Pivotal and robust subvector inference in structural models

Bertille Antoine *
Department of Economics, Simon Fraser University

Pascal Lavergne †
Toulouse School of Economics, Université Toulouse Capitole

July 15, 2025

Abstract

We propose a new specification test for a parametric structural model defined by conditional moments. Our procedure is robust to weak identification and heteroskedasticity of unknown form, and detects nonparametric deviations from the null of correct specification. The test statistic builds on the ICM statistics of Bierens (1982) and Antoine and Lavergne (2023), but directly accounts for heteroskedasticity of unknown form. Our procedure is omnibus and uniformly controls size irrespective of identification strength. It is also powerful irrespective of the precise form of the link between instruments and endogenous variables. In addition, our test statistic is compatible with identification-robust subvector inference without maintaining any additional (identification) assumption on the remaining parameters.

Our inference procedure and specification test are computationally-friendly and competitive with existing procedures in simulations and applications.

Keywords: Weak Identification, Specification Testing, Subvector Inference.

*Bertille Antoine acknowledges funding from The Social Sciences and Humanities Research Council of Canada.

†Pascal Lavergne acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010).

1 Introduction

We consider cross-section data observations and the linear model popular from micro-econometrics

$$y_i = Y_{2i}'\beta + X_{1i}'\gamma + u_i \qquad \mathbb{E}\left(u_i | X_{1i}, X_{2i}\right) = 0 \quad i = 1, \ldots n\,, \tag{1}$$

where Y_2 are endogenous variables, X_1 are exogenous control variables, and X_2 are exogenous instrumental variables. Over the last 30 years, it has become clear that standard asymptotic approximations may reflect poorly what is observed even for large samples when there is weak correlation between instrumental variables and endogenous explanatory variables. Alternative asymptotic frameworks have then been developed to account for potentially weak identification and tests have been proposed that deliver reliable inference about parameters of interest, see e.g. Staiger and Stock (1997), Stock and Wright (2000), Moreira (2003), Kleibergen (2002, 2005), Andrews and Cheng (2012), Andrews and Guggenberger (2019), Andrews (2016), and Andrews and Mikusheva (2016a,b). Surveys on weak identification issues include Stock et al. (2002), Dufour (2003), Hahn and Hausman (2003), and Andrews and Stock (2007). Existing inference procedures are robust to identification strength and uniformly control size, but rely on a parametric first-stage, and often on a linear projection of endogenous variables on instruments. From an empirical perspective, Dieterle and Snell (2016) have documented significant nonlinearities in first-stage regression in several applied microeconomics papers. Since practitioners typically have little prior information on the form of the relation between endogenous variables and instruments, one may consider estimating the reduced form nonparametrically, e.g. using an increasing number of approximating series. However, nonparametrically estimated instruments cannot be relied upon under weak identification, see Jun and Pinkse (2012) and Mikusheva and Sun (2020). Indeed, if identification is not strong enough, the statistical variability of a nonparametric estimator will dominate the signal we aim to estimate.

In a recent work, building on the Integrated Conditional Moment (ICM) principle originally proposed by Bierens (1982), Antoine and Lavergne (2023) develop two inference procedures that are easy to implement, robust to any identification pattern and unknown heteroskedasticiy, and that do not rely on a linear projection in the first-stage equation. In particular, they study an ICM test which tests at the same time for the value of the parameter and the specification of the model. Our present work elaborates on theirs. We propose a heteroskedastic version of their ICM statistic, labelled HICM. A key feature of our new statistic is that, under the null hypothesis $H_0 : \beta = \beta_0$, its asymptotic distribution is independent of β_0. Since building a confidence interval for β involves inverting a test statistic for H_0 a large number of times, using critical values that are independent of parameter values significantly speeds up the process. An additional advantage of this feature is that it allows to devise a *pure* specification test, which is valid - and powerful - independently of identification weakness and of the particular functional form of the reduced form that links instruments and endogenous variables. We show that our specification test is omnibus and that it uniformly controls size irrespective of identification strength. We also study its power properties along sequences of local alternatives.

Furthermore, our procedure can also be easily adapted to handle subvector inference. That is, delivering identification-robust inference on some of the parameters of interest, without maintaining any additional assumption on the parameters that are not under test: specifically, our procedure remains valid regardless of the underlying identification of the parameters that are not under test.

We illustrate the finite sample properties of our procedures in a series of simulations. When the model is correctly specified, we find that the level of our HICM-based inference procedure is well controlled and that it has significant power advantages compared to existing procedures such as that of Stock and Wright (2000) when the reduced form equation is nonlinear. It also displays competitive power for a linear reduced form. Similarly, when testing whether the model is correctly specified, our HICM-based specification test demonstrates excellent size and power properties. We also consider two empirical applications: (i) a study on the effects of population decline in Mexico on land concentration in the sixteenth century, using the data and framework of Sellars and Alix-Garcia (2018) and (ii) a study on the determinants of risk and time preferences, using the data and framework of Tanaka et al. (2010). In both cases, our procedures are competitive and provide empirically valuable inference.

Our paper is organized as follows. In section 2, we introduce our framework and our new HICM test statistic. We discuss how it can be used for powerful and identification-robust inference that is compatible with subvector inference, as well as for specification testing. The asymptotic properties of our inference procedure and of our specification test based on HICM are studied in section 3. Their finite sample properties are investigated in a series of Monte-Carlo experiments in section 4 and in two empirical applications in section 5.

2 Framework

The influence of exogenous control variables X_1 can be projected out through orthogonal projection in (1), which does not influence our reasoning, but simplifies exposition. Hence, in what follows, we consider a structural equation of the form

$$y_i = Y_{2i}'\beta + u_i \qquad \mathbb{E}\left(u_i|Z_i\right) = 0 \quad i = 1, \ldots n\,. \tag{2}$$

This is augmented by a first-stage reduced form equation for Y_2

$$Y_{2i} = \Pi(Z_i) + V_{2i} \qquad \mathbb{E}\left(V_{2i}|Z_i\right) = 0\,. \tag{3}$$

Hence,

$$y_i - Y_{2i}'\beta_0 = \Pi'(Z_i)\left(\beta - \beta_0\right) + \varepsilon_i, \qquad \text{where} \quad \varepsilon_i = u_i + V_{2i}'\left(\beta - \beta_0\right) \quad \text{and} \quad \mathbb{E}\left(\varepsilon_i|Z_i\right) = 0\,.$$

The variables Z include the instruments X_2 but also the exogenous X_1 to account for potential nonlinearities in X_1 in the function $\Pi(\cdot)$.

2.1 ICM statistic

The ICM statistic introduced by Antoine and Lavergne (2023) is a test statistic for

$$\tilde{H}_0 : \mathbb{E}\left(y - Y_2'\beta_0|Z\right) = 0 \quad \text{a.s.} \tag{4}$$

which considers at the same time $H_0 : \beta = \beta_0$ and the correct specification of the model, in the same way the Anderson and Rubin (1950) (AR hereafter) test does. A result of Bierens (1982) states that $\tilde{H}_0$ holds if and only if

$$\mathbb{E}\left[(y - Y_2'\beta_0)\exp(is'Z)\right] = 0 \quad \forall s \in \mathbb{R}^k\,. \tag{5}$$

To test this hypothesis, Bierens' Integrated Conditional Moment (ICM) statistic is

$$\int_{\mathbb{R}^k} |n^{-1/2}\sum_{j=1}^{n}\left(y_j - Y_{2j}'\beta_0\right)\exp(is'Z_j)|^2\,d\mu(s)\,, \tag{6}$$

where μ is some symmetric measure with support $\mathbb{R}^k$. Let us define

$$w(z) = \int_{\mathbb{R}^k}\exp(is'z)\,d\mu(s) = \int_{\mathbb{R}^k}\cos(s'z)\,d\mu(s)\,,$$

due to the symmetry of μ. We can then rewrite the statistic (6) as

$$\begin{aligned}
&\int_{\mathbb{R}^k} n^{-1}\sum_{j=1}^{n}\sum_{m=1}^{n}(Y_j'b_0)(Y_m'b_0)\exp(is'(Z_j - Z_m))\,d\mu(s) \\
= \;& n^{-1}\sum_{j=1}^{n}\sum_{m=1}^{n}(Y_j'b_0)(Y_m'b_0)\int_{\mathbb{R}^k}\exp(is'(Z_j - Z_m))\,d\mu(s) = b_0'Y'WYb_0\,,
\end{aligned} \tag{7}$$

where W is a matrix with generic element $n^{-1}w(Z_j - Z_m)$. The condition for μ to have support $\mathbb{R}^k$ translates into the restriction that $w(\cdot)$ should have a strictly positive Fourier transform almost everywhere. Examples include products of triangular, normal, logistic, see Johnson et al. (1995, Section 23.3), Student, including Cauchy, see Dreier and Kotz (2002), or Laplace densities. To achieve scale invariance, we recommend, as in Bierens (1982), to scale the exogenous instruments by a measure of dispersion, such as their empirical standard deviation. The role of the function $w(\cdot)$ resembles the one of the kernel in nonparametric estimation, but, in contrast, it is a *fixed user-chosen function that does not vary with the sample size*. To make this explicit, we will impose that the squared integral of $w(\cdot)$ equals one.[1]

If Z has bounded support, results from Bierens (1982) yield that $\tilde{H}_0$ holds if and only if $\mathbb{E}\left[(y - Y_2'\beta_0)\exp(s'Z)\right] = 0$ for all s in a (arbitrary) neighborhood of 0 in $\mathbb{R}^q$. Hence μ in (6) can be taken as any symmetric probability measure that contains 0 in the interior of its support. For instance, we can consider the product of uniform distributions on $[-\pi, \pi]$, so that $w(\cdot)$ is the product of sinc functions. As noted by Bierens (1982), there is no loss of generality to assume a bounded support, as his above-mentioned equivalence result equally applies to a one-to-one transformation of Z, which can be chosen with bounded image. Moreover, if it is known that

$$\mathbb{E}\left(y - Y_2'\beta_0 | Z\right) = \mathbb{E}\left(y - Y_2'\beta_0 | \Psi(Z)\right),$$

for some known dimension-reducing function $\Psi(\cdot)$, then W could be defined using this transformation instead.

The ICM principle replaces conditional moment restrictions by a continuum of unconditional moments such as (5). Other functions have been used beyond the complex exponential, see Bierens (1990) and Bierens and Ploberger (1997). Stinchcombe and White (1998) give a characterization of a large class of functions that could generate an equivalent set of unconditional moments. As detailed by Lavergne and Patilea (2013), this yields a full collection of potential estimators under strong (or semi-strong) identification, such as the ones developed by Dominguez and Lobato (2004), Antoine and Lavergne (2014), and Escanciano (2018) among others. This would also yield a collection of test statistics that could be used under weak identification, see Chen et al. (2021) for a recent instance. Here, we focus on a particular application of the ICM which is suitable for theoretical investigation and practical implementation, and we leave for future work the investigation of the relative merits of these different ICM-type tests.

[1]A more involved restriction would be to impose a similar condition on the Frobenius norm of W.

Let $\widehat{\omega}$ be a (semiparametric) estimator of $\omega = \mathbb{E}\left(\text{Var}(Y|Z)\right)$. Antoine and Lavergne (2023) ICM test statistic is defined as

$$\text{ICM}(\beta_0) = \frac{b_0' Y' W Y b_0}{b_0' \widehat{\omega} b_0} \qquad \text{with} \qquad b_0 = (1, -\beta_0')' \quad \text{and} \quad Y = \begin{bmatrix} y_1 & Y_{21}' \\ \vdots & \vdots \\ y_n & Y_{2n}' \end{bmatrix}. \tag{8}$$

It corresponds to the ICM statistic (7) with the value of the parameter set at β_0 which is standardized by an estimator of the variance of $Y_i' b_0$. It resembles the AR statistic, with W replacing P_Z, the orthogonal projection on Z. The statistic is also related to Antoine and Lavergne (2014) Weighted Minimum Distance objective function, though they chose a different normalization and only consider semi-strong identification. As apparent from its construction, ICM is designed to test the correct specification of the model together with the parameter value, as does the AR test under a linear reduced form. Since ICM equals (6) up to the positive term $b_0' \widehat{\omega} b_0$, it is non-negative, and the test rejects the null hypothesis for large positive values of the statistic.

2.2 Heteroskedasticity-robust ICM statistic

We now construct a version of ICM which is robust to heteroskedastic, so-called HICM. Accounting for unknown heteroskedasticity requires estimating the conditional variance $\Omega(Z)$ of Y given Z. One should note that weak identification does not preclude consistent estimation of these quantities. The conditional variance can be estimated parametrically if one is ready to make an assumption on its functional form. Otherwise, we can resort to nonparametric conditional variance estimation. Several consistent ones have been developed for a univariate Y, and generalize easily. To make things concrete, we focus on kernel smoothing, which is used in our simulations and application. Let

$$\overline{Y}(z) = (n b_n)^{-1} \sum_{i=1}^{n} Y_i K\left((Z_i - z)/b_n\right)$$

based on the n iid observations (Y_i, Z_i), a kernel $K(\cdot)$, and a bandwidth b_n. With $e = (1, \ldots 1)'$, let $\widehat{f}(z) = \overline{e}(z)$ and $\widehat{Y}(z) = \overline{Y}(z)/\widehat{f}(z)$. The conditional variance estimator of Y is defined as

$$\widehat{\Omega}(z) = (n b_n)^{-1} \frac{\sum_{i=1}^{n} \left(Y_i - \widehat{Y}(Z_i)\right)\left(Y_i - \widehat{Y}(Z_i)\right)' K\left((Z_i - z)/b_n\right)}{\widehat{f}(z)}. \tag{9}$$

This estimator, studied by Yin et al. (2010), is a generalization of the kernel conditional variance, and is positive definite whenever $K(\cdot)$ is positive. Note that we could equivalently consider an estimator of the uncentered moment $\mathbb{E}\,(Y'Y)$ and then avoid preliminary estimation of $\mathbb{E}\,(Y|Z)$.

Now it is clear that our null hypothesis of interest (4) also writes

$$\tilde{H}_0 : \mathbb{E}\,\left[(b_0'\Omega(Z)b_0)^{-1/2}(y - Y_2'\beta_0)|Z\right] = 0 \quad \text{a.s.} \tag{10}$$

Following our previous reasoning, we define

$$\begin{aligned}\text{HICM}(\beta_0) &= n^{-1}\sum_{j=1}^{n}\sum_{m=1}^{n}(b_0'\widehat{\Omega}(Z_j)b_0)^{-1/2}(Y_j'b_0)(b_0'\widehat{\Omega}(Z_m)b_0)^{-1/2}(Y_m'b_0)w(Z_j - Z_m)\\ &= b_0'Y'\left[(e\otimes b_0)'\widehat{\Omega}(e\otimes b_0)\right]^{-1/2}W\left[(e\otimes b_0)'\widehat{\Omega}(e\otimes b_0)\right]^{-1/2}Yb_0\,,\end{aligned} \tag{11}$$

where $\widehat{\Omega} = \text{diag}\left(\widehat{\Omega}(Z_i), i = 1,\ldots n\right)$ with $\hat{\Omega}(.)$ as in (9) and e the n-vector of ones.

Our new HICM statistic can be used to construct a confidence region for β, in the same way an ICM-based confidence region is built, see Antoine and Lavergne (2023). If the model is correctly specified with parameter β_0, then, as we will show, HICM(β_0) asymptotically follows the same distribution as $G'WG$, where $G \sim N(\mathbf{0}, \mathbf{I})$. This distribution is independent of the particular value of β_0. We can then simulate the distribution of our statistic under H_0 and recover a critical value as the $(1-\alpha)$-quantile of the distribution of $G'WG$, denoted as $c_{1-\alpha}(Z)$. The confidence set obtained by inverting the HICM test is $\{\beta_0 : \text{HICM}(\beta_0) < c_{1-\alpha}(Z)\}$. Since building such a confidence set for β involves inverting the HICM test a large number of times - testing many candidate values β_0 - the use of critical values that are independent of β_0 is computationally advantageous.

Next, we explain how to adapt our procedure to deliver subvector inference, before introducing our *pure* specification test in section 2.4.

2.3 Subvector inference

In this subsection, we return to the complete model (1) to introduce an HICM-based inference procedure compatible with subvector inference,

$$y_i = Y_{2i}'\beta_0 + X_{1i}'\gamma_0 + u_i \qquad \mathbb{E}\,(u_i|X_{1i}, X_{2i}) = 0 \quad i = 1,\ldots n\,. \tag{12}$$

We consider two cases of interest for subvector inference:

(i) inference on β_0, the entire vector of slope parameters associated with endogenous variables;

(ii) inference on some components of β_0 only, say $\beta_{0,1}$, with $\beta_0 = [\beta_{0,1}',\ \beta_{0,2}']'$.

2.3.1 Inference on β_0

We first consider the case where we are interested in delivering identification-robust inference on β_0 in model (12), and not on (β_0, γ_0). Since the exogeneity of X_1 is maintained throughout, we can obviously partial it out from the model and follow the HICM-based inference strategy introduced in section 2.2. We propose here an alternative strategy which relies on the strong identification of γ_0 and on the fact that it can be easily and reliably estimated.

Let us introduce the following notations: Z collects all the exogenous variables, $Z = (X_1', X_2')'$, while $\widetilde{b}_0 \equiv (b_0', \gamma_0')' = (1, -\beta_0', -\gamma_0')'$, and

$$\widetilde{Y} = \begin{bmatrix} y_1 & Y_{21}' & X_{11}' \\ \vdots & \vdots & \\ y_n & Y_{2n}' & X_{1n}' \end{bmatrix},$$

allow us to compute $\widetilde{Y}\widetilde{b}_0$, the vector of generic components $y_i - Y_{2i}'\beta_0 - X_{1i}'\gamma_0 = u_i$ for a given (β_0, γ_0). The associated HICM statistic thus writes

$$\mathrm{HICM}(\beta_0, \gamma_0) = \widetilde{b}_0'\widetilde{Y}'\widetilde{\Omega}_0^{-1/2} W \widetilde{\Omega}_0^{-1/2}\widetilde{Y}\widetilde{b}_0 = \widetilde{b}_0'\widetilde{Y}' W_0 \widetilde{Y}\widetilde{b}_0 \,, \tag{13}$$

where $\widetilde{\Omega}_0 = (e \otimes b_0)'\widehat{\Omega}(e \otimes b_0)$, and $W_0 = \widetilde{\Omega}_0^{-1/2} W \widetilde{\Omega}_0^{-1/2}$. Note that γ_0 has no bearing on $\widehat{\Omega}$.

Now, for a fixed β_0, there is no weak identification, and one can estimate γ_0 by minimizing the above statistic (13) (with respect to γ), which gives

$$\widehat{\gamma} = \left(X'W_0X\right)^{-1} X'W_0 Y b_0 \,, \quad \text{with} \quad X = \begin{bmatrix} X_{11}' \\ \vdots \\ X_{1n}' \end{bmatrix}.$$

It is then easy to deduce that the associated minimum value, $\min_\gamma \mathrm{HICM}(\beta_0, \gamma)$, is

$$\begin{aligned}
\mathrm{HICM}(\beta_0, \widehat{\gamma}) &= b_0' Y' \widetilde{\Omega}_0^{-1/2} \left[W - W\widetilde{\Omega}_0^{-1/2} X \left(X'\widetilde{\Omega}_0^{-1/2} W \widetilde{\Omega}_0^{-1/2} X \right)^{-1} X' \widetilde{\Omega}_0^{-1/2} W \right] \widetilde{\Omega}^{-1/2} Y b_0 \\
&= \widetilde{b}_0' \widetilde{Y}' \widetilde{\Omega}_0^{-1/2} \left[W - W\widetilde{\Omega}_0^{-1/2} X \left(X'\widetilde{\Omega}_0^{-1/2} W \widetilde{\Omega}_0^{-1/2} X \right)^{-1} X' \widetilde{\Omega}_0^{-1/2} W \right] \widetilde{\Omega}^{-1/2} \widetilde{Y}\widetilde{b}_0 \\
&= \widetilde{b}_0' \widetilde{Y}' \widetilde{\Omega}_0^{-1/2} \widetilde{W}_0 \widetilde{\Omega}^{-1/2} \widetilde{Y}\widetilde{b}_0 \,.
\end{aligned}$$

If the model is correctly specified with parameter (β_0, γ_0), then $\min_\gamma \mathrm{HICM}(\beta_0, \gamma)$ asymptotically follows the same distribution as $G'\widetilde{W}_0 G$, where $G \sim N(\mathbf{0}, \mathbf{I})$. We can then simulate the distribution of our statistic under $H_0 : \beta = \beta_0$ and recover a critical value as the $(1-\alpha)$-quantile of the distribution of $G'\widetilde{W}_0 G$, denoted as $\tilde{c}_{0,1-\alpha}(Z)$. The confidence region obtained by inverting the corresponding HICM test is then $\{\beta_0 : \mathrm{HICM}(\beta_0, \hat{\gamma}) < \tilde{c}_{0,1-\alpha}(Z)\}$.

In our simulation study (see section 4), we find that $\text{HICM}(\beta_0, \hat{\gamma})$ is less undersized and delivers more powerful inference than $\text{HICM}(\beta_0)$ introduced in subsection 2.2. The main advantage of $\text{HICM}(\beta_0)$ is computational, in the sense that its critical values do not depend on β_0: a significant advantage when β_0 is multidimensional and confidence regions are obtained by inverting the associated test over a multidimensional grid of candidates. Accordingly, in practice, we recommend using $\text{HICM}(\beta_0)$ when β_0 is multidimensional and computational efficiencies are at stake; otherwise, we recommend using $\text{HICM}(\beta_0, \hat{\gamma})$.

2.3.2 Inference on $\beta_{0,1}$

We now consider the case where we are interested in delivering identification-robust inference on some components of β_0 only, say $\beta_{0,1}$. We rewrite and partition our model accordingly as follows:

$$y_i = Y_{2i,1}'\beta_{0,1} + Y_{2i,2}'\beta_{0,2} + X_{1i}'\gamma_0 + u_i \qquad \mathbb{E}\left(u_i | X_{1i}, X_{2i}\right) = 0 \quad i = 1, \ldots n\,. \tag{14}$$

In this model, we cannot reliably estimate $(\beta_{0,2}, \gamma_0)$ for two main reasons: first, $\beta_{0,2}$ may only be weakly identified; and second, without knowing / fixing $\beta_{0,2}$, the approach highlighted in the previous subsection is infeasible when it comes to the estimation of γ_0. We resort instead to the following *hybrid* approach which combines: (i) partialing out X_1 from (14); (ii) minimizing $\text{HICM}(\beta_{0,1}, \beta_{0,2})$ with respect to $\beta_{0,2}$ for given $\beta_{0,1}$; and (iii) inverting the associated test using a conservative upper-bound.

With obvious notations, partialing out X_1 from model (14) yields:

$$y_i^{\perp} = Y_{2i,1}^{\perp\prime}\beta_{0,1} + Y_{2i,2}^{\perp\prime}\beta_{0,2} + u_i \qquad \mathbb{E}\left(u_i | X_{1i}, X_{2i}\right) = 0 \quad i = 1, \ldots n\,. \tag{15}$$

Following our derivations from section 2.2 and using similar notations, the associated HICM statistic writes

$$\text{HICM}(\beta_{0,1}, \beta_{0,2}) = b_0' Y^{\perp\prime} \left[(e \otimes b_0)'\widehat{\Omega}(e \otimes b_0)\right]^{-1/2} W \left[(e \otimes b_0)'\widehat{\Omega}(e \otimes b_0)\right]^{-1/2} Y^{\perp\prime} b_0$$

with $b_0 = (1, -\beta_{0,1}', -\beta_{0,2}')$. If the model is correctly specified with parameter β_0, then, as previously explained, $\text{HICM}(\beta_{0,1}, \beta_{0,2})$ asymptotically follows the same distribution as $G'WG$, where $G \sim N(\mathbf{0}, \mathbf{I})$. This distribution is independent of the particular value of β_0. As a result, even though $\text{HICM}(\beta_{0,1}, \beta_{0,2})$ is infeasible under $H_0 : \beta_1 = \beta_{0,1}$ (since $\beta_{0,2}$ is now unknown under the null of interest), we can still design an identification-robust inference procedure for β_1 without maintaining any additional restrictions on the identification of β_2. Our inference is now based on

$$\text{HICM}^*(\beta_{0,1}) \equiv \min_{\beta_2} \text{HICM}(\beta_{0,1}, \beta_2) \leq \text{HICM}(\beta_{0,1}, \beta_{0,2})\,,$$

and it relies on a (conservative) critical value obtained as the $(1-\alpha)$-quantile of the distribution of $G'WG$, still denoted as $c_{1-\alpha}(Z)$. The confidence set obtained by inverting the corresponding test is $\{\beta_{0,1} : \text{HICM}^*(\beta_{0,1}) < c_{1-\alpha}(Z)\}$.

2.4 Specification testing

For simplicity, we now introduce our *pure* specification test in the context of the linear model where the exogenous have been partialed out,

$$y_i = Y_{2i}'\beta + u_i \qquad \mathbb{E}\left(u_i|Z_i\right) = 0 \quad i = 1, \ldots n\,.$$

To test the correct specification of this model, we define

$$\text{HICM}^* = \min_{\beta} \text{HICM}(\beta)\,.$$

Under correct specification, there is a β_0 such that (10) holds. Since, by definition,

$$\text{HICM}^* \leq \text{HICM}(\beta_0)\,,$$

we can rely on the simulated null distribution of $\text{HICM}(\beta_0)$, which is independent of β_0, to bound the distribution of HICM^*. Our asymptotic test rejects the correct specification of the model whenever $\text{HICM}^* > c_{1-\alpha}(Z)$. From the above inequality, we have control of asymptotic size, and the test is conservative.

If the model is incorrectly specified, then there is no β_0 such that

$$\mathbb{E}\left[\Omega^{-1/2}(Z)\left(y - Y_2'\beta_0\right)\exp(is'Z)\right] = 0 \quad \forall s \in \mathbb{R}^k\,.$$

Hence, under strong identification, we expect

$$\int_{\mathbb{R}^k} |n^{-1}\sum_{j=1}^{n}(b_0'\widehat{\Omega}(Z_j)b_0)^{-1/2}\left(y_j - Y_{2j}'\beta_0\right)\exp(is'Z_j)|^2\,d\mu(s) \tag{16}$$

to be bounded away from zero uniformly in β_0, and our test to be consistent. Under weak identification, (16) should be bounded away from zero uniformly in β_0, and our specification test should have non trivial power.

3 Asymptotics

3.1 Uniform asymptotic validity

We consider the following assumptions.

Assumption A. *(i) The observations* (y_i, Y_{2i}, Z_i) *form a rowwise independent triangular array that follows (2) and (3), where the marginal distribution of* Z *remains unchanged. (ii) For some* $\delta > 0$ *and* $M' < \infty$, $\sup_z \mathbb{E}\left(\|Y\|^{2+\delta}|Z = z\right) \leq M'$ *uniformly in* n.

The assumption of a constant distribution for Z could be weakened, but is made to formalize that identification strength is related to the conditional distribution of Y given Z only. For the sake of simplicity, we will not use a double index for observations and will denote by $\{Y_1, \ldots, Y_n\}$ the independent copies from Y for a sample size n. We denote by $\mathcal{P}$ the class of distributions on which our observations lie.

Let $\mathcal{E}$ be a class of vector-valued functions $\Pi(\cdot)$ and let $N(\varepsilon, \mathcal{E}, L_2(Q))$ be the covering number of $\mathcal{E}$, that is the minimum number of $L_2(Q)$ ε-balls needed to cover $\mathcal{E}$, where an $L_2(Q)$ ε-ball around $\Pi(\cdot)$ is the set of vector functions $\left\{h \in L_2(Q) \ : \ \int \|h - \Pi\|^2 \, dQ < \varepsilon\right\}$.

Assumption B. *The conditional expectation vector* $\mathbb{E}(Y_2|Z = \cdot)$ *belongs to a class of vector functions* $\mathcal{E}$ *such that* $\forall\, \Pi(\cdot) \in \mathcal{E}$, $\|\Pi(\cdot)\| \leq F(\cdot)$ *with*

$$\lim_{M\to\infty} \sup_{\mathcal{P}} \mathbb{E}\left[F^2(Z)\mathbb{I}\left(F(Z) > M\right)\right] = 0$$

and

$$\log N\left(\varepsilon \mathbb{E}^{1/2}\left(F^2(Z)\right), \mathcal{E}, L^2(P)\right) \leq K\varepsilon^{-V} \quad \text{for some } V < 2\,,$$

for all $P \in \mathcal{P}$ *and some* K, V *independent of* P.

Andrews (1994) and van der Vaart (1994), among others, exhibit classes of smooth functions that fulfill the above conditions.

Let $\mathcal{O}$ be a class of matrix-valued functions and let $N(\varepsilon, \mathcal{O}, L_2(Q))$ be the covering number of $\mathcal{O}$, defined similarly as above.

Assumption C. *(i)* $\sup_{P\in\mathcal{P}} \Pr\left[\|\widehat{\Omega} - \Omega\| > \varepsilon\right] \to 0 \quad \forall \varepsilon > 0.$

(ii) $\Omega(\cdot)$ *belongs to a class of matrix functions* $\mathcal{O}$ *such that* $O < \underline{\lambda} \leq \inf_z \lambda_{\min}(\Omega(z)) \leq \sup_z \lambda_{\max}(\Omega(z)) \leq \overline{\lambda} < \infty$ *for all* $\Omega(\cdot) \in \mathcal{O}$ *and*

$$\log N\left(\varepsilon, \mathcal{O}, L^2(P)\right) \leq K\varepsilon^{-V} \quad \text{for some } V < 2\,,$$

for all $P \in \mathcal{P}$ *and some* K, V *independent of* P.

(iii) $\sup_{P\in\mathcal{P}} \Pr\left(\widehat{\Omega}(\cdot) \in \mathcal{O}\right) \to 1$ *as* $n \to \infty$

(iv) $\sup_{P\in\mathcal{P}} \int \|\widehat{\Omega}(Z) - \Omega(Z)\|^2 \, dP(Z) \xrightarrow{p} 0.$

This assumption entails, in particular, that conditional variance estimation does not affect the asymptotic behavior of our statistics. There is a tension between the generality of the class of functions $\mathcal{O}$ and the class of possible distributions $\mathcal{P}$. When $\Omega(\cdot)$ is of a parametric form, Assumption C will be satisfied for a large class of distributions. When $\Omega(\cdot)$ is considered nonparametric and estimated accordingly, one typically assumes that its components are smooth functions, and to prove (iii) one has to show that $\widehat{\Omega}(\cdot)$ also satisfies the same smoothness conditions with probability converging to 1. Such results have been derived, see e.g. Andrews (1995) for kernel estimators or Cattaneo and Farrell (2013) for partitioning estimators. Uniform convergence of nonparametric regression estimators (and their derivatives) generally requires the domain of the functions to be bounded and the absolutely continuous components of the distributions of the conditioning variables to have densities bounded away from zero on their support. When they are not, Andrews (1995) discusses the use of a vanishing trimming that is compatible with the stochastic equicontinuity results of Andrews (1994). Condition (iv) is dealt with in the literature on honest confidence intervals using L^2 norm, see e.g. Robins and van der Vaart (2006) and the references therein.

Assumption D. *$w(\cdot)$ is a symmetric, bounded density with $\int w^2(x)\,dx = 1$. Its Fourier transform is a density, which is positive almost everywhere, or whose support contains a neighborhood of the origin if Z is bounded.*

We denote by $c_{1-\alpha}(Z)$ the critical value of HICM obtained by the simulation-based method detailed above.[2] Let $\mathcal{P}_{\beta_0}$ be the subset of distributions in $\mathcal{P}$ such that $\beta = \beta_0$. The following result establishes that our tests control size uniformly over a large class of probability distributions.

Theorem 3.1. *Under Assumptions A, B, C, and D,*

$$\limsup_{n\to\infty} \sup_{\beta_0} \sup_{P\in\mathcal{P}_{\beta_0}} \Pr\left[\mathrm{HICM}(\beta_0) > c_{1-\alpha}(Z)\right] \leq \alpha\,.$$

Our theorem readily implies that our test is asymptotically valid whatever identification strength. Indeed, for any sequence $\Pi_n(\cdot), n \geq 1$, of functions in $\mathcal{E}$, that can decrease in norm to zero arbitrarily fast, our result yields asymptotic validity under this sequence, see e.g. van der Vaart and Wellner (2000, Chap. 2.8). The result also readily implies the uniform asymptotic validity of our specification test.

[2]We neglect the approximation error due to a finite number of simulations by assuming the number of simulations is infinite so that the critical values are exact.

Theorem 3.2. *Under Assumptions A, B, C, and D,*

$$\lim_{n\to\infty}\sup \sup_{\beta_0} \sup_{P\in\mathcal{P}_{\beta_0}} \Pr\left[\text{HICM}^* > c_{1-\alpha}(Z)\right] \leq \alpha\,.$$

3.2 Asymptotic power

We adopt here a large local alternatives setup similar to Bierens and Ploberger (1997).

Assumption E. *There exists a fixed matrix $C(\cdot)$ such that $\mathbb{E}\, C(Z)C'(Z)$ is bounded and positive definite, and either (i) $\Pi(Z) = \tilde{c}_n\, \frac{C(Z_i)}{\sqrt{n}}$ or (ii) $\Pi(Z) = C(Z_i)$.*

Condition (i) allows to study the power of our tests against weak and semi-strong identification, when considering a test of $H_0:\ \beta = \beta_1$ where $\beta_1 \neq \beta_0$, the true parameter value. Condition (ii) is the strong identification case and we consider local alternatives of the type $H_{1n}:\ \beta_{1n} = \beta_0 + \tilde{c}_n \frac{\delta}{\sqrt{n}}$, where $\delta \neq 0$ is fixed. In both cases, the object of interest is the asymptotic power of our two tests when $\tilde{c}_n$ becomes large.

Theorem 3.3. *Under Assumptions A, C, and D,*

(i) under Assumption E-(i), for any fixed $\beta_1 \neq \beta_0$,

$$\liminf_{\tilde{c}_n\to\infty} \inf_{P\in\mathcal{P}_{\beta_0}} \Pr\left[\text{HICM}(\beta_1) > c_{1-\alpha}(Z)\right] = 1\,.$$

(ii) under Assumption E-(ii), for $\beta_{1n} = \beta_0 + \tilde{c}_n \frac{\delta}{\sqrt{n}}$ and a fixed $\delta \neq 0$,

$$\liminf_{\tilde{c}_n\to\infty} \inf_{P\in\mathcal{P}_{\beta_0}} \Pr\left[\text{HICM}(\beta_{1n}) > c_{1-\alpha}(Z)\right] = 1\,.$$

Result (i) shows that, under weak identification, power is non trivial for a large enough $\tilde{c}_n$. Result (ii) implies that, under strong identification, power is non trivial under a sequence of Pitman local alternatives for $\tilde{c}_n$ large enough.

3.3 Asymptotic power of specification test

Consider the sequence of local alternatives

$$H_{1,n}:\ \min_{\beta} \int \left|\mathbb{E}\ \left[(b_0'\Omega(Z)b_0)^{-1/2}(y - Y_2'\beta_0)\exp(is'Z)\right]\right|^2\, d\mu(s) = \tilde{d}_n/\sqrt{n}\,,$$

where $\tilde{d}_n, n = 1,\ldots$ is a real sequence uniformly bounded away from zero.

Theorem 3.4. *Under Assumptions A, C, and D,*

$$\liminf_{\tilde{d}_n\to\infty} \inf_{P\in\mathcal{P}\cap H_{1,n}} \Pr\left[\text{HICM}^* > c_{1-\alpha}(Z)\right] = 1\,.$$

4 Small Sample Behavior

We generated data following the model

$$\begin{aligned} y_i &= \alpha_0 + Y_{2i}\beta_0 + \delta Z_i + \sigma(Z_i)u_i\,, \\ Y_{2i} &= \gamma_0 + \frac{c}{\sqrt{n}} f(Z_i) + \sigma(Z_i)v_{2i}\,. \end{aligned} \tag{17}$$

where c is a constant that controls the strength of the identification and Y_{2i} is univariate. The joint distribution of (u_i, v_{2i}) is a bivariate normal with mean $\mathbf{0}$, unit unconditional variances, and unconditional correlation ρ. We set $\alpha_0 = \beta_0 = \gamma_0 = 0$ and $\rho = 0.8$. We consider three different specifications for the function $f(\cdot)$: (i) a polynomial function of degree 3 proportional to $z - 2z^3/5$, (ii) a linear function, and (iii) a function compatible with first-stage group heterogeneity, see Abadie et al. (2024), proportional to $(2z_2 - 1)(z_1 - 2z_1^3/5)$. Here Z (or Z_1) is deterministic with values evenly spread between -2 and 2, and Z_2 follows a Bernoulli with probability 1/2. Also, $f(Z)$ is centered and scaled to have variance one to ensure that the different cases are comparable. We consider heteroskedasticity depending on the first component of Z of the form

$$\sigma(z) = \sqrt{\frac{3(1+z^2)}{7}}\,.$$

Finally, δ controls the degree of misspecification. When $\delta = 0$, the model that excludes Z from the structural equation is well-specified, while it is misspecified when $\delta \neq 0$.

We focus on the 10% asymptotic level tests for the correct specification of the model that excludes Z from the structural equation. In all our experiments, $w(\cdot)$ is a triangle density, and conditional covariances are estimated through kernel smoothing with Gaussian kernel and rule-of-thumb bandwidth. We consider 5,000 replications and 999 simulated values of the statistic to compute the tests' p-values.

4.1 Inference on parameters in a correctly specified model

In this subsection, we keep $\delta = 0$ to ensure that the model is always correctly specified and we focus on delivering inference on β_0. We compare the performance of three inference procedures: (i) HICM introduced in this paper; (ii) ICM introduced in Antoine and Lavergne (2023); and (iii) S introduced in Stock and Wright (2000). Our benchmark is the heteroskedastic version of the polynomial model with a degree of weakness $c = 3$ and a sample size $n = 101$. We consider three versions of S, respectively with 1, 3 or 7 instruments in the polynomial and linear models. These instruments are obtained by fitting piecewise linear functions on intervals defined by the quartiles of Z (or Z_1): e.g. the three considered instruments are $1(z \leq 0)$, $z \times 1(z \leq 0)$, and z. For the group heterogeneity model, we implement S based on a reduced form with 3 instruments, namely the continuous instrument Z_1, the discrete one Z_2, and an interaction term. We then consider increasing the number of instruments to 7 and 15. We construct these instruments as piecewise linear and interaction terms on intervals defined by the quartiles of Z_1. E.g. the seven considered instruments are $1(z_1 \leq 0)$, $z_1 \times 1(z_1 \leq 0)$, z_1, $z_2 \times 1(z_1 \leq 0)$, z_2, $z_2 \times z_1 \times 1(z_1 \leq 0)$, $z_2 \times z_1$.

In Figures 1 and 2, we present the power curves of these three inference procedures when testing the null hypothesis that $H_0 : \beta = \bar{\beta}$ with $\bar{\beta} \in [-1.5, 1.5]$ for various specifications; recall that the true unknown value is $\beta_0 = 0$. The empirical sizes are reported in Table 5.

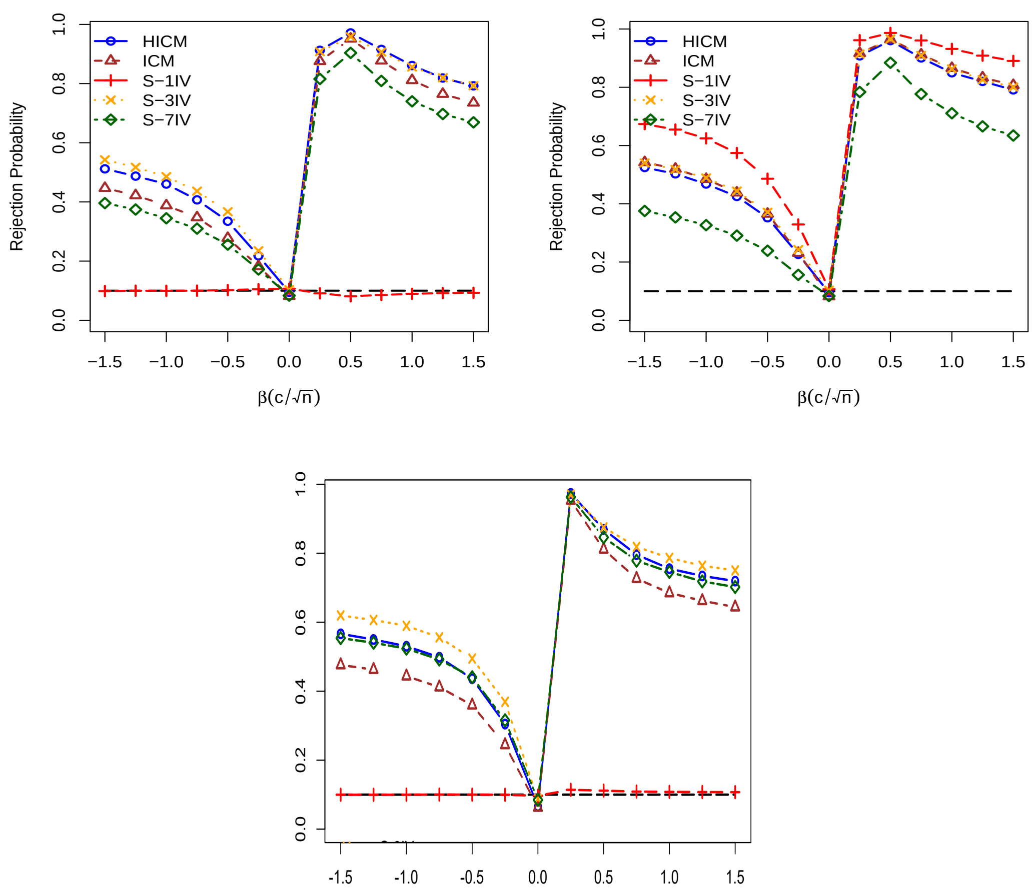

Figure 1: Power curves for HICM (blue circle dash), ICM (brown triangle dash), and S with 1 IV (red + solid), with 3 IV (orange x dot), and with 7 IV (green lozenge dot) for the polynomial and the linear models with (top left and right, respectively), and for the polynomial model with a sample size of 401 (bottom).

Throughout all specifications, size is controlled by the three procedures. Notably, HICM and ICM tend to be more conservative than S. That being said, when it comes to power, HICM - and, to some extent, ICM - display the best performance. In particular, in the benchmark model, S with 1 IV does not have any power to reject the model, even for large values of β whereas HICM and ICM do. Increasing the sample size does not help improve the power of S. What helps is an increase of the number of instruments[3]: e.g. with 3 instruments, the power of S is slightly above that of HICM. However, when the number of instruments is *too large* the power of S worsens: e.g. with 7 instruments, the power of S is slightly below that of HICM.

With a linear model, S with 1 IV is the most powerful procedure - as expected. Interestingly, HICM and ICM have power properties that are not too different and comparable to S with 3 IV; also, both dominate S with 7 IV.

With a model compatible with group heterogeneity, S with 1 IV does not have any power. S with 7 IV slightly dominates HICM and ICM, while the power of S with 15 IV falls below that of HICM and ICM.

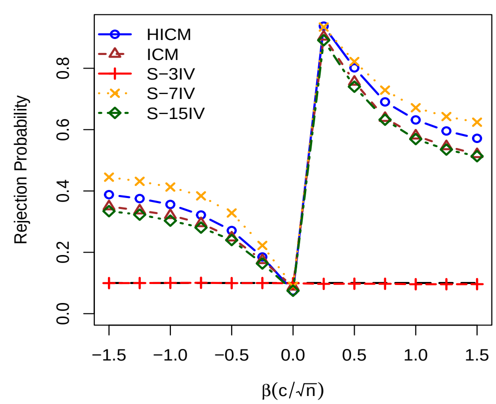

Figure 2: Power curves for HICM (blue circle dash), ICM (brown triangle dash), and S with 3 IV (red + solid), with 7 IV (orange x dot), and with 15 IV (green lozenge dot) for the case with first-stage group heterogeneity with a sample size of 201.

Overall, HICM (and ICM) perform extremely well. And both have the advantage that their power properties do not depend on the number of chosen instruments. This is in sharp contrast with S: as can be seen from our experiments, it may lack power when too few instruments are used; at the same time, its power properties worsen if too many

[3]It is important to mention that neither HICM nor ICM depend on the number of instruments - or moments - which are derived from the conditional mean independence.

instruments are used. Since it is never clear how to choose the number of instruments in practice, our HICM inference procedure appears particularly attractive.

4.2 Specification testing

In this subsection, we consider a misspecified model with $\delta \neq 0$ and our goal is to test the null hypothesis that the model is correctly specified.

We compare our specification test based on the minimum of HICM (hereafter HICM-min) with two other specification tests: the jackknife T-specification test proposed by Chao et al. (2014) (hereafter Jack-T) and the J-test of overidentification based on the continuously updated GMM with conservative critical values (hereafter J-CUE). Both procedures rely on unconditional moments obtained from a given set of instruments.

The Jack-T test is based on: (i) estimating β_0 by HFULL, a heteroskedasticity-robust version of the Fuller estimator proposed by Hausman et al. (2012); and (ii) considering a jackknife version of the overidentification statistic, based on the objective function of the JIVE2 estimator proposed by Angrist et al. (1999). The associated test statistic is asymptotically distributed as a chi-square with degrees of freedom equal to the degree of overidentification when the number of instruments increases with the sample size, or when it is fixed under maintained homoskedasticity. This test is valid under the framework of *many weak instruments* - but, strictly speaking, maybe not under the traditional framework of weak identification considered here.

The J-CUE test is based on the objective function of the continuously updated GMM with critical values obtained from the chi-square distribution with degrees of freedom equal to the number of instruments[4]. This test is expected to be conservative, but it remains valid under arbitrary weak identification.

In Figure 3, we present the power curves of these three competing procedures for our three underlying DGPs: (i) the polynomial model with a sample size of 101 (top left); (ii) the linear model with a sample size of 101 (top right); and (iii) the group heterogeneity model with a sample size of 201 (bottom). We consider three versions of Jack-T and J-CUE, respectively with 3, 7 or 11 instruments for models (i) and (ii) and with 3, 7, or 15 instruments with model (iii); see section 4.1 for details. The empirical sizes are reported in Table 6.

Throughout all our designs, our specification test HICM-min always controls size (when

[4]Regardless of the identification properties of the parameters, the CUE objective function is asymptotically upper-bounded by a random variable distributed as a chi-square with degrees of freedom equal to the number of instruments.

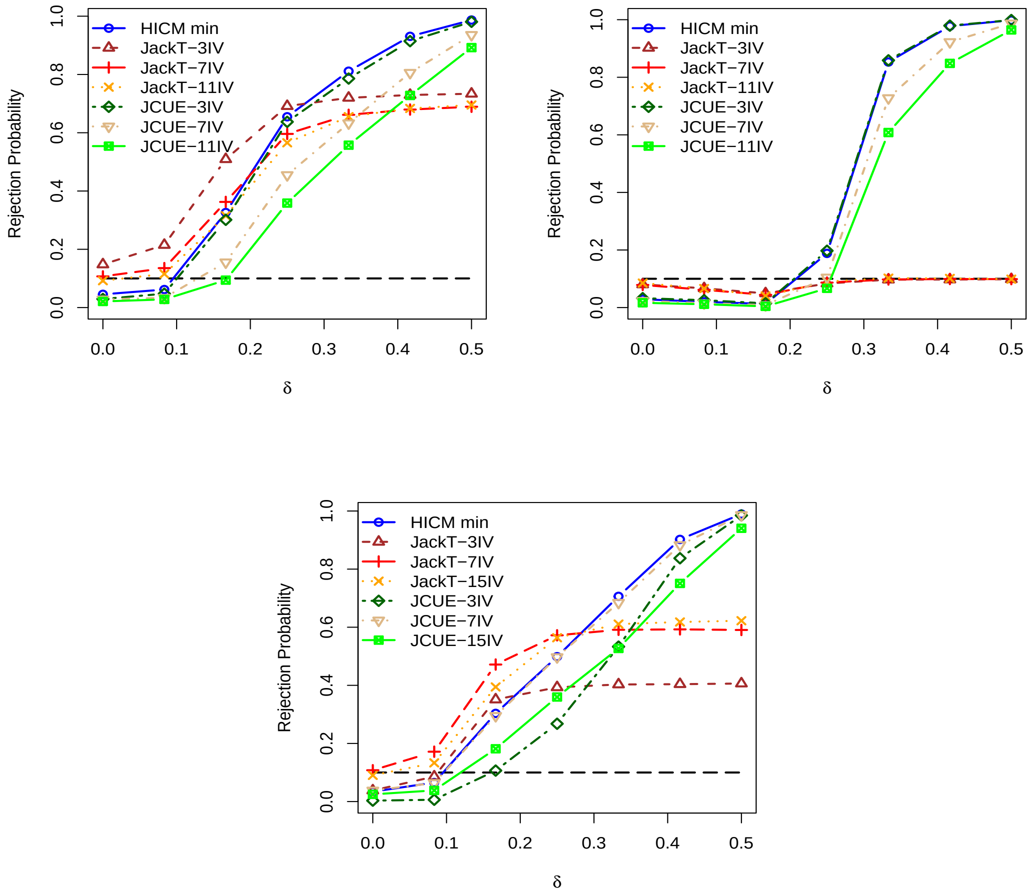

Figure 3: Power curves for (i) polynomial Model (n=101) (ii) linear model (n=101) (iii) group heterogeneity model (n=201).

$\delta = 0$), and it also consistently displays excellent power properties - even in the linear model. As expected, the properties of both Jack-T and J-CUE are affected by the number of underlying instruments. While J-CUE always controls size, it is not the case with Jack-T (see e.g. the polynomial model with 3 instruments). Interestingly, increasing the number of instruments does not always yield improved power properties, either for Jack-T or J-CUE. It is worth to note that Jack-T does not have any power in the linear model, and that it is not consistent in the other cases, regardless of the number of instruments. As previously mentioned, we suspect that this last point may be related to the preliminary estimation of the model with HFULL which is not robust to weak identification. In our last experiment, we consider a polynomial model that is more strongly identified with $c = 7$ (instead of $c = 3$), while everything else remains the same. The associated power curves are presented side-by-side in Figure 4: (i) on the left hand-side with $c = 3$; (ii) on the right hand-side with $c = 7$. Jack-T is now consistent, but it remains oversized with 3 instruments.

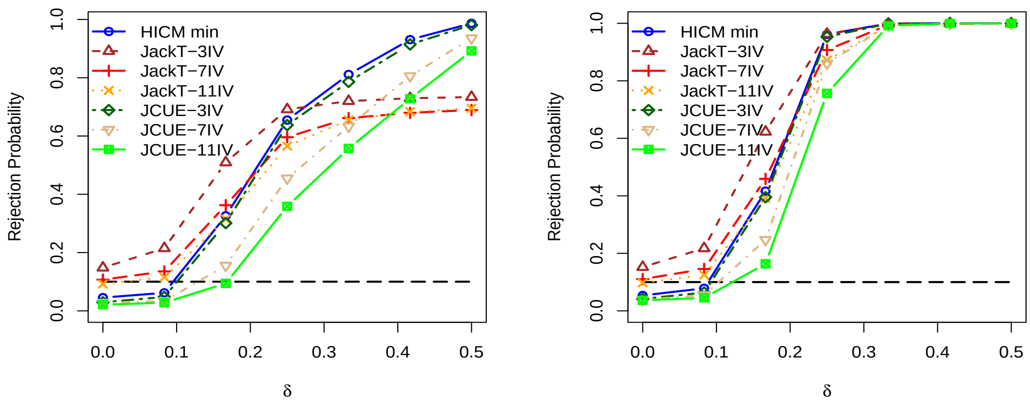

Figure 4: Power curves for polynomial model (n=101) with (i) $c = 3$ (left) and (ii) $c = 7$ (right).

Overall, our proposed specification test based on HICM-min performs extremely well. It always controls size and displays competitive power properties regardless of the identification strength and the particular functional form that links instruments and endogenous variables.

5 Empirical illustrations

5.1 Demographic collapse in 16th-century Mexico

We revisit Antoine and Lavergne (2023)'s empirical application which extends some of the results presented in Sellars and Alix-Garcia (2018). They study the impact of a large population collapse in 16th-century Mexico on land institutions by adopting an instrumental-variables empirical strategy based on the characteristics of a massive epidemic in the mid-1570s which is believed to have been caused by a rodent-transmitted pathogen that emerged after several years of drought were followed by a period of above-average rainfall. Sellars and Alix-Garcia (2018) use proxies for these climate conditions as their three excluded instrumental variables: (i) drought, the sum of the 2 lowest consecutive PDSI[5] values between 1570 and 1575 (more negative numbers indicate severe and prolonged drought), (ii) rainfall, the maximum PDSI between 1576 and 1580 (as a measure of excess rainfall),

[5]The Palmer Drought Severity Index (PDSI) is a normalized measure of soil moisture that captures deviations from typical conditions at a given location.

and (iii) gap, the difference between the minimum PDSI between 1570 and 1575 and the maximum between 1576 and 1580.

Using the data constructed by Sellars and Alix-Garcia (2018), we estimate the short-term effects of the population collapse[6], that is:

$$y_i = \beta_0 + \beta_1 Y_{2i} + \gamma' X_{1i} + u_i \,, \qquad E(u_i|X_{1i}, X_{2i}) = 0 \,, \tag{18}$$

where y_i is the inverse hyperbolic sine of the percent rural population living in hacienda communities in 1900, Y_{2i} is the population decline in municipality measured as the log ratio of 1650 and 1570 density, X_{2i} represents the vector of the 3 climate instruments, and X_{1i} is a vector of control variables of geographic features related to population and agriculture[7]. We report confidence regions obtained with three inference procedures that simulatenously test the null and the validity of the model: (i) HICM introduced in this paper; (ii) ICM introduced in Antoine and Lavergne (2023); and (iii) S introduced in Stock and Wright (2000). When using all three climate instruments, the model is rejected, even when adding regional dummies. When using the 2 most reliable instruments (gap and drought), the model is once again rejected, unless regional dummies are added to the set of controls. In Table 1, we report confidence regions obtained with either one or 2 climate instruments (gap and/or drought) including the full set of full set of 12 control variables from Sellars and Alix-Garcia (2018) and regional dummies. When using both gap and drought, all three procedures report a significant and negative effect of the population decline on the percent rural population living in hacienda - as expected: this means that a decrease in the ratio of 1650 to 1570 density increases the likelihood of having more large estates per area in 1900. The confidence set obtained with HICM is the narrowest: in addition, it is fully enclosed in the one obtained with S and that obtained with ICM.

Computationally, the advantage of HICM is clear. Since its critical values do not depend on the tested value, they only need to be computed once for the entire grid of candidates. Accordingly - and similar to S - the associated confidence set can be obtained much faster than that obtained with ICM: under 50 seconds for HICM, and over 52 minutes with ICM when considering a grid of 2,500 candidate points! It is also important to mention that, unlike S, the critical values of HICM are not tied to the number of instruments, but rather to the number of unknown parameters.

[6]As explained in details in Sellars and Alix-Garcia (2018), the sharp decline in population lowered the costs and increased the benefits of acquiring land from indigenous villages in many areas.

[7]This specification corresponds to Column 6 in Table 2 in Sellars and Alix-Garcia (2018). It includes their full set of 12 control variables, the standard deviation of PDSI, a measure of maize productivity, various measures of elevation and slope, as well as the log of tributary density in 1570 and governorship-level fixed effects. See also Column 1 in Table 2 in Antoine and Lavergne (2023)

	2 climate IV (drought, gap)	**Computational time** (in sec.)	**1 climate IV** (gap)	**Computational time** (in sec.)
HICM	[-1.256, -0.834]	46.66	[-3.837, 0.444]	95.39
ICM	[-2.209, -0.506]	3,140.22	[-4.395, 0.626]	6,527.91
S	[-1.369, -0.682]	46.66	[-2.193, -0.636]	95.39
F-stat	145.25		55.8	
Adj. R^2	0.22		0.05	

Table 1: 95% Confidence Intervals for the population collapse, using either 1 or 2 climate instruments over the full sample of size equal to 1030. All specifications include the full set of 12 control variables as well as regional dummies. Confidence regions are obtained using a grid of 2,500 evenly spread points over $[-2.5, 0]$ (with 2 IV) and 5,100 evenly spread points over $[-4.4, 0.7]$ (with 1 IV).

Sellars and Alix-Garcia (2018) - and Antoine and Lavergne (2014) - assume throughout a linear structural equation; see equation (18) above. To relax the assumption that the effect of the population collapse, Y_{2i}, on land tenure is linear[8], we consider instead the following nonlinear specifications:

$$S1: \quad y_i = b_0 + b_1 Y_{2i} + b_2 Y_{2i} \times I_{Y_{2i}<-2} + \gamma_b' X_{1i} + u_{b,i}\,, \qquad E(u_{b,i}|X_{1i}, X_{2i}) = 0 \quad (19)$$

$$S2: \quad y_i = c_0 + c_1 Y_{2i} + c_2 Y_{2i}^3 + \gamma_c' X_{1i} + u_{c,i}\,, \qquad E(u_{c,i}|X_{1i}, X_{2i}) = 0 \quad (20)$$

In Figure 5, we report the confidence regions for the (nonlinear) effect of the population collapse obtained with HICM and S for both specifications using 2 climate instruments over the full sample. We include the full set of 12 control variables as well as regional dummies as previously discussed. Bivariate confidence regions are obtained (jointly) using a 2-dimensional grid of 401 evenly spread points over $[-1.75, -0.5] \times [-0.5, 0.5]$ (for specification S1) and over $[-2, 2] \times [-0.5, 0.5]$ (for specification S2). For both specifications, the HICM confidence regions are small and bounded, while those obtained with S are unbounded.

In Table 2, we report the 95% and 90% confidence intervals obtained with HICM by projecting the bivariate confidence regions. They reveal some nonlinearities in the effect of the population collapse on land tenure for both specifications. Specifically, with the estimation of specification S1 (19), we find that the first parameter b_1 is negative (at 95%) while the second parameter b_2 is positive (at 90%) and much smaller in magnitude. Overall, the effect of the population collapse at the mean remains negative and significant, and very much in line with previously reported results; see Table 3. Our analysis suggests that more negative values of the population collapse - that is, more severe population declines - are

[8]We thank I. Andrews for this suggestion.

Specification	S1 (Indicator)			S2 (Cubic)	
Level	95%	90%		95%	90%
b_1	[-1.68, -0.73]	[-1.26, -1.09]	c_1	[-1.78, 1.61]	[-1.23, 0.80]
b_2	[-0.83, 1.62]	[0.01, 0.41]	c_2	[-0.28, 0.05]	[-0.20, -0.00]

Table 2: 95% and 90% Confidence Intervals for the (nonlinear) effect of the population collapse using 2 climate instruments over the full sample of size equal to 1030. All specifications include the full set of 12 control variables as well as regional dummies. Confidence intervals are computed by projection from the bivariate confidence regions obtained (jointly) using a 2-dimensional grid of 401 evenly spread points over $[-1.75, -0.5] \times [-0.5, 0.5]$ (for specification S1) and over $[-2, 2] \times [-0.5, 0.5]$ (for specification S2).

Specification	S0 (Linear)	S1 (Indicator)	S2 (Cubic)
Mean/median $Y_2 = -1.4$	[-1.256, -0.834]	[-1.68, -0.73]	[-2.530, 1.744]
1st-decile $Y_2 = -2.37$	[-1.256, -0.834]	[-2.51, 0.89]	[-4.787, 2.147]

Table 3: 95% confidence Intervals for the marginal effect of a one-unit increase in Y_2, the population decline, using three specifications, S0 (linear), S1 (indicator), and S2 (cubic). We report the marginal effect computed at the mean/median and at the 1st-decile. We use 2 climate instruments over the full sample of size equal to 1030 and all specifications include the full set of 12 control variables as well as regional dummies.

not driving the results as they seem to be associated with a slightly smaller (negative) effect. With the estimation of specification S2 (20), we find that the first parameter c_1 is not significantly different from 0, while the second parameter c_3 is negative (at 90%) and small in magnitude.

In Table 3, we report the marginal effects of a unit increase in Y_2, the population decline, at the mean/median value of Y_2 (equal to -1.4) as well as at the first-decile of Y_2 (equal to -2.37). For both nonlinear specifications, the confidence interval obtained for the marginal effect at the mean is much narrower than - and fully enclosed within - that obtained at the fist-decile. Interestingly, the confidence intervals obtained with specification S2 are the widest: they fully contain those obtained with specification S1, which contain those obtained with the linear specification.

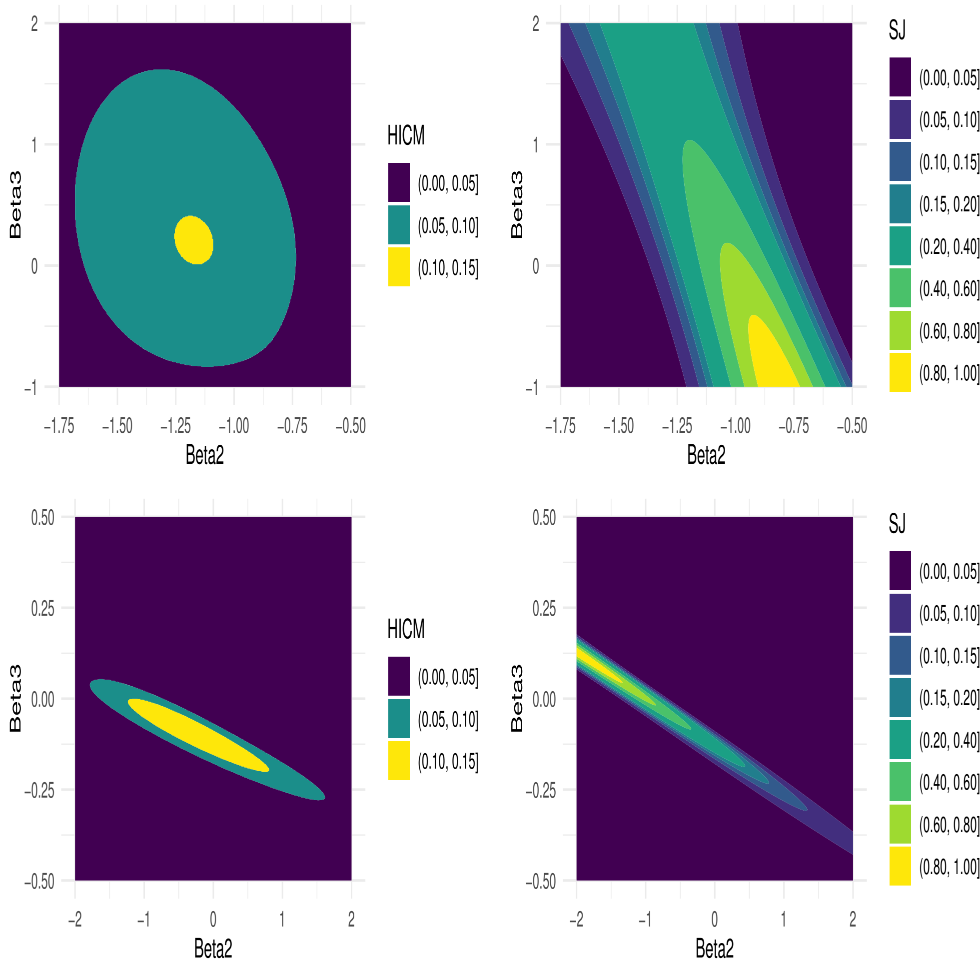

Figure 5: Confidence regions for the (nonlinear) effect of the population collapse obtained with HICM (left column) and S (right column). We estimate specification 1 (top row) and specification 2 (bottom row) using 2 climate instruments over the full sample of size equal to 1030. We include the full set of 12 control variables as well as regional dummies. Bivariate confidence regions obtained (jointly) using a 2-dimensional grid of 401 evenly spread points over $[-1.75, -0.5] \times [-0.5, 0.5]$ (for specification S1) and over $[-2, 2] \times [-0.5, 0.5]$ (for specification S2).

5.2 Determinants of risk and time preferences

We revisit Guggenberger et al. (2019)'s empirical illustration which extends some of the results presented in Tanaka et al. (2010). Using experimental data collected from Vietnamese villages, they estimate linear IV regressions to study the determinants of risk and time preferences. Their dependent variable y is the curvature of the utility function, and they consider two specifications with the same exogenous covariates X_1 (Chinese, Age, Gender, Education, Distance to market and South), the same excluded exogenous variables used as instruments X_2 (Rainfall and Head-of-household-can't-work), and either one or two endogenous variables Y_2: the first specification contains Income as the single endogenous variable, while the second one relies instead on a decomposition of income into two components, the village mean's income and the relative income within the village.

In Table 4, we report the 95% confidence intervals of the slope of the income variables for both specifications obtained with four inference procedures: (i) HICM after partialing out the exogenous controls (as in section 2.2), (ii) subvector-HICM (as in section 2.3), (iii) S introduced in Stock and Wright (2000), and (iv) conditional Anderson-Rubin introduced by Guggenberger et al. (2019). For the first specification, all four procedures report a small and insignificant effect of income - though all intervals contain 0, they are asymmetric and suggest a positive effect of income. Subvector-HICM delivers the narrowest confidence interval which is fully enclosed in those obtained with S and with HICM(β_0). This example also illustrates the efficiency gains obtained by implementing subvector-HICM rather than HICM(β_0).

	Specification #1	Specification #2
HICM(β_0)	[-0.021, 0.101]	
HICM($\beta_0, \hat{\gamma}$)	[-0.005, 0.022]	
S	[-0.005, 0.033]	
Conditional AR	[-0.002, 0.044]	

Table 4: 95% Confidence Intervals for the slope of the endogenous variables measuring income in both specification, using 2 instruments and 6 included control variables. We report confidence intervals obtained by: (i) HICM(β_0), HICM after partialing out the controls; (ii) HICM($\beta_0, \hat{\gamma}$), subvector-HICM; (iii) S; and (iv) Conditional AR. Confidence intervals are obtained using a grid of 2,000 evenly spread points over [-1,1] for all procedures except Conditional AR.

To be completed

References

Abadie, A., Gu, J., and Shen, S. (2024). Instrumental variable estimation with first-stage heterogeneity. *Journal of Econometrics*, 240(2):105425.

Anderson, T. W. and Rubin, H. (1950). The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, 21(4):570–582.

Andrews, D. W. K. (1994). Empirical Process Methods in Econometrics. In *Handbook of Econometrics*, volume 4, pages 2247 – 2294. Elsevier.

Andrews, D. W. K. (1995). Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, 11(03):560.

Andrews, D. W. K. and Cheng, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211.

Andrews, D. W. K. and Guggenberger, P. (2019). Identification and singularity-robust inference for moment condition models. *Quant. Econ.*, 10(4):1703–1746.

Andrews, D. W. K. and Stock, J. H. (2007). Inference with weak instruments. In *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, volume Volume 3 of *Econometric Society Monograph Series*, chapter 8. Cambridge University Press.

Andrews, I. (2016). Conditional linear combination tests for weakly identified models. *Econometrica*, 84(6):2155–2182.

Andrews, I. and Mikusheva, A. (2016a). Conditional Inference With a Functional Nuisance Parameter. *Econometrica*, 84(4):1571–1612.

Andrews, I. and Mikusheva, A. (2016b). A Geometric Approach to Nonlinear Econometric Models. *Econometrica*, 84(3):1249–1264.

Angrist, J. D., Imbens, G. W., and Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14:57–67.

Antoine, B. and Lavergne, P. (2014). Conditional Moment Models under Semi-Strong Identification. *J. Econometrics*, 182(3):59–69.

Antoine, B. and Lavergne, P. (2023). Identification-robust nonparametric inference in a linear IV model. *Journal of Econometrics*, 235(1):1–24.

Bierens, H. (1982). Consistent Model Specification Tests. *J. Econometrics*, 20:105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58(6):1443–1458.

Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65(5):1129–1151.

Cattaneo, M. D. and Farrell, M. H. (2013). Optimal convergence rates, Bahadur representation, and asymptotic normality of partitioning estimators. *J. Econometrics*, 174(2):127–143.

Chao, J. C., Hausman, J. A., Newey, W. K., Swanson, N. R., and Woutersen, T. (2014). Testing overidentifying restrictions with many instruments and heteroskedasticity. *Journal of Econometrics*, 178:15–21.

Chen, X., Lee, S., and Seo, M. H. (2021). Powerful Inference. *arXiv:2008.11140 [econ, math, stat]*. arXiv: 2008.11140.

Dieterle, S. G. and Snell, A. (2016). A Simple Diagnostic to Investigate Instrument Validity and Heterogeneous Effects When Using a Single Instrument. *Labour Econ*, 42(C):76–86.

Dominguez, M. A. and Lobato, I. N. (2004). Consistent Estimation of Models Defined by Conditional Moment Restrictions. *Econometrica*, 72(5):1601–1615.

Dreier, I. and Kotz, S. (2002). A note on the characteristic function of the t-distribution. *Statistics & Probability Letters*, 57(3):221 – 224.

Dufour, J.-M. (2003). Identification, Weak Instruments, and Statistical Inference in Econometrics. *Can J Economics*, 36:767–808.

Escanciano, J. C. (2018). A simple and robust estimator for linear regression models with strictly exogenous instruments. *Econom. J.*, 21(1):36–54.

Guggenberger, P., Kleibergen, F., and Mavroeidis, S. (2019). A more powerful subvector anderson-rubin test in linear variables regression. *Quantitative Economics*, 10:487–526.

Hahn, J. and Hausman, J. (2003). Weak Instruments: Diagnosis and Cures in Empirical Economics. *Am. Econ. Rev.*, 93:118–125.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., and Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quant. Econ.*, 3(2):211–255.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2 of *Wiley series in probability and mathematical statistics: Applied probability and statistics*. Wiley & Sons.

Jun, S. J. and Pinkse, J. (2012). Testing under weak identification with conditional moment restrictions. *Econometric Theory*, 28(6):1229–1282.

Kasy, M. (2018). Uniformity and the Delta Method. *J. Econom. Methods*, 8(1).

Kleibergen, F. (2002). Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, 70(5):1781–1803.

Kleibergen, F. (2005). Testing Parameters in GMM Without Assuming that They Are Identified. *Econometrica*, 73:1103–1123.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer-Verlag New York.

Lavergne, P. and Patilea, V. (2013). Smooth Minimum Distance Estimation and Testing with Conditional Estimating Equations: Uniform in Bandwidth Theory. *J. Econometrics*, 177(1):47–59.

Mikusheva, A. and Sun, L. (2020). Inference with Many Weak Instruments. Working paper, MIT.

Moreira, M. J. (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, 71:1027–1048.

Robins, J. and van der Vaart, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.*, 34(1):229–253.

Sellars, E. A. and Alix-Garcia, J. (2018). Labor scarcity, land tenure, and historical legacy: Evidence from Mexico. *J Dev Econ*, 135:504–516.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65:557–586.

Stinchcombe, M. B. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 14(3):295–325.

Stock, J. H. and Wright, J. H. (2000). GMM with Weak Identification. *Econometrica*, 68:1055–1096.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, 20:518–529.

Tanaka, T., Camerer, C., and Ngyuen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from vietnam. *American Economic Review*, 100:557–571.

van der Vaart, A. (1994). Bracketing Smooth Functions. *Stoch Proc Appl*, 52(1):93–105.

van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.

Yin, J., Geng, Z., Li, R., and Wang, H. (2010). Nonparametric covariance model. *Statist. Sinica*, 20:469–479.

6 Proofs

6.1 Proof of Theorem 3.1

To simplify exposition, we consider the case where Ω is known and the statistic is based on $S = Yb_0\,(b_0'\Omega b_0)^{-1/2}$. It is easy to adapt our reasoning to account for a consistent estimator of Ω using Assumption C-(iv). However, we do not assume that the conditional variance $\Omega(\cdot)$ is known.

6.1.1 Uniform Convergence of Processes

The class of functions $\left\{s'Z, s \in \mathbb{R}^k\right\}$ has Vapnik-Červonenkis dimension $k+2$ and thus has bounded uniform entropy integral (BUEI). Since the functions $t \to \cos(t)$ and $t \to \sin(t)$ are bounded Lipschitz with derivatives bounded by 1, the class $\left\{\cos(s'Z), \sin(s'Z), s \in \mathbb{R}^k\right\}$ is BUEI, see Kosorok (2008, Lemma 9.13).

By Assumption B, the class $\mathcal{E}$ is BUEI. From Kosorok (2008, Theorem 9.15), the class $\left\{\Pi(Z)\cos(s'Z), \Pi(Z)\sin(s'Z), \Pi(\cdot) \in \mathcal{E}, s \in \mathbb{R}^k\right\}$ is BUEI, and from van der Vaart and Wellner (2000, Lemma 2.8.3)

$$\left(\begin{array}{c} n^{-1/2}\sum_{i=1}^n \left[\mathbb{E}\left(Y_i|Z_i\right)\cos(s'Z_i) - \mathbb{E}\left(Y\cos(s'Z)\right)\right] \\ n^{-1/2}\sum_{i=1}^n \left[\mathbb{E}\left(Y_i|Z_i\right)\sin(s'Z_i) - \mathbb{E}\left(Y\sin(s'Z)\right)\right] \end{array}\right) \rightsquigarrow \left(\begin{array}{c} \mathbb{G}_1(s) \\ \mathbb{G}_2(s) \end{array}\right),$$

uniformly in $P \in \mathcal{P}$ where $(\mathbb{G}_1'(\cdot), \mathbb{G}_2'(\cdot))$ is a vector Gaussian process with mean $\mathbf{0}$. Formally weak convergence uniform in P means that

$$\sup_{P\in\mathcal{P}} d_{BL}(\mathbb{G}_n, \mathbb{G}) \to 0 \quad \text{where} \quad d_{BL}(\mathbb{G}_n, \mathbb{G}) = \sup_{f\in BL_1} \left|\mathbb{E} f\left(\mathbb{G}_n\right) - \mathbb{E} f\left(\mathbb{G}\right)\right|$$

is the bounded Lipschitz metric, that is BL_1 is the set of real functions bounded by 1 and whose Lipschitz constant is bounded by 1. This implies that

$$n^{-1/2}\sum_{i=1}^n \left[\mathbb{E}\left(Y_i|Z_i\right)\exp(is'Z_i) - \mathbb{E}\left(Y\exp(is'Z)\right)\right] \rightsquigarrow \mathbb{G}_1(s) + i\ \mathbb{G}_2(s) \tag{21}$$

Since $\mathbb{E}\|Y\|^{2+\delta} < \infty$, and because $\mathcal{E}$ is BUEI,

$$n^{-1/2}\sum_{i=1}^n \left(Y_i - \mathbb{E}\left(Y_i|Z_i\right)\right)\exp(is'Z_i) \rightsquigarrow \mathbb{G}_3(s) + i\ \mathbb{G}_4(s) \tag{22}$$

Since $\Omega(\cdot)$ is a variance matrix with uniformly bounded elements, the functions $a'\Omega(\cdot)b$ for $\|a\|, \|b\| \leq M$, and $\Omega \in \mathcal{O}$ satisfies

$$|a'\Omega_1(\cdot)b - a'\Omega_2(\cdot)b| \leq \|a\|\|b\|\|\Omega_1 - \Omega_2\| \leq M^2\|\Omega_1 - \Omega_2\|.$$

From Assumption C and Kosorok (2008, Lemma 9.13), these functions forms a BUEI class. Consider now the class of functions $\mathcal{B} = \{a'\Omega(\cdot)b/b'\Omega(\cdot)b, \|a\|, \|b\| \leq M, \Omega \in \mathcal{O}\}$. Since the function $\phi(f, g) = f/g$ is Lipschitz for f, g uniformly bounded and g uniformly bounded away from zero, $\mathcal{B}$ is a BUEI class. Gathering results, for $B \in \mathcal{B}$

$$\mathbb{G}_n(B, s) = n^{-1/2}\sum_{i=1}^n B(Z_i)\left(Y_i - \mathbb{E}\left(Y_i|Z_i\right)\right)\exp(is'Z_i) \rightsquigarrow \mathbb{G}(B, s), \tag{23}$$

converges uniformly in $P \in \mathcal{P}$ to a centered Gaussian vector process. The joint uniform convergence of the processes in (21)–(23) follows.

Now let us show that replacing $\Omega(\cdot)$ by its estimator, or replacing $B(\cdot) = a'\Omega(\cdot)b/b'\Omega(\cdot)b$ by $\widehat{B}(\cdot) = a'\widehat{\Omega}(\cdot)b/b'\widehat{\Omega}(\cdot)b$, does not change the uniform weak limit of the process. From Assumption C-(iii) and (iv), it is sufficient to show that

$$\sup_{P\in\mathcal{P}} \Pr\left[\sup_{m\geq n}\sup_{s} \|\mathbb{G}_m(\widehat{B}_m, s) - \mathbb{G}_m(B, s)\|_{\mathcal{B}} > \varepsilon\right] \to 0 \qquad \forall\varepsilon > 0.$$

This follows as $\mathbb{G}_n(B, s)$ is asymptotically equicontinuous uniformly in P, see van der Vaart and Wellner (2000, Theorem 2.8.2).

6.1.2 Notations and Preliminary Results

For vector complex-valued functions $h_1(s)$ and $h_2(s)$, define the scalar product

$$\langle h_1, h_2\rangle = \frac{1}{2}\left(\int \left(\overline{h}_1'(s)h_2(s) + h_1'(s)\overline{h}_2(s)\right) d\mu(s)\right)$$

and the norm $\|h_1\| = \langle h_1, h_1\rangle^{1/2}$. Denote

$$h_{\beta_0,S}(s) \equiv n^{-1/2}\sum_{i=1}^{n} S_i \exp(is'Z_i)\,,$$

and note that $\|h_{\beta_0,S}\|^2 = S'WS$, so that we can write $\mathrm{ICM}(\beta_0) = \mathrm{ICM}(h_{\beta_0,S}) = \|h_{\beta_0,S}\|^2$. Let

$$h_{\beta_0,T}(s) \equiv n^{-1/2}\sum_{i=1}^{n} T_i \exp(is'Z_i)\,.$$

Lemma 6.1. *Over the set $\{h : \|h\| \leq C\}$, $\mathrm{ICM}(h)$ is bounded and Lipschitz continuous in h.*

Proof. (a) Boundedness is trivial. For Lipschitz continuity,

$$\begin{aligned}|\mathrm{ICM}(h_1) - \mathrm{ICM}(h_2)| &= \left|\|h_1\|^2 - \|h_2\|^2\right| = |\langle h_1 - h_2, h_1 + h_2\rangle| \\ &\leq \|h_1 - h_2\|\|h_1 + h_2\| \leq \|h_1 - h_2\|(\|h_1\| + \|h_2\|) \leq 2\,C\,\|h_1 - h_2\|\,.\end{aligned}$$

□

Lemma 6.2. *Under Assumption A and D,*

$$\lim_{M\to\infty}\sup_{\beta_0}\sup_{P\in\mathcal{P}_{\beta_0}} \Pr\left[\mathrm{ICM}(\beta_0) > M\right] \to 0\,.$$

Proof. By definition

$$\mathrm{ICM}(\beta_0) = S'WS = n^{-1}\sum_{i=1}^{n} S_i^2 w(0) + n^{-1}\sum_{i=1}^{n}\sum_{j\neq i} S_i S_j w(Z_i - Z_j)\,.$$

Hence, for some constants $C, C', C'' > 0$ independent of $P \in \mathcal{P}_{\beta_0}$ and of β_0,

$$\Pr\left[n^{-1}\sum_{i=1}^{n} S_i^2 w(0) > M/2\right] \leq 2w(0)\frac{\mathbb{E}\,S_1^2}{M} \leq \frac{C}{M}$$

$$\Pr\left[n^{-1}\sum_{i=1}^{n}\sum_{j\neq i} S_i S_j w(Z_i - Z_j) > M/2\right] \leq 4C'\frac{\mathbb{E}^{\,2}(S_1^2)}{M^2} \leq \frac{C''}{M}\,,$$

using the boundedness of $w(\cdot)$ and Markov's inequality. □

6.1.3 Main proof

Let $\mathcal{P}_{\beta_0} = \{P \in \mathcal{P} : \beta = \beta_0\}$. From (22),

$$h_{\beta_0,S}(s) \rightsquigarrow \mathbb{G}_S(s)\,, \tag{24}$$

uniformly in $P \in \mathcal{P}_{\beta_0}$ and in β_0, where $\mathbb{G}_S(s)$ is a centered complex Gaussian process. Let $\widehat{\Omega}_i = \widehat{\Omega}(Z_i)$ and

$$\widehat{G}_i = (b_0'\Omega b_0)^{-1/2}\left(b_0'\widehat{\Omega}_i b_0\right)^{1/2}\varepsilon_i\,,$$

where the ε_i are independent $N(0,1)$. From our results in Section 6.1.1,

$$h_{\widehat{G}}(s) = n^{-1/2}\sum_{i=1}^{n}\widehat{G}_i\exp(is'Z_i) \rightsquigarrow \mathbb{G}_S(s)\,,$$

uniformly in $P \in \mathcal{P}$. We say that $h_{\beta_0,S}$ *uniformly weakly converges* to $h_{\widehat{G}}$ in $P \in \mathcal{P}$, i.e.

$$\sup_{\beta_0}\sup_{P\in\mathcal{P}_{\beta_0}} d_{BL}(h_{\beta_0,S}, h_{\widehat{G}}) \to 0\,,$$

see Kasy (2018) for a similar terminology. Let $F(x) = \mathbb{I}\left[x < C_1\right] + \frac{C_2-x}{C_2-C_1}\mathbb{I}\left[C_1 \le x \le C_2\right]$ for some $0 < C_1 < C_2$ and consider the continuous truncation of $\mathrm{ICM}(h_S)$ defined by $\mathrm{ICM}_F(h_S) = \mathrm{ICM}(h_S)F(\|h_S\|)$. Consider the conditional quantile of $\mathrm{ICM}_F(h)$

$$c_{F,1-\alpha}(h) = \inf\left\{c : \Pr\left[\mathrm{ICM}_F(h) \le c\right] \ge 1-\alpha\right\}\,.$$

Lemma 6.1 ensures that $\mathrm{ICM}_F(h)$ is Lipschitz, and it follows that $c_{F,1-\alpha}(h)$ is also Lipschitz. Indeed,

$$\begin{aligned} 1-\alpha &\le \Pr\left[\mathrm{ICM}_F(h_1) \le c_{F,1-\alpha}(h_1)\right] \\ &\le Pr\left[\mathrm{ICM}_F(h_2) \le c_{F,1-\alpha}(h_1) + K\|h_1 - h_2\|\right]\,, \end{aligned}$$

so that $c_{F,1-\alpha}(h_2) \le c_{F,1-\alpha}(h_1) + K\|h_1 - h_2\|$ for some constant $K > 0$. Interverting the role of h_1 and h_2 we get $c_{F,1-\alpha}(h_1) \le c_{F,1-\alpha}(h_2) + K\|h_1 - h_2\|$, so $c_{F,1-\alpha}(h)$ is Lipschitz in h.

Assume now that the conclusion of Theorem 3.1 does not hold. Then there exists some $\delta > 0$, an infinitely increasing subsequence of sample sizes n_j and a sequence of probability measures $P_{n_j} \in \mathcal{P}_{\beta_{0,n_j}}$, with corresponding sequences of β_{0,n_j} and $\Pi_{n_j}(\cdot)$, such that

$$\Pr_{n_j}\left[\mathrm{ICM}(h_{\beta_{0,n_j},S}) > c_{1-\alpha}(h_{\widehat{G}})\right] > \alpha + 3\delta \qquad \forall n_j\,.$$

Choose C_1 such that $\Pr_{n_j}\left[\mathrm{ICM}(h_{\beta_{0,n_j},S}) \ge C_1\right] < \delta$, which is possible from Lemma 6.2. Now

$$\Pr\left[\mathrm{ICM}(h_{\beta_0,S}) > x\right] \le \Pr\left[\mathrm{ICM}_F(h_{\beta_0,S}) > x\right] + \Pr\left[\mathrm{ICM}(h_{\beta_0,S}) \ge C_1\right]$$

for any β_0 and any P_{β_0}, and $c_{F,1-\alpha}(h) \leq c_{1-\alpha}(h)$, so that

$$\Pr_{n_j}\left[\mathrm{ICM}_F(h_{\beta_{0,n_j},S}) > c_{F,1-\alpha}(h_{\widehat{G}})\right] > \alpha + 2\delta \qquad \forall n_j\,.$$

As $\mathrm{ICM}_F(h)$ is bounded and Lipschitz in h, by the uniform convergence of $h_{\beta_0,S}$ to $h_{\widehat{G}}$,

$$\sup_{\beta_0}\sup_{P\in\mathcal{P}_{\beta_0}}\sup_{x}\left|\Pr\left[\mathrm{ICM}_F(h_{\beta_0,S}) > x\right] - \Pr\left[\mathrm{ICM}_F(h_{\widehat{G}}) > x\right]\right| \to 0\,.$$

Therefore for n_j large enough

$$\Pr_{n_j}\left[\mathrm{ICM}_F(h_{\widehat{G}}) > c_{F,1-\alpha}(h_{\widehat{G}})\right] \geq \alpha + \delta\,,$$

which contradicts the definition of $c_{F,1-\alpha}(h_{\widehat{G}})$.

7 Empirical Results

7.1 Simulation results

Polynomial model (i)	HICM	ICM	S-1IV	S-3IV	S-7IV
Level 5%					
Sample size of 101, $c = 3$	5.56	4.22	5.48	5.00	3.26
Sample size of 201, $c = 3$	3.82	3.36	4.98	4.72	4.14
Sample size of 401, $c = 3$	3.18	3.00	4.96	4.20	3.82
Level 10%					
Sample size of 101, $c = 3$	9.46	8.18	10.68	10.34	8.34
Sample size of 201, $c = 3$	7.34	7.18	9.62	9.82	9.44
Sample size of 401, $c = 3$	6.60	6.30	9.68	8.72	8.48

Linear model (ii)	HICM	ICM	S-1IV	S-3IV	S-7IV
Level 5%					
Sample size of 101, $c = 3$	5.56	4.22	5.48	5.00	3.26
Sample size of 201, $c = 3$	3.82	3.36	4.98	4.72	4.14
Sample size of 101, $c = 0$	5.56	4.22	5.48	5.00	3.26
Sample size of 201, $c = 0$	3.82	3.36	4.98	4.72	4.14
Level 10%					
Sample size of 101, $c = 3$	9.46	8.18	10.68	10.34	8.34
Sample size of 201, $c = 3$	7.34	7.18	9.62	9.82	9.44
Sample size of 101, $c = 0$	9.46	8.18	10.68	10.34	8.34
Sample size of 201, $c = 0$	7.34	7.18	9.62	9.82	9.44

Group heterogeneity model (iii)	HICM	ICM	S-1IV	S-3IV	S-7IV
Level 5%					
Sample size of 201, $c = 3$	3.58	4.34	4.68	4.36	2.76
Sample size of 101, $c = 0$					
Level 10%					
Sample size of 101, $c = 3$					
Sample size of 201, $c = 3$	7.42	8.78	9.92	8.96	7.48
Sample size of 101, $c = 0$					

Table 5: Empirical sizes for the three inference procedures, HICM, ICM and S, considered in section 4.1 for various DGPs and with significance level of 5% or 10%.

Polynomial model (i)	HICM-min	Jack-T			J-CUE		
		3-IV	7-IV	11-IV	3-IV	7-IV	11-IV
Level 5%							
Sample size of 101, $c = 3$	2.38	8.38	5.48	4.68	1.22	0.82	0.66
Sample size of 201, $c = 3$	1.34	8.40	4.84	4.78	1.48	1.44	0.88
Sample size of 101, $c = 7$	3.16	8.30	5.42	4.84	1.50	1.32	1.16
Level 10%							
Sample size of 101, $c = 3$	4.54	14.80	10.70	9.22	2.94	2.78	2.10
Sample size of 201, $c = 3$	3.26	15.70	9.74	8.86	3.40	3.84	6.62
Sample size of 101, $c = 7$	5.34	15.22	11.04	9.76	4.16	3.88	3.64

Linear model (ii)	HICM-min	Jack-T			J-CUE		
		3-IV	7-IV	11-IV	3-IV	7-IV	11-IV
Level 5%							
Sample size of 101, $c = 3$	1.38	4.04	3.90	4.22	1.14	0.64	0.40
Sample size of 201, $c = 3$	1.02	4.18	4.64	4.62	1.44	1.26	0.82
Sample size of 101, $c = 7$	2.00	2.78	3.08	3.84	1.58	1.26	1.08
Level 10%							
Sample size of 101, $c = 3$	2.88	8.04	7.86	8.50	3.26	2.30	1.66
Sample size of 201, $c = 3$	2.44	7.84	8.28	8.40	3.62	3.76	3.58
Sample size of 101, $c = 7$	3.64	6.72	6.96	8.02	4.02	3.92	3.22

Group heterogeneity model (iii)	HICM-min	Jack-T			J-CUE		
		3-IV	7-IV	11-IV	3-IV	7-IV	11-IV
Level 5%							
Sample size of 201, $c = 3$	1.28	1.56	5.30	4.58	0.04	1.24	0.66
Level 10%							
Sample size of 201, $c = 3$	3.40	3.80	10.78	9.02	0.32	3.80	2.48

Table 6: Empirical sizes for the three specification tests, HICM-min, Jack-T and J-CUE, considered in section 4.2 for various DGPs and with a significance level of 5% or 10%.