamberlain 1982, Newey 1990). This IV estimator uses the optimal instruments $Z_t^* := \sigma^{-2}(Z_t)m(Z_t)$, where $\sigma^2(Z_t) := E[\varepsilon_t^2 \mid Z_t]$. It is interesting to notice that this identification result of efficient estimators in linear models does not translate into non-linear models, as recently shown by Dominguez and Lobato (2004). That is, for linear models not only the optimal IV estimator is convenient in terms of minimum asymptotic variance, but it also relies on minimal identification assumptions, which makes this estimator a natural benchmark to compare with.

A drawback of optimal IV is that it requires nonparametric smoothing estimation of $m(Z_t)$ and $\sigma^{-2}(Z_t)$; see the aforementioned references. In contrast, the IIV estimator in (3) guarantees identification under CI without the need to estimate nonparametric local quantities. Although the proposed estimator is not as efficient as optimal IV asymptotically,
it can easily outperform optimal IV in finite-samples of moderate size when conditional identification is weak or the sample size is small relative to the number of instruments \( k \); as shown in our simulations below.

The rest of the article is organized as follows. Section 2 introduces formally the IIV estimator, and provides intuition as to why the new estimator guarantees identification under CI. In Section 3 we establish the asymptotic distribution theory for the IIV estimator. Section 4 reports the results of some Monte Carlo experiments that suggest that the proposed estimator performs well in finite-samples and compares favorably with IV estimators, including optimal ones. Section 5 contains an application to estimating the elasticity of intertemporal substitution for international data. Finally, Section 6 concludes and discusses future research.

2 The Integrated Instrumental Variables Estimator

This section introduces the IIV estimator. To that end, we shall assume that the following moment is well defined

\[
h_y(v) := E[Y_t \exp(iv'Z_t)], \quad v \in \mathbb{R}^k,
\]

where \( i = \sqrt{-1} \). Likewise, define \( h_x(v) := E[X_t \exp(iv'Z_t)] \) and \( h_e(v) := E[\varepsilon_t \exp(iv'Z_t)] \).

Note that by (2) and the law of iterated expectations, \( h_x(\cdot) \equiv 0 \). Since \( h_y(v) \) is a linear transformation in \( Y_t \), we obtain

\[
h_y^c(v) = h_x^c(v)\theta_0. \tag{5}
\]

Multiplying both sides of (5) by \( h_x(v) \), evaluating \( v \) at \( v = V \), with \( V \) a standard Gaussian random vector in \( \mathbb{R}^k \), and taking expectations with respect to \( V \) we get

\[
E[h_x(V)h_y^c(V)] = E[h_x(V)h_x^c(V)]\theta_0 \tag{6}
\]

Note that \( E[h_x(V)h_y^c(V)] \) and \( E[h_x(V)h_y^c(V)] \) are both real-valued and exist under the following moment assumption:

**Assumption A1:** \( E[|Y_t|^2] < \infty \) and \( E[|X_t|^2] < \infty \).

Next proposition proves that Assumption CI is equivalent to

\[
E[h_x(V)h_y^c(V)] \text{ is p.d.} \tag{7}
\]

**Proposition 1:** Under A1, Assumption CI is equivalent to (7).
Proof of Proposition 1: We shall prove that the negation of CI is equivalent to the negation of (7). That is, the negation CI implies that there exists a non-zero vector, say $\alpha \in \mathbb{R}^p$, such that $\alpha' E[m(Z_t)m'(Z_t)]\alpha = 0$, that is,

$$E[\alpha' X_t \mid Z_t] = 0 \text{ a.s.} \quad (8)$$

Now, by the law of iterated expectations and Theorem 1 in Bierens (1982), (8) is equivalent to

$$\alpha' h_x(v) = E[E[\alpha' X_t \mid Z_t] \exp(iv'Z_t)] = 0 \text{ almost everywhere (a.e.)}$$

Hence, $\alpha' E[h_x(V)h_x^c(V)]\alpha = 0$. $\square$

In view of Proposition 1, under CI, $\theta_0$ is identified as $\theta_0 = (E[h_x(V)h_x^c(V)])^{-1} E[h_x(V)h_{y}^c(V)]$. The IIV estimator is then defined as

$$\hat{\theta}_{IIV} := \left(E[\hat{h}_x(V)\hat{h}_x^c(V)]\right)^{-1} E[\hat{h}_x(V)\hat{h}_y^c(V)], \quad (9)$$

where $\hat{h}_y(v) := n^{-1}\sum_{s=1}^{n} Y_s \exp(iv'Z_s)$ and $\hat{h}_x(v) := n^{-1}\sum_{s=1}^{n} X_s \exp(iv'Z_s)$. For computational purposes we recommend to compute $\hat{\theta}_{IIV}$ as in (3). As evidenced from (9), the estimator $\hat{\theta}_{IIV}$ has the interpretation of the population slope of the least squares regression of $\hat{h}_y(V)$ on $\hat{h}_x(V)$.

Some remarks about our estimator are in order. First, it should be noted that our approach is related to that in Carrasco (2008), although her interest is in efficiency, rather than in (conditional or unconditional) identification. In particular, one of her estimators can be considered as an efficient version of $\hat{\theta}_{IIV}$ by regularizing a suitable inverse operator. She shows that the resulting estimator is equivalent to an optimal IV with a series estimator for the optimal instrument; see also Donald and Newey (2001). Our estimator is considerably simpler, as it does not require any regularization, at the cost of being less efficient than optimal IV estimators in large samples. Our Monte Carlo evidence suggests, however, that our estimator can outperform optimal IV estimates when conditional identification is weak, which motivates the present approach. We stress that this finding is robust to alternative implementations of optimal IV estimators such as series or kernel nonparametric estimates.

Second, we point out that our theory can be extended to other weighting functions $W$. For instance, if we use the transformation $E[Y_11(Z_t \leq v)]$ in (4) we get an estimator that is closely related to one proposed by Dominguez and Lobato (2004) in a more general set-up.\footnote{In this case the estimator is computed as in (3) but with $W = PP'$, where $P$ is the $n \times n$ matrix with elements $(1(Z_t \leq Z_s)_{t,s})$, with $1(\cdot)$ the indicator function, i.e. $1(A) = 1$ if $A$ occurs, 0 otherwise.} These authors’ interest was in non-linear models and lack of identification of the unconditional
model. In contrast, our interest here is in linear models and conditional identification. Our choice of the exponential function over the indicator function is motivated by the fact that using the indicator \(1(Z_t \leq v)\) can lead to misleading inferences in applications where we have a large number of instruments at our disposal, i.e. large \(k\). More precisely, when \(k\) is moderate or large, say greater than 3 for commonly used sample sizes, it is likely that in the given sample most of the indicators \(1(Z_t \leq v)\) are zero when \(v\) is evaluated at the sample observations \(Z_s\). As a consequence, the integrated estimator using \(1(Z_t \leq v)\) will present finite-sample identification problems due to singularity of the indicators. This is an integrated version of the so called “curse of dimensionality” problem in nonparametric inference. The exponential weight is more robust to this “curse of dimensionality” problem. Nevertheless, there are other weighting schemes that share with the exponential function this robust-to-dimensionality property, see e.g. Escanciano (2006). Although all our results can be extended to general weighting schemes \(E[Y_t w(Z_t, v)]\) for suitable transformations \(w\) satisfying some conditions, see Stinchcombe and White (1998) and Bierens and Ploberger (1997) for such conditions, in this article we have preferred to keep the exposition simpler and we have restricted ourselves to exponential weights for the sake of clarity.

3 Asymptotic Theory for the IIV Estimator

This section establishes the asymptotic distribution of \(\hat{\theta}_{IIV}\). To that end, it is convenient for our purposes to make use of the asymptotic theory for random elements taking values in a suitable Hilbert space; see Chen and White (1996, 1998) for applications of this approach in econometrics. We need some basic notation. Let \(L_2(\phi)\) be the Hilbert space of all complex-valued and square \(\phi\)-integrable functions, where \(\phi\) is the \(k\)-dimensional multivariate standard Gaussian density. In \(L_2(\phi)\) we define the inner product

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

\(L_2(\phi)\) is endowed with the natural Borel \(\sigma\)-field induced by the norm \(\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}\). We denote by \(L_2 \xrightarrow{\text{convergence in probability}}\) convergence in probability in \(L_2(\phi)\), i.e., \(h_n \xrightarrow{L_2} h \iff \|h_n - h\| \xrightarrow{P} 0\).

Using the introduced notation, we can write

\[
\hat{\theta}_{IIV} = \theta_0 + \langle \hat{h}_x, \hat{h}_x \rangle^{-1}\langle \hat{h}_x, \hat{h}_e \rangle,
\]

where, with some abuse of notation we have denoted by \(\langle \hat{h}_x, \hat{h}_x \rangle\) the \(n \times n\) matrix with elements \(\langle \hat{h}_{xi}, \hat{h}_{xj} \rangle\), where \(\hat{h}_{xi}\) is the \(i - th\) component of \(\hat{h}_x\). Now, consistency of \(\hat{\theta}_{IIV}\) will
follow from a law of large numbers in $L_2(\phi)$ for $\hat{h}_x$ and $h_\varepsilon$, i.e. from $\hat{h}_x \xrightarrow{L_2} h_x$ and $\hat{h}_\varepsilon \xrightarrow{L_2} h_\varepsilon$, and the continuity of the inner product. In this article we shall use a law of large numbers for strong mixing sequences, but the more general results in Chen and White (1996) can also be applied.

Similarly, our assumptions will imply that

$$\sqrt{n}(\hat{\theta}_{IV} - \theta_0) = (h_x, h_x)^{-1}\sqrt{n}(h_x, \hat{h}_\varepsilon) + o_P(1).$$

Notice that the term $\sqrt{n}(h_x, \hat{h}_\varepsilon)$ is simply a standardized sample mean, i.e.

$$\sqrt{n}(h_x, \hat{h}_\varepsilon) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t H_x(Z_t),$$

where $H_x(v) := \langle h_x, \exp(-i \cdot v) \rangle = E[X_s \exp(-0.5 |v - Z_s|^2)]$.

By the Central Limit Theorem (CLT) for mixing sequences, $\sqrt{n}(\hat{\theta}_{IV} - \theta_0)$ will converge to a multivariate normal distribution with zero mean and variance matrix $\Sigma^{-1} \Lambda \Sigma^{-1}$, where $\Sigma := \langle h_x, h_x \rangle$ and

$$\Lambda := \sum_{j=-\infty}^{\infty} E[\varepsilon_t \varepsilon_{t-j} H_x(Z_t) H'_x(Z_{t-j})].$$

Let $\mathcal{F}_t := \sigma(W_t, W_{t-1}, \ldots)$ be the $\sigma$-field generated by the time series vector $W_t$ up to time $t$. If $\varepsilon_t$ is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$, then $\Lambda$ boils down to $\Lambda = E[\varepsilon_t^2 H_x(Z_t) H'_x(Z_t)]$. If $\varepsilon_t$ presents some correlation with elements in $\mathcal{F}_{t-1}$, some assumptions are necessary for the existence of $\Lambda$ and the verification of the CLT. In this article we assume that $\{W_t\}$ is a strictly stationary and $\alpha$-mixing process, with strong ($\alpha$-) mixing coefficients

$$\alpha_j := \sup_t \sup_{A \in \mathcal{F}_{-\infty}, B \in \mathcal{F}_{t+j}} |P(A \cap B) - P(A)P(B)|,$$

satisfying the following assumption:

**Assumption A2:** $\sum_{j=1}^{\infty} \alpha_j^{\delta/(\delta+2)} < \infty$, where $\delta > 0$ is such that $E[|\varepsilon_t H_x(Z_t)|^{2+\delta}] < \infty$.

Assumption A2 is standard in the literature. In particular, it implies that $\Lambda$ exists and that a CLT for (10) holds; see Theorem 2.1 in Politis and Romano (1994). For extensions to heterogenous time series see Chen and White (1998).

**Theorem 1:** Suppose Assumptions A1 and A2 hold. Under (2) and CI

$$\sqrt{n}(\hat{\theta}_{IV} - \theta_0) \rightarrow_d N(0, \Sigma^{-1} \Lambda \Sigma^{-1}).$$
Proof of Theorem 1: By the law of large numbers for $\alpha$-mixing sequences in $L_2(\phi)$, it holds $\hat{h}_x \xrightarrow{L_2} h_x$. On the other hand, the CLT for mixing sequences implies $\sqrt{n}\hat{h}_x = O_P(1)$. These asymptotic results and the continuity of the inner product yield

$$\sqrt{n}(\hat{\theta}_{IIV} - \theta_0) = \langle h_x, h_x \rangle^{-1}\sqrt{n}\langle h_x, \hat{h}_x \rangle + o_P(1).$$

A further application of the CLT and Slutsky’s Theorem imply the asymptotic normality of $\sqrt{n}(\hat{\theta}_{IIV} - \theta_0)$. □

The identification assumption CI is minimal, as discussed above. This condition implies that $\Sigma$ is non-singular (cf. Proposition 1). The estimation of the asymptotic variance is straightforward: $\Sigma$ can be estimated by $\hat{\Sigma} = (X'WX)/n^2$, whereas $\Lambda$ can be estimated by any heteroskedasticity and autocorrelation consistent (HAC) estimator, see e.g. Andrews (1991) and Newey and West (1994). We refer the reader to the aforementioned references for results on consistency on the these variance estimators. We remark that if $\varepsilon_i$ is a martingale difference sequence, then HAC is not needed and a natural estimator becomes $\hat{\Lambda} = (X'W\hat{\Omega}WX)/n^3$ with $\hat{\Omega}$ a diagonal matrix with components $\left\{\hat{\varepsilon}_t^2\right\}_{t=1}^n$.

4 Monte Carlo Simulations

This section investigates the finite-sample performance of the proposed estimator in comparison with other popular methods in two simulation experiments. Specifically, we compare the performance of our 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 we consider a slight variation of Example 1 from the Introduction. We generate data $Y$, $X$ and $Z$ according to the model

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

with $Z \sim N(\mu, 1)$ drawn independently of $\varepsilon$ and $u$. The true parameter is zero, i.e. $\theta_0 = 0$. We consider several values for the parameters $(\mu, \gamma, \rho)$. Since $E[XZ] = \gamma E[Z^3] = \gamma(\mu^3 + 3\mu)$, $\mu$ and $\gamma$ measure jointly the level of identification of the IV estimator using the instrument $Z$. Also, $\gamma$ alone measures the level of conditional identification, since $E[m^2(Z)] = \gamma^2 E[Z^4]$. The correlation parameter $\rho$ measures the level of endogeneity. In the simulations we 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 we consider 1000 Monte Carlo replications.
The optimal estimator $\hat{\theta}_{OIV}$ requires nonparametric estimation of $E[X \mid Z]$ and $\sigma^2(Z) = E[\varepsilon^2 \mid Z]$. 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 squared error of the resulting estimator, see Linton (2002). Here $\hat{\sigma}$ is the sample standard deviation of $\{Z_i\}_{i=1}^n$. The term $1.06\hat{\sigma}$ is not necessarily the optimal one, but we prefer to keep the exposition simple. We considered other bandwidth choices, but they led to qualitatively similar conclusions.

Table I 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 I 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. In particular, the results for the IIV estimator are not sensitive to the level of endogeneity $\rho$. For small and moderate values of $\mu$ and $\gamma$ the SE of IV and optimal IV are much higher than that of the proposed estimator. In particular, there are several cases where the RMS of our estimator is smaller than 100 times those of the optimal IV and IV estimators. The new estimator has in most cases lower biases than optimal IV and IV estimators. The RMS errors of the three estimators decrease with the level of conditional identification for a given level of endogeneity, that is, the larger $\gamma$ the lower the RMS, for a given value of $\rho$.

Tables I also illustrates the difference between the classical identification assumption and the conditional identification in CI. It is interesting to see the bad performance of the IV estimator when there is weak linear identification ($\mu$ is small) but moderate or strong conditional identification ($\gamma$ is large), as evidenced in the case $(\mu, \gamma, \rho) = (0.1, 1, 0.9)$. Note also how the optimal IV estimator is highly affected by the weak conditional identification, to the extent that the non-optimal IV leads in several cases to smaller RMS. For completeness, we also report at the bottom of Table I the case with no conditional identification ($\gamma = 0$). Again, the proposed estimator has the lowest RMS.

Table II contains the results for $n = 200$. In general, we observe a reduction in biases and standard errors when conditional identification holds ($\gamma > 0$). But somewhat unexpectedly, there are cases where the SE of optimal IV increases as the sample size increases, even when $\gamma > 0$, see for instance the case $(\mu, \gamma) = (0, 0.1)$. An interesting feature of Table II is the low bias of the proposed estimator relative to IV methods.

In sum, these results show an omnibus outperformance of the IIV estimator over more traditional estimators, specially when identification is weak or moderate.
## TABLE I

**Bias, Standard Error and Root Mean Square Error. n=50**

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(\gamma)</th>
<th>(\rho)</th>
<th>BIAS</th>
<th>SE</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>-6.376</td>
<td>0.121</td>
<td>0.274</td>
</tr>
<tr>
<td>0.9</td>
<td>10.033</td>
<td>0.659</td>
<td>0.731</td>
<td>355.403</td>
<td>10.698</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.493</td>
<td>-0.085</td>
<td>0.011</td>
<td>20.204</td>
</tr>
<tr>
<td>0.9</td>
<td>0.666</td>
<td>0.141</td>
<td>0.070</td>
<td>9.597</td>
<td>10.622</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.098</td>
<td>-0.004</td>
<td>0.004</td>
<td>4.539</td>
</tr>
<tr>
<td>0.9</td>
<td>0.195</td>
<td>-0.037</td>
<td>0.013</td>
<td>20.415</td>
<td>0.204</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>-6.829</td>
<td>-0.243</td>
<td>0.229</td>
</tr>
<tr>
<td>0.9</td>
<td>0.542</td>
<td>2.526</td>
<td>0.712</td>
<td>19.377</td>
<td>101.543</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-1.026</td>
<td>0.096</td>
<td>0.033</td>
<td>37.674</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.266</td>
<td>-0.143</td>
<td>0.062</td>
<td>21.297</td>
<td>6.866</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.577</td>
<td>-0.004</td>
<td>0.010</td>
<td>26.838</td>
</tr>
<tr>
<td>0.9</td>
<td>11.746</td>
<td>-0.045</td>
<td>0.066</td>
<td>41.650</td>
<td>0.242</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.616</td>
<td>3.202</td>
<td>0.603</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.088</td>
<td>-0.004</td>
<td>0.014</td>
<td>2.817</td>
<td>0.742</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.093</td>
<td>-0.071</td>
<td>0.030</td>
<td>1.440</td>
<td>0.366</td>
</tr>
<tr>
<td>1</td>
<td>-0.012</td>
<td>-0.003</td>
<td>-0.000</td>
<td>0.223</td>
<td>0.102</td>
</tr>
<tr>
<td>0.9</td>
<td>11.746</td>
<td>0.016</td>
<td>0.011</td>
<td>0.339</td>
<td>0.111</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.552</td>
<td>-3.274</td>
<td>0.112</td>
</tr>
<tr>
<td>0.9</td>
<td>0.169</td>
<td>2.213</td>
<td>0.320</td>
<td>13.435</td>
<td>31.377</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.003</td>
<td>-0.010</td>
<td>0.003</td>
<td>0.111</td>
<td>0.175</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.009</td>
<td>-0.031</td>
<td>0.007</td>
<td>0.115</td>
<td>0.146</td>
</tr>
<tr>
<td>1</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.000</td>
<td>0.053</td>
<td>0.065</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.002</td>
<td>-0.003</td>
<td>0.003</td>
<td>0.054</td>
<td>0.064</td>
</tr>
<tr>
<td>0</td>
<td>0.3</td>
<td>0.215</td>
<td>0.049</td>
<td>0.284</td>
<td>25.548</td>
</tr>
<tr>
<td>0.9</td>
<td>-0.625</td>
<td>1.674</td>
<td>0.910</td>
<td>57.407</td>
<td>17.477</td>
</tr>
</tbody>
</table>
### TABLE II

**Bias, Standard Error and Root Mean Square Error. n=200**

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\gamma$</th>
<th>$\rho$</th>
<th>IV</th>
<th>OIV</th>
<th>IIIV</th>
<th>IV</th>
<th>OIV</th>
<th>IIIV</th>
<th>IV</th>
<th>OIV</th>
<th>IIIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.785</td>
<td>-0.831</td>
<td>0.137</td>
<td>26.080</td>
<td>35.078</td>
<td>0.519</td>
<td>2.609</td>
<td>35.880</td>
<td>0.536</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.530</td>
<td>-0.472</td>
<td>0.439</td>
<td>17.464</td>
<td>16.911</td>
<td>0.413</td>
<td>17.472</td>
<td>16.918</td>
<td>0.603</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>0.063</td>
<td>-0.014</td>
<td>-0.000</td>
<td>14.995</td>
<td>0.120</td>
<td>0.142</td>
<td>14.995</td>
<td>0.121</td>
<td>0.142</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.236</td>
<td>-0.039</td>
<td>0.008</td>
<td>17.791</td>
<td>0.155</td>
<td>0.130</td>
<td>17.792</td>
<td>0.160</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.512</td>
<td>0.001</td>
<td>0.004</td>
<td>7.750</td>
<td>0.062</td>
<td>0.071</td>
<td>7.767</td>
<td>0.062</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.187</td>
<td>-0.008</td>
<td>0.001</td>
<td>8.526</td>
<td>0.070</td>
<td>0.068</td>
<td>8.528</td>
<td>0.071</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.3</td>
<td>-0.038</td>
<td>-0.070</td>
<td>0.165</td>
<td>14.205</td>
<td>6.538</td>
<td>0.528</td>
<td>14.205</td>
<td>6.539</td>
<td>0.554</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1.158</td>
<td>-1.307</td>
<td>0.415</td>
<td>15.816</td>
<td>67.639</td>
<td>0.400</td>
<td>15.858</td>
<td>67.651</td>
<td>0.577</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.302</td>
<td>-0.003</td>
<td>0.011</td>
<td>6.454</td>
<td>0.127</td>
<td>0.132</td>
<td>6.461</td>
<td>0.127</td>
<td>0.132</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.033</td>
<td>-0.016</td>
<td>0.014</td>
<td>9.226</td>
<td>0.135</td>
<td>0.128</td>
<td>9.226</td>
<td>0.136</td>
<td>0.128</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.337</td>
<td>0.000</td>
<td>0.001</td>
<td>14.414</td>
<td>0.061</td>
<td>0.071</td>
<td>14.418</td>
<td>0.061</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.097</td>
<td>-0.006</td>
<td>0.003</td>
<td>4.220</td>
<td>0.065</td>
<td>0.067</td>
<td>4.221</td>
<td>0.066</td>
<td>0.067</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.431</td>
<td>0.213</td>
<td>0.064</td>
<td>6.924</td>
<td>7.055</td>
<td>0.426</td>
<td>6.937</td>
<td>7.058</td>
<td>0.430</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.179</td>
<td>0.732</td>
<td>0.252</td>
<td>14.682</td>
<td>87.858</td>
<td>0.356</td>
<td>14.683</td>
<td>87.861</td>
<td>0.436</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.007</td>
<td>-0.006</td>
<td>0.000</td>
<td>0.110</td>
<td>0.125</td>
<td>0.104</td>
<td>0.111</td>
<td>0.126</td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.013</td>
<td>-0.019</td>
<td>0.001</td>
<td>0.109</td>
<td>0.116</td>
<td>0.099</td>
<td>0.110</td>
<td>0.118</td>
<td>0.099</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.002</td>
<td>-0.004</td>
<td>0.001</td>
<td>0.051</td>
<td>0.064</td>
<td>0.048</td>
<td>0.051</td>
<td>0.064</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.001</td>
<td>-0.003</td>
<td>0.001</td>
<td>0.053</td>
<td>0.053</td>
<td>0.051</td>
<td>0.053</td>
<td>0.053</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>-0.034</td>
<td>0.142</td>
<td>0.023</td>
<td>0.325</td>
<td>22.013</td>
<td>0.292</td>
<td>0.326</td>
<td>22.013</td>
<td>0.293</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.078</td>
<td>-0.191</td>
<td>0.045</td>
<td>0.340</td>
<td>6.398</td>
<td>0.288</td>
<td>0.349</td>
<td>6.401</td>
<td>0.291</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>-0.000</td>
<td>-0.003</td>
<td>0.001</td>
<td>0.052</td>
<td>0.076</td>
<td>0.061</td>
<td>0.052</td>
<td>0.076</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.003</td>
<td>-0.005</td>
<td>0.000</td>
<td>0.052</td>
<td>0.092</td>
<td>0.061</td>
<td>0.052</td>
<td>0.092</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.025</td>
<td>0.036</td>
<td>0.030</td>
<td>0.025</td>
<td>0.036</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.024</td>
<td>0.039</td>
<td>0.029</td>
<td>0.024</td>
<td>0.039</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.3</td>
<td>0.429</td>
<td>-1.773</td>
<td>0.328</td>
<td>7.622</td>
<td>26.902</td>
<td>0.805</td>
<td>7.634</td>
<td>26.961</td>
<td>0.870</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>-0.572</td>
<td>0.349</td>
<td>0.903</td>
<td>57.482</td>
<td>12.834</td>
<td>0.395</td>
<td>57.485</td>
<td>12.839</td>
<td>0.986</td>
<td></td>
</tr>
</tbody>
</table>
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

\begin{align*}
\text{DGP}2: \begin{cases} 
Y = \theta_{01} + X\theta_{02} + \varepsilon \\
X = 1(\alpha_{01} + \alpha_{02}Z + u > 0) \\
(\varepsilon, u) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)
\end{cases}
\end{align*}

with $Z \sim N(0,1)$ independent of $\varepsilon$ and $u$. The true parameters are $\theta_{01} = \theta_{02} = 1$. The parameter $\alpha_{01}$ is fixed at $\alpha_{01} = 1$, but we consider different values for the parameters $(\alpha_{02}, \rho)$. As discussed in Example 2, the parameters $\theta_0 = (\theta_{01}, \theta_{02})'$ are identified if the propensity score $p(z) = P(X = 1 \mid Z = z)$ is nonlinear in $z$. Since $p(z) = \Phi(\alpha_{01} + \alpha_{02}z)$, conditional identification holds provided $\alpha_{02} \neq 0$. Thus, $|\alpha_{02}|$ measures the level of (conditional) identification. The correlation parameter $\rho$ measures the level of endogeneity. In the simulations we consider several combinations of the parameter values $\alpha_{02} \in \{1, 0.5, 0.1\}$ and $\rho \in \{0.2, 0.8\}$.

We compare the performance of our 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. We follow Newey (1990) and estimate $p(z)$ by $\hat{p}(z) = A^q(z)'\gamma_n$, where $A^q(z) = (A_1(z),...,A_k(z))$, $A_j(z) = z^{j-1}$, $j = 1,...,q$, $\gamma_n = [S_n' S_n]^{-1}S_n' Y$;

\[
Y := \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad \text{and} \quad S_n := \begin{bmatrix} A^q(Z_1)' \\ \vdots \\ A^q(Z_n)' \end{bmatrix}.
\]

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

Table III reports results for estimates of $\theta_{01}$ and $\theta_{02}$ for a sample size of $n = 100$. For all parameter values considered our 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 conditional 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 conditional identification.
TABLE III
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></td>
<td></td>
<td>IV</td>
<td>OIV</td>
<td>IV</td>
</tr>
<tr>
<td>$\theta_{01}$</td>
<td>1</td>
<td>0.2</td>
<td>1.010</td>
<td>1.760</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>1.029</td>
<td>1.762</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2</td>
<td>1.082</td>
<td>1.813</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.710</td>
<td>1.781</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
<td>5.732</td>
<td>2.049</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.687</td>
<td>1.514</td>
<td>0.098</td>
</tr>
<tr>
<td>$\theta_{02}$</td>
<td>1</td>
<td>0.2</td>
<td>0.986</td>
<td>0.284</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.967</td>
<td>0.308</td>
<td>0.978</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2</td>
<td>0.901</td>
<td>0.211</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>1.356</td>
<td>0.795</td>
<td>1.009</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
<td>-4.589</td>
<td>-0.072</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>2.991</td>
<td>4.000</td>
<td>2.074</td>
</tr>
</tbody>
</table>

To summarize, the new estimator performs quite well in finite samples of small and moderate size when conditional and unconditional identification is weak. In contrast, IV and optimal IV estimates are rather sensitive to weak identification. 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 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

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

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 (11) by several estimation strategies; see Hansen and Singleton (1983) and Hall (1988). We can estimate $\theta_0$ by a two-stage least squares (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.

We revisit this issue using our new tools for 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). In addition, a longer time series is available at annual frequency for Sweden, the United Kingdom and the United States. 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 and frequency, see Table IV. 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), 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, $\pi_t$ 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 IV here, and used inference methods (confidence intervals and test statistics) that were robust to identification failure. His main conclusion, in agreement with previous findings, was that the EIS is small and not significant across the eleven countries considered.

The weak identification 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 article emphasizes, it is possible that these variables are linearly unpredictable but

---

4 The data set is available at Motohiro Yogo’s web page. We thank Motohiro for making the data available.
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 investigate the use of this possible nonlinearity in identification of the linearized CCAPM. Our theoretical results imply that the parameter $\theta_0$ is identified in the regression (11) if and only if

$$P(E[r_{t+1} \mid Z_t] \neq E[r_{t+1}]) > 0.$$  \hspace{1cm} (12)

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

$$E[r_{t+1} - \beta - (1/\psi)\Delta c_{t+1} \mid Z_t] = 0$$

provided $P(E[\Delta c_{t+1} \mid Z_t] \neq E[\Delta c_{t+1}]) > 0$. Fortunately, consistent tests for conditional moment restrictions such as (12) are available in the literature, see e.g. Dominguez and Lobato (2003). Here, we consider Cramer-von Mises (CvM) tests for the null hypothesis of lack of identification, i.e. $H_0 : E[Y_{t+1} \mid Z_t] = E[Y_{t+1}]$ a.s., where $Y_{t+1}$ denotes either $\Delta c_{t+1}$, $r_{f,t+1}$ or $r_{e,t+1}$, against the alternative of identification $H_1 : P(E[Y_{t+1} \mid Z_t] \neq E[Y_{t+1}]) > 0$. 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

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

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

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, we can approximate the asymptotic distribution of $CvM_n$ by that of

$$CvM_n^* = \frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2 n},$$

where $\hat{u}^* = (\hat{u}_1^*, ..., \hat{u}_n^*)'$, $\hat{u}_t^* = (V_t\hat{u}_t - c_n)$, $c_n = n^{-1} \sum_{t=1}^{n} V_t\hat{u}_t$, $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} \hat{u}_t^2$, and where $\{V_t\}_{t=1}^{n}$ is a sequence of independent and identically distributed (i.i.d) random variables with zero mean, unit variance, bounded support and also independent of the sequence...
\(\{Y_t, X_t, Z_t\}_{t=1}^{n+1}\). Examples of \(\{V_t\}\) sequences are i.i.d. Bernoulli variates 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 \(P(V_t = 1) = 0.5\) 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 identification 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, we 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 IV 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 we also report the values of the first-stage F test from Yogo (2004). This is useful in order to compare unconditional and conditional identification failures. The F tests suggest low linear predictability for all countries and both frequencies in consumption growth and stock returns, but high predictability of interest rates. When looking for nonlinear 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 our analysis: (i) the evidence of weak linear identification in international data found in Yogo (2004) and others can be extended to weak conditional identification when considering stock returns, and (ii) when using interest rates, it appears that \(\psi\) and \(1/\psi\) are conditionally identified for UK and USA data, but \(1/\psi\) is weakly identified by IV methods. Thus, the UK and USA quarterly data highlights the difference between linear and nonlinear predictability, and hence the difference between lack of identification by linear methods and our conditional identification assumption. For that reason, in what follows we restrict our attention to these two countries and interest rates data. It is interesting to note that our 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”.

17
## TABLE IV

### Conditional Identification Tests. Bootstrap P-values

<table>
<thead>
<tr>
<th>Country</th>
<th>Sample Period</th>
<th>$\Delta c_{t+1}$</th>
<th>$r_{f,t+1}$</th>
<th>$r_{e,t+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUL</td>
<td>1970.3-1998.4</td>
<td>0.350 (1.79)</td>
<td>0.000 (21.81)</td>
<td>0.345 (1.82)</td>
</tr>
<tr>
<td>CAN</td>
<td>1970.3-1999.1</td>
<td>0.446 (3.03)</td>
<td>0.000 (15.37)</td>
<td>0.599 (2.51)</td>
</tr>
<tr>
<td>FR</td>
<td>1970.3-1998.3</td>
<td>0.901 (0.17)</td>
<td>0.000 (38.43)</td>
<td>0.614 (3.09)</td>
</tr>
<tr>
<td>GER</td>
<td>1979.1-1998.3</td>
<td>0.541 (0.83)</td>
<td>0.005 (17.66)</td>
<td>0.820 (0.69)</td>
</tr>
<tr>
<td>ITA</td>
<td>1971.4-1998.1</td>
<td>0.678 (0.73)</td>
<td>0.000 (19.01)</td>
<td>0.819 (1.10)</td>
</tr>
<tr>
<td>JAP</td>
<td>1970.3-1998.4</td>
<td>0.501 (1.18)</td>
<td>0.000 (8.64)</td>
<td>0.344 (3.49)</td>
</tr>
<tr>
<td>NTH</td>
<td>1977.3-1998.2</td>
<td>0.599 (0.89)</td>
<td>0.000 (12.05)</td>
<td>0.946 (0.73)</td>
</tr>
<tr>
<td>SWD</td>
<td>1970.3-1999.2</td>
<td>0.622 (0.48)</td>
<td>0.000 (17.08)</td>
<td>0.151 (2.24)</td>
</tr>
<tr>
<td>SWT</td>
<td>1976.2-1998.4</td>
<td>0.832 (0.97)</td>
<td>0.005 (8.55)</td>
<td>0.975 (0.11)</td>
</tr>
<tr>
<td>UK</td>
<td>1970.3-1999.1</td>
<td>0.075 (2.52)</td>
<td>0.003 (17.04)</td>
<td>0.600 (2.62)</td>
</tr>
<tr>
<td>USA</td>
<td>1947.3-1998.4</td>
<td>0.003 (2.93)</td>
<td>0.000 (15.53)</td>
<td>0.196 (2.88)</td>
</tr>
<tr>
<td>SWD</td>
<td>1921-1994</td>
<td>0.805 (1.02)</td>
<td>0.000 (5.50)</td>
<td>0.439 (1.67)</td>
</tr>
<tr>
<td>UK</td>
<td>1921-1994</td>
<td>0.407 (1.93)</td>
<td>0.001 (4.87)</td>
<td>0.009 (4.18)</td>
</tr>
<tr>
<td>USA</td>
<td>1891-1995</td>
<td>0.959 (1.55)</td>
<td>0.058 (2.87)</td>
<td>0.439 (1.00)</td>
</tr>
</tbody>
</table>

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

We now proceed to estimate $\psi$ in (11) and $1/\psi$ in (13) using our IIV estimator and interest rates for UK and USA data. For completeness we also report results for annual data. We compare in Table V our 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. Our IIV estimates for $\psi$ with quarterly data are considerably larger than those of the TSLS and LIML estimates, and vice versa for annual data. The case of quarterly UK data 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 are conditionally identified by exploiting nonlinear dependence between interest rates and instruments. Our estimate of the EIS for quarterly UK data is 0.5, and is significantly different from zero. The TSLS estimate is much smaller (0.17) and not significant at 5%. Unlike with TSLS, we do not get contradictory results when estimating $1/\psi$, with an estimate of 1.94 that is fairly consistent with the estimate of $\psi$. For USA quarterly data, we obtain an estimate of 0.66, that although larger than the TSLS and LIML estimates, it is not
significantly different from zero. Estimates of the EIS with annual data are small and no significantly different from zero, in agreement with previous studies.

**TABLE V**

**Estimates of Interest Rates**

<table>
<thead>
<tr>
<th>Country</th>
<th>Sample Period</th>
<th>( \psi )</th>
<th>( 1/\psi )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IIV</td>
<td>TSLS</td>
<td>LIML</td>
</tr>
<tr>
<td>UK 1970.3-1999.1</td>
<td>0.50</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.13)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>USA 1947.3-1998.4</td>
<td>0.66</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.09)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>UK 1921-1994</td>
<td>0.08</td>
<td>0.26</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.12)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>USA 1891-1995</td>
<td>-0.28</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.11)</td>
<td>(0.16)</td>
</tr>
</tbody>
</table>

We now compare 95% asymptotic confidence intervals for the EIS based on the IIV estimates 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 quarterly UK data, IIV provides confidence intervals consistent with larger values for the EIS than those provided by fully robust methods. In particular, according to the CLR the EIS is not significantly different from zero, in stark contrast with our finding using IIV estimates. Our confidence intervals suggest that for USA data the EIS is not significantly different from zero at 5%. Moreover, it appears that there are serious identification problems with linear methods in the annual USA data as evidenced by the uninformative confidence intervals of the CLR test. On the contrary, our results indicate that the parameter \( \psi \) is identified, although takes small values. It should be also noted that for quarterly USA data the null hypothesis of \( \psi = 1 \) is not rejected at 5% with our methods but it is rejected with the robust procedures. This hypothesis is of great economic interest because with Epstein-Zin preferences, an investor’s optimal consumption choice is a constant fraction of wealth when the EIS is equal to 1.
Finally, we close this application by checking the exogeneity assumption in (11). To that end, we apply the test for conditional moment restrictions proposed in Escanciano (2009). The test statistics are of CvM type for testing 

\[ H_0: E[\Delta c_{t+1} - \alpha - \psi r_{f,t+1} | Z_t] = 0 \text{ a.s.}, \]

against the general nonparametric alternative 

\[ H_1: P(E[\Delta c_{t+1} - \alpha - \psi r_{f,t+1} | Z_t] \neq 0) > 0. \]

The CvM test is computed as a quadratic form in the residuals \( \hat{u}_t = (\hat{u}_1, ..., \hat{u}_n)' \), with \( \hat{u}_t = \Delta c_{t+1} - \hat{\alpha} - \hat{\psi} r_{f,t+1}, \hat{\theta}_{IV} = (\hat{\alpha}, \hat{\psi})' \), simply as

\[
CvM_n = \frac{1}{\sigma^2 n^2} \hat{u}' P P' \hat{u},
\]

where \( P = HW_{\text{ind}}, H = I_n - X(X'X)^{-1}X' \), \( I_n \) is the \( n \times n \) identity matrix, and \( W_{\text{ind}} \) is the \( n \times n \) matrix with elements \( w_{ts} = 1(Z_t \leq Z_s) \). The asymptotic distribution of \( CvM_n \) is approximated by that of the bootstrap analogue \( CvM^*_n = n^{-2} \hat{u}' P P' \hat{u}^* \), where \( \hat{u}^* = (V_1 \hat{u}_1, ..., V_n \hat{u}_n)' \), and \( \{V_t\}_{t=1}^n \) are generated from (14); see Escanciano (2009) for a formal justification of this bootstrap procedure. The conclusion from applying these consistent exogeneity tests is that the strict exogeneity assumption is not rejected for all data sets. The bootstrap p-values for quarterly data are 0.252 and 0.282 for UK and USA, respectively. For annual data the bootstrap p-values are 0.146 and 0.703, respectively. Hence, these tests validate previous inferences.

To sum, this application has shown that the new IIV estimator, and the inference procedures derived from it, can complement existing inferential procedures under weak instruments in applications of economic interest. In addition, we have shown that is possible to test for conditional identification by means of consistent tests of conditional moment restrictions. The application to quarterly UK data provides an example where the EIS is weakly identified by linear methods, but strongly identified with our methods. Making it possible the application of pointwise estimation procedures based on IIV even in situations where instruments are weak in the classical sense.
6 Conclusions and Extensions

In this article we have emphasized minimal conditions for identification when a strict exogeneity assumption holds. Attention has been paid to the fact that this minimal assumption does not preclude having weak or no unconditional identification, as defined in terms of the classical IV identification assumption. This distinction suggests the possibility of inference procedures, in particular estimation, robust to weak or no unconditional identification when conditional identification holds. We have proposed an estimator achieving this goal. The new estimator should be appealing to practitioners given its simplicity and robustness. Our results are complementary, rather than competing, to existing robust methods under weak identification for hypothesis testing and computation of confidence bands. Moreover, our results can provide a potential solution to the less explored problem of pointwise estimation under weak instruments when endogenous and exogenous variables are nonlinearly related.

Optimal IV estimators, optimal in the sense of minimum asymptotic variance, also rely on the minimal identification assumption in linear models, but as shown in our simulations, they are very sensitive to the level of conditional identification. The IIV estimator seems to outperform the optimal IV estimator for moderate and small sample sizes when conditional identification is weak.

It remains a topic of future research to investigate the behaviour under weak conditional identification of other, more robust-to-identification, estimates such as the Fuller-$k$ class; see Section 6 in Stock et al. (2002) for a survey. Note that, unlike the new estimator, the vast majority of existing robust procedures do not exploit the possible nonlinearity in the conditional mean of endogenous variables given regressors. Finite sample comparisons will be investigated in future work.

An interesting extension of our results is to the development a parallel theory to that of traditional weak instruments, as discussed in Staiger and Stock (1997), for the conditional identification assumption. In our conditional framework, the weak identification can be parametrized by a direction of local departure from zero of the conditional moment $m$, such as $m(Z_t) = b_n a(Z_t)$ for some sequence of positive numbers $b_n \to 0$, and global departures from zero can be measure with consistent tests for the conditional moment restriction $m \equiv 0$.

Finally, another important issue that we have omitted is the possibility of testing for conditional identification in linear models with multivariate covariates. The negation of the conditional identification assumption 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. In turn, testing for the latter identification involves a smaller model, so one can proceed by an inductive reasoning. This issue is an interesting topic for future research.

References


