10ISSN 1183-1057

SIMON FRASER UNIVERSITY

Department of Economics

Working Papers

12-17

"Testing Identification Strength"

Bertille Antoine and
Eric Renault

September, 2012

SFU Economics

Testing Identification Strength*

Bertille Antoine[†] and Eric Renault[‡]

September 2012

Abstract

We consider models defined by a set of moment restrictions that may be subject to weak identification. Following the recent literature, the identification of the structural parameters is characterized by the Jacobian of the moment conditions. We unify several definitions of identification that have been used in the literature, and show how they are linked to the consistency and asymptotic normality of GMM estimators. We then develop two tests to assess the identification strength of the structural parameters. Both tests are straightforward to apply. In simulations, our tests are well-behaved when compared to contenders, both in terms of size and power.

Keywords: GMM; Weak IV; Test; Misspecification.

JEL classification: C32; C12; C13; C51.

1 Introduction

Hansen's (1982) seminal paper on the Generalized Method of Moments (GMM) has provided a unified asymptotic theory that encompasses the classical econometric tools of estimation

*This paper has been circulating under the title "Specification tests for strong identification".

[†]*Simon Fraser University. Email: bertille_antoine@sfu.ca.* *The author gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada and the Economics Department at Brown University where part of this work was completed.*

[‡]*Brown University. Email: eric_renault@brown.edu*

and inference about structural parameters based on instrumental variables (IV). Such a unified asymptotic theory delivers asymptotically normal estimators with asymptotic variances that can be estimated easily to build Wald-type confidence sets. However, over the last 30 years, the practice of GMM has shown that these asymptotically normal distributions may be poor approximations of the actual distributions met in finite samples. Such distortions, observed even for relatively large samples, might be explained by the weak correlation between instruments and explanatory variables.

The weak instrument literature has proposed two kinds of alternative asymptotics to capture actual finite samples estimator's distributions. On the one hand, Staiger and Stock's (1997) asymptotic approximation (see also Stock and Wright (2000) for its non-linear generalization) is such that IV estimators have non-standard distributions. On the other hand, a more recent literature still considers the IV estimators as approximately normal, but such that the standard asymptotic variance estimators may not be as reliable as in the strong instrument approach. While several authors, including Hansen, Hausman and Newey (2008) in the linear case, and Newey and Windmeijer (2009) in the non-linear case, have justified such adjustments of Gaussian-based confidence intervals by the so-called many-instrument asymptotics, others are more agnostic and simply acknowledge that slower rates of convergence towards normality may occur: see e.g. Hahn and Kuersteiner (2002) in the linear case, Antoine and Renault (2009) and Caner (2010) in the non-linear case. In this respect, the fact that the number of instruments may be seen as going to infinity with the sample size is only one possible interpretation of these non-standard rates.

More generally, the weak instrument literature can be understood by considering the reduced rank setting as the limit of a sequence of Data Generating Processes (DGP) indexed by the sample size. Antoine and Renault (2009) have characterized how various degrees of identification weakness (as defined by the rate of convergence towards reduced rank along the sequence of drifting DGPs) lead to various rates of convergence for estimators of structural parameters. Besides the extreme case of weak identification studied by Staiger and Stock (1997) and Stock and Wright (2000), we show that only a slightly less severe identification issue, or so-called near-weak identification, ensures asymptotic normality (albeit at a slower rate than standard root-T), allowing almost standard GMM inference; see also Antoine and Renault (2009).

Our first contribution is to unify several definitions of identification that have been used in the literature, and to show how they are linked to the asymptotic properties of GMM estimators. We also discuss the identification of subvectors. From there, our second contribution is to propose simple test procedures to assess the identification strength of the structural parameters and the validity of the moment restrictions. Such tests should provide user-friendly guidelines to practitioners. For instance, our tests for identification strength aim at concluding that, up to some type 1 error, standard GMM inference can be applied safely (and possibly efficiently). All the tests we propose are straightforward to apply since based on squared norms of the moment conditions computed at a suitable estimator, in the spirit of Sargan (1958) and Hansen's (1982) J-test for overidentification. In addition, our tests have good power properties as illustrated in our Monte-Carlo simulations. This is in contrast with what one commonly thinks about J-tests. Beyond the Monte-Carlo evidence, we also provide some theoretical arguments to understand why our testing procedures may have power to identify instruments' strength or validity. The key intuition is that the GMM optimal weighting matrix automatically sets the focus on the most informative moment conditions. As a result, an inflation of the number of moment conditions may not hurt so much the power of the test.

Our tests for identification strength are inspired by Dufour's (1997) seminal observation that, when the degree of overidentification can be arbitrarily small, valid confidence sets should be infinite with a positive probability. In terms of tests, it is akin to consider that a null hypothesis written as an infinite distortion of the true value may deliver a positive p-value. This is the key idea behind our two tests for identification. Both tests consider the null hypothesis that no standard asymptotic theory (as discussed above) is reliable.

In linear settings, we simply test the null hypothesis of weak identification against near-weak identification. We consider a J-test statistic of overidentification computed at a distorted GMM estimator. The distortion is such that it cannot be detected under the null, but allows the test to reject consistently under the alternative in spite of the fact that the test can be conservative. Practical considerations for the choice of such distortion are also discussed.

In non-linear settings, we explain why we cannot directly test the null hypothesis of weak identification against near-weak identification. As a result, we have to rely on the sufficient condition for near-weak identification put forward by Antoine and Renault (2009), namely near-strong identification. Our second test is a conservative test regarding the null that some components of the vector of structural parameters are not near-strongly identified. We also use a J-test statistic of overidentification computed at a distorted GMM estimator. Both tests are conservative since, under the null, no standard asymptotic theory is available. By contrast, when the null hypothesis is rejected, the practitioner can safely apply standard inference procedures, like the overidentification test, or the Wald test, since studentization protects her against possibly slower rates of convergence (see e.g. Antoine and Renault (2009) and Newey and Windmeijer (2009)). Note also that both our tests set the focus on testing identification strength of subvectors which is in contrast with respect to common practice.

Other test procedures have been proposed in the literature to detect weak identification in linear settings. In simulations, we compare our tests to detect identification strength to the rule-of-thumb proposed by Staiger and Stock (1997) and to the tests based on the 2SLS bias and size distortion proposed by Stock and Yogo (2005). The test proposed by Hahn and Hausman (2003) tests the null of strong identification and is not considered here. Detecting weak identification in non-linear settings is an open area of research[1], and our second test provides some partial answers to this issue.

The paper is organized as follows. In section 2, we introduce our framework and characterize the identification strength of structural parameters through the Jacobian of the moment restrictions. We also show how it is linked to the asymptotic properties of GMM estimators. In section 3, we propose two tests to assess the identification strength of the structural parameters in linear and non-linear settings. In section 4, we illustrate the finite samples performance of our tests though Monte-Carlo simulations. We consider the linear IV regression model and a (persistent) AR(1) model calibrated to interest rate data. Section 5 concludes. The proofs of the main results are gathered in the Appendix.

The following notation is used throughout the paper. The symbols "$\stackrel{p}{\to}$" and "$\stackrel{d}{\to}$" denote convergence in probability and in distribution, while "$\texttt{Plim}$" denotes the probability limit of a random expression. $o_p(1)$ denotes a random variable that converges to 0 in probability, whereas $\mathcal{O}_p(1)$ denotes a random variable that is bounded in probability. For any (k,p)-

[1]The test proposed by Wright (2003) using the estimated curvature of the objective function is actually a test for non-identification.

matrix M, "M'" denotes the transpose matrix of M, $\texttt{Rank}(M)$ denotes the rank of M, and $\|M\| \equiv \max\{\sqrt{\lambda}, \lambda \text{ is an eigenvalue of } M'M\}$. $\mathbf{I}_q$ denotes the identity matrix of size q. $\chi^2(k)$ denotes the central chi-square random variable with k degrees of freedom. "With respect to" is written "w.r.t".

2 General framework

2.1 Identification strength

We consider the true unknown value θ^0 of the parameter $\theta \in \Theta \subset \mathbb{R}^p$ defined as the solution of the moment conditions,

$$E[\phi_t(\theta)] = 0 \quad \text{for some known function } \phi_t(.) \text{ of size } K. \tag{2.1}$$

Since the seminal work of Stock and Wright (2000), the weakness of the moment conditions (or instrumental variables) is usually captured through a drifting DGP such that the informational content of the estimating equations shrinks towards zero (for all θ) while the sample size T grows to infinity. The strength of identification of the parameters is then reflected by the Jacobian of the moment equations with respect to the parameters. We maintain the assumptions that the moment function $\phi_t(.)$ is continuously differentiable with respect to θ on the interior of the set of possible parameter values Θ, $\texttt{int}(\Theta)$, and that the true unknown value θ^0 belongs to $\texttt{int}(\Theta)$. We now unify several definitions of identification strength of θ that have been used in the literature.

Definition 2.1. *(Identification strength of θ)*
The identification strength of θ is characterized by a sequence M_T of deterministic nonsingular matrices of size p such that

$$\Gamma(\theta^0) \equiv \texttt{Plim}\left[\frac{\partial\overline{\phi}_T(\theta^0)}{\partial\theta'} M_T\right] \qquad \textit{exists and is full-column rank.} \tag{2.2}$$

We borrow the terminology "identification strength" to Kleibergen and Mavroeidis (2009) who set the focus (see their Assumption 6) on the special case where

$$\frac{\partial\overline{\phi}_T(\theta^0)}{\partial\theta'} = \Gamma(\theta^0) M_T^{-1} .$$

They stress the importance of characterizing the identification strength of θ to draw valid inference about some other parameters of the model of interest. The faster the sequence of matrices M_T diverges to infinity, the lesser θ is identified. It is actually strongly identified when M_T can be taken as the identity matrix. The concept of identification strength has been extensively studied in Antoine and Renault (2009, 2010). In the context of many instruments, it is revisited in Assumption 1(ii) in Newey and Windmeijer (2009). An important message of all these papers is that different linear combinations of θ may display different strengths (or degrees) of identification. More generally, the identification strength of the possible linear combinations of θ is tightly related to the rate of convergence of the eigenvalues of $(M_T M_T')$ to infinity, while the eigenvectors describe the linear combinations corresponding to different degrees of identification (see Antoine and Renault (2009, 2010) and assumption 4 below). The role of the sequence of matrices M_T in the asymptotic distributional theory of the GMM estimator of θ is well-understood, at least in the linear case, as first pointed out by Staiger and Stock (1997) and reminded in the example below.

Example 2.1. *(Linear IV regression)*

We consider the following structural linear equation,

$$y_t = x_t'\theta^0 + u_t \qquad \text{for } t = 1, \cdots, T,$$

where the p explanatory variables x_t may be endogenous. The true unknown value θ^0 of the structural parameter is identified through $K \geq p$ instrumental variables z_t uncorrelated with u_t. In other words, the estimating equations for standard IV estimation are

$$\overline{\phi}_T(\hat{\theta}_T) = \frac{1}{T} Z' \left(y - X\hat{\theta}_T \right) = 0\,, \tag{2.3}$$

where X (respectively Z) is the (T, p) (respectively (T, K)) matrix which contains the available observations of the p explanatory variables (respectively the K instrumental variables) and $\hat{\theta}_T$ denotes the standard IV estimator of θ. The reduced form equation for X can be written as

$$X = Z\Pi_T + V\,, \tag{2.4}$$

where the K columns of Z and the p columns of V are uncorrelated. Note that, at the price of more tedious notations, one could easily accommodate the general model considered in

Staiger and Stock (1997) where additional exogenous variables W show up in both the structural and the reduced form equation. Actually, everything can be understood in the more general setting by considering orthogonal projections on the orthogonal space of the range of W. Then, we have

$$\frac{\partial\overline{\phi}_T(\theta)}{\partial\theta'} = -\frac{Z'X}{T} = -\frac{Z'Z}{T}\Pi_T - \frac{Z'V}{T}\,. \tag{2.5}$$

Under standard regularity conditions, $(Z'V/T)$ and $(Z'Z/T)$ converge respectively towards zero and a nonsingular matrix Σ_Z. Therefore, (2.5) can be used to reinterpret the above definition 2.1 in terms of an asymptotic specification of the matrix Π_T,

$$\Pi_T = \Pi M_T^{-1} \quad \text{with} \quad \texttt{Rank}(\Pi) = p\,. \tag{2.6}$$

In other words, instead of a fixed full-column rank matrix Π, drifting reduced form parameters Π_T are used to capture the fact that identification may be weaker than usual (e.g. when coefficients of M_T go to infinity). Staiger and Stock (1997) actually define weak identification by considering $\Pi_T = \Pi/\sqrt{T}$, that is $M_T = \sqrt{T}\mathbf{I}_p$. The conformity between condition (2.6) and definition 2.1 follows from (2.5) by noting that we have for all $\theta \in \Theta$,

$$\frac{\partial\overline{\phi}_T(\theta)}{\partial\theta'}M_T = -\frac{Z'Z}{T}\Pi - \frac{Z'V}{\sqrt{T}}\frac{M_T}{\sqrt{T}}\,.$$

In the weak identification case ($M_T/\sqrt{T} = \mathbf{I}_p$), an extension of definition 2.1 would lead to consider a random matrix $\Gamma(\theta^0)$ since the Jacobian matrix rescaled by $\sqrt{T}$ is asymptotically normal as $Z'V/\sqrt{T}$. The effects of this randomness have been documented by Kleibergen (2005) for a score test on the whole parameter vector. By contrast, such randomness is implicitly precluded in Kleibergen and Mavroeidis (2009) (see their Assumption 6) for the Jacobian matrix of parameters not under test. The only way to ensure that the matrix $\Gamma(\theta^0)$ in definition 2.1 is not random is to assume, in addition, that

$$\lim_T \left(\frac{M_T}{\sqrt{T}}\right) = 0\,. \tag{2.7}$$

Condition (2.7) has been dubbed near-weak identification by Hahn and Kursteiner (2002) who typically consider

$$M_T = T^{\lambda}\mathbf{I}_p \quad \text{with} \quad 0 < \lambda < 1/2\,.$$

The extreme cases $\lambda = 0$ and $\lambda = 1/2$ correspond respectively to strong and weak identification. Precluding the extreme case of weak identification, or, in other words, maintaining the rank condition (2.6) with the upper bound (2.7) on the rate of weakness is key to get asymptotic normality of the IV estimator $\hat{\theta}_T$ with standard studentized statistics. To see this, simply rewrite (2.3) as

$$\begin{aligned} & \frac{Z'X}{T}(\hat{\theta}_T - \theta^0) = \frac{Z'u}{T} \\ \Leftrightarrow\quad & \frac{Z'Z}{T}\Pi\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0) + \frac{Z'V}{\sqrt{T}}\frac{M_T}{\sqrt{T}}\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0) = \frac{Z'u}{\sqrt{T}}\,. \end{aligned} \tag{2.8}$$

Under standard regularity conditions, (2.8) delivers asymptotic normality of

$$\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0)\,,$$

after noting that, thanks to (2.7), we have

$$\Sigma_Z\Pi\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0) = \frac{Z'u}{\sqrt{T}} + o_P(1)\,, \tag{2.9}$$

with $\Sigma_Z\Pi$ full-column rank and $Z'u/\sqrt{T}$ asymptotically normal. Of course, since near-weak identification entails some coefficients of the matrix M_T diverging to infinity (albeit not as fast as $\sqrt{T}$), the rate of convergence of the IV estimator to normality may be slower than $\sqrt{T}$. The fact that the matrix M_T may not be proportional to the identity matrix (and not even diagonal) allows for linear combinations of θ to have different degrees of identification. Assuming M_T diagonal means that different identification strengths are assigned to different columns of $Z\Pi$, and not to different instruments (or columns of Z). The two are not equivalent in the overidentified case since Π is not a square matrix. Therefore, maintaining the diagonality of M_T, albeit commonly done, would be overly restrictive[2].

Assumption 1 in Newey and Windmeijer (2009) highlights the importance of the near-weak identification condition (see their condition $\mu_{jn}/\sqrt{n} \to 0$; see also Hansen, Hausman

[2]See further assumption 4 for a convenient generalization of diagonality.

and Newey (2008) for the linear case). As already explained, this assumption allows, at least in the linear case, to get consistent asymptotically normal estimators whose rates of convergence are described by the sequence of matrices $\sqrt{T}M_T^{-1}$ (see (2.8)). The non-linear case for near-weak identification, as studied by Antoine and Renault (2009, 2010) and Caner (2010) works similarly. It can be shown under very general conditions (see Antoine and Renault (2009, 2010)) that any GMM estimator of θ will display a rate of convergence at least equal to $\sqrt{T}/\|M_T\|$. It is worth stressing that while we allow moments to display some singularities, the GMM estimators we consider are all defined in a standard way from a positive definite weigthing matrix.

Definition 2.2. *(GMM estimator)*
For any sequence of possibly random symmetric matrices Ω_T of size K that converges in probability towards a positive definite matrix Ω, a GMM estimator $\hat{\theta}_T$ is defined as any solution of

$$\min_{\theta\in\Theta}\left[\overline{\phi}_T(\theta)'\Omega_T\overline{\phi}_T(\theta)\right] .$$

Regarding asymptotic normality of such GMM estimators, the proof requires a Taylor expansion of the first-order conditions to get a linear representation of the GMM estimator that generalizes the linear case (2.9). Non-linearity may then entail an additional technical difficulty due to the fact that the concept of identification strength may not be robust to plugging in the Jacobian matrix a consistent estimator of the true unknown value. This is why we consider the following high-level condition that strengthens definition 2.1.

Definition 2.3. *(Near-weak identification)*
In the context of definition 2.1, θ is said near-weakly identified if there exists a sequence M_T of deterministic nonsingular matrices of size p such that

$$\lim_T\left(\frac{M_T}{\sqrt{T}}\right) = 0\,,$$

and, for any GMM estimator $\hat{\theta}_T$ as in definition 2.2 and any sequence θ_T between θ^0 and $\hat{\theta}_T$ component by component[3]*, we have*

$$\Gamma(\theta^0) = \texttt{Plim}\left[\frac{\partial\overline{\phi}_T(\theta_T)}{\partial\theta'}M_T\right]\,,$$

[3]Hereafter, we use the notation $\theta_T\in\ [\theta^0,\hat{\theta}_T]$.

where $\Gamma(\theta^0)$ *is the full-column rank matrix introduced in (2.2).*

When definition 2.1 is fulfilled with $M_T/\sqrt{T}$ going to zero, near-weak identification of θ will be warranted in many cases. It is worth realizing that going from the former to the latter only amounts to assume that

$$\theta_T \in [\theta^0, \hat{\theta}_T] \quad \Rightarrow \quad \texttt{Plim}\left[\left(\frac{\partial \overline{\phi}_T(\theta_T)}{\partial \theta'} - \frac{\partial \overline{\phi}_T(\theta^0)}{\partial \theta'}\right) M_T\right] = 0\,. \tag{2.10}$$

The required zero-limit in (2.10) is obviously ensured for the components of the moment vector $\overline{\phi}_T(.)$ that are linear with respect to the parameters θ. For those which are not linear with respect to some subset θ_1 of components of θ, the issue at stake is to know whether their rate of convergence along the sequence θ_T is sufficient to supersede the possible convergence to infinity of the sequence M_T. As explained above, we expect for the GMM estimator $\hat{\theta}_T$ (and thus also for $\theta_T \in [\theta^0, \hat{\theta}_T]$) a rate of convergence at least equal to $\sqrt{T}/\left\|M_T\right\|$. More precisely, we expect, as in the linear case, that $\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0) = \mathcal{O}_P(1)$. Hence, the validity of (2.10) would mean that in relevant directions, the convergence to zero of $M_T/\sqrt{T}$ dominates the convergence to infinity of the sequence M_T. Roughly speaking, $\|M_T\|$ should not blow-up as fast as $T^{1/4}$. This threshold is the key concept of near-strong identification promoted by Antoine and Renault (2009, 2010) as a sufficient condition for near-weak identification. A less restrictive, albeit related, point of view will be warranted in section 3 when testing for the identification strength of subvectors θ in non-linear settings. We first present the asymptotic theory for GMM estimators under the high-level assumption of near-weak identification. While a careful study of rates of convergence of GMM estimators is provided in Antoine and Renault (2009, 2010), we simplify the exposition by directly maintaining the required high-level assumptions.

Assumption 1. θ *is near-weakly identified as in definition 2.3.*

Assumption 2. *In the context of assumption 1, any GMM estimator* $\hat{\theta}_T$ *as in definition 2.2 is such that* $\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0) = \mathcal{O}_P(1)$.

2.2 Asymptotic Theory

As usual, asymptotic normality of a GMM estimator results from a central limit theorem applied to the moment conditions evaluated at the true unknown value of the parameters.

Assumption 3. *$\sqrt{T}\phi_T(\theta^0)$ converges in distribution towards a normal distribution with mean zero and variance $S(\theta^0)$.*

The following theorem extends the asymptotic normality result given in the linear case, as well as results previously given in Antoine and Renault (2009, 2010).

Theorem 2.1. *(Asymptotic normality)*
Let $\hat{\theta}_T$ denote any GMM estimator as in definition 2.2. Under assumptions 1 to 3,

$$\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0)$$

is asymptotically normal with mean zero and variance

$$\Sigma(\theta^0) = \left[\Gamma'(\theta^0)\Omega\Gamma(\theta^0)\right]^{-1}\Gamma'(\theta^0)\Omega S(\theta^0)\Omega\Gamma(\theta^0)\left[\Gamma'(\theta^0)\Omega\Gamma(\theta^0)\right]^{-1} .$$

As already acknowledged, assumptions 1 and 2 are high-level assumptions, and we refer the interested reader to Antoine and Renault (2010) for more primitive conditions. In any case, Theorem 2.1 paves the way for a concept of efficient estimation. By a common argument, the unique limit weighting matrix Ω minimizing the above covariance matrix is clearly $\Omega = [S(\theta^0)]^{-1}$.

Theorem 2.2. *(Efficient GMM estimator)*
Under the assumptions of Theorem 2.1, any GMM estimator $\hat{\theta}_T$ as in definition 2.2 with a weighting matrix $\Omega_T = S_T^{-1}$, where S_T denotes a consistent estimator of $S(\theta^0)$, is such that

$$\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0)$$

is asymptotically normal with mean zero and variance $\left[\Gamma'(\theta^0)S^{-1}(\theta^0)\Gamma(\theta^0)\right]^{-1}$.

In our framework, the terminology efficient GMM must be carefully qualified. For all practical purposes, Theorem 2.2 states that, for T large enough, $\sqrt{T}M_T^{-1}(\hat{\theta}_T - \theta^0)$ can be seen as a Gaussian vector with mean zero and variance consistently estimated by

$$M_T^{-1}\left[\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}S_T^{-1}\frac{\partial\overline{\phi}_T(\hat{\theta}_T)}{\partial\theta'}\right]^{-1}M_T^{-1\prime}\,, \tag{2.11}$$

since $\Gamma(\theta^0) = \texttt{Plim}\left[\frac{\partial\overline{\phi}_T(\theta^0)}{\partial\theta'}M_T\right]$. However, it is incorrect to deduce from formula (2.11) that $\sqrt{T}\left(\hat{\theta}_T - \theta^0\right)$ can be seen (for T large enough) as a Gaussian vector with mean zero and variance consistently estimated by

$$\left[\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}S_T^{-1}\frac{\partial\overline{\phi}_T(\hat{\theta}_T)}{\partial\theta'}\right]^{-1} . \tag{2.12}$$

The above matrix (2.12) is actually the inverse of an asymptotically singular matrix. In this sense, a truly standard GMM theory does not apply and some components of $\sqrt{T}(\hat{\theta}_T - \theta^0)$ actually blow-up. Quite fortunately, standard inference procedures work, albeit for non-standard reasons. For all practical purposes related to inference about the structural parameter θ, the knowledge of the matrix M_T is not required; see also the discussion in Antoine and Renault (2010). Of course, the asymptotic singularity of (2.12) means that the actual rate of convergence may vary depending on the linear combinations of the structural parameter vector θ. The following high-level assumption[4] helps to characterize the relevant directions in the parameter space.

Assumption 4. *The sequence of matrices M_T such that assumptions 1 and 2 are fulfilled can be chosen as*

$$M_T = R\Lambda_T \, ,$$

for some fixed non-singular matrix R and a sequence Λ_T of diagonal matrices.

As shown in the appendix, the main intuition is that, from an initial sequence of matrices M_T that does not fulfill assumption 4, we can build the diagonal coefficients of Λ_T as the singular values of the matrix M_T (as the square-roots of eigenvalues of $(M_T M'_T)$), while the matrix R is the limit of a sequence of orthogonal matrices of eigenvectors of $(M_T M'_T)$. Then, Theorems 2.1 and 2.2 characterize the asymptotic normal distribution of $\sqrt{T}\Lambda_T^{-1}R^{-1}(\hat{\theta}_T - \theta^0)$. In other words, when considering the reparametrization $\eta \equiv R^{-1}\theta$, the j-th component of $\hat{\eta}_T \equiv R^{-1}\hat{\theta}_T$ is a consistent asymptotically normal estimator of η_j^0 (with $\eta^0 \equiv R^{-1}\theta^0$) with a rate of convergence $\sqrt{T}/\lambda_{jT}$ (with λ_{jT} the j-th diagonal coefficient of Λ_T). Moreover, Antoine and Renault (2010) have shown that this rate of convergence is not impacted by

[4]More primitive justifications are provided in the appendix

a preliminary consistent estimation of the matrix R, making this asymptotically normal estimation feasible. The characterization of the different rates of convergence through the sequence of diagonal coefficients of the matrix Λ_T may matter for interpretation, even though their knowledge is not necessary to run Wald inference from (2.11). A similar discussion may be relevant to interpret the outcome of a J-test for overidentification while taking into account the heterogeneous strengths of instruments. We can already note that the J-test for overidentification can be performed as usual thanks to the following result.

Theorem 2.3. *(J-test)*
Under the assumptions of Theorem 2.2, for any GMM estimator as in definition 2.2 with a weighting matrix $\Omega_T = S_T^{-1}$, where S_T denotes a consistent estimator of $S(\theta^0)$, we have

$$T\overline{\phi}_T'(\hat{\theta}_T)S_T^{-1}\overline{\phi}_T(\hat{\theta}_T) \xrightarrow{d} \chi^2(K-p)\,.$$

2.3 Identification of subvectors

When testing for identification strength, we will check whether our data set allows us to reject the null hypothesis that some components of the vector θ of structural parameters are (very) poorly identified. Throughout, θ_1 denotes a vector of p_1 components of θ, while $\theta_{\backslash 1}$ collects the p_2 $(= p-p_1)$ remaining components of θ that are not included in θ_1. For simplicity, θ_1 corresponds hereafter to the first p_1 components of θ, that is $\theta = (\theta_1' \ \theta_{\backslash 1}')'$. Typically, we consider cases where the econometrician's is concerned about the poor identification of the components of $\theta_{\backslash 1}$ while prior knowledge warrants "sufficiently strong" identification of θ_1.[5]

Note that it is only when a sequence of matrices M_T characterizing the identification strength of θ is block-diagonal,

$$M_T = \begin{bmatrix} M_{1T} & \mathbf{0} \\ \mathbf{0} & M_{\backslash 1T} \end{bmatrix}, \qquad (2.13)$$

that we can deduce the identification strengths of θ_1 and $\theta_{\backslash 1}$ from the identification strength of θ. A well-known example is the setting put forward by Stock and Wright (2000) where

[5]Note that our test does not exclude the case where $\theta_{\backslash 1} = \theta$.

the two subsets of components of θ are disentangled as follows[6],

$$\overline{\phi}_T(\theta) = \overline{\phi}_{1T}(\theta_1) + \frac{1}{T^\lambda}\overline{\phi}_{2T}(\theta), \quad \text{with} \;\; 0 < \lambda \leq 1/2\,,$$

$$\texttt{Rank}\left(\frac{\partial\overline{\phi}_{1T}(\theta_1^0)}{\partial\theta_1'}\right) = p_1 \quad \text{and} \quad \texttt{Rank}\left(\frac{\partial\overline{\phi}_{2T}(\theta^0)}{\partial\theta_{\backslash 1}'}\right) = p_2\,.$$

M_T can then be defined as in (2.13) with $M_{1T} = \mathbf{I}_{p_1}$ and $M_{\backslash 1T} = T^\lambda \mathbf{I}_{p_2}$.

It is worth pointing out that, up to a convenient reparameterization, the above block-diagonal structure of (2.13) is not really restrictive. From assumption 4, we define the new vector of parameters, $\eta \equiv R^{-1}\theta$ whose identification strength (in the sense of definition 2.1) is described through the sequence of diagonal matrices Λ_T. Hence, the maintained assumption (2.13) simply means that such a reparameterization is possible with a block-diagonal matrix R. It makes sense to question the identification strength of $\theta_{\backslash 1}$ while maintaining a (near-weak) identification assumption on θ_1, precisely because the two subvectors are disentangled in the classification of directions as regards identification strength.

Our tests for identification strength will provide some (partial) answers to the following question: taking for granted that θ_1 is near-weakly identified, do our data confirm the identification of a larger vector of unknown parameters? Our null hypothesis will be devised such that, failing to reject it means that we cannot rely upon standard inference based on the Gaussian asymptotic theory of section 2.2 when any of the parameters in $\theta_{\backslash 1}$ are considered as unknown. In front of such a negative evidence, only two strategies are available:

- either one resorts to inference procedures that are robust to weak identification. Of course, robustness has a cost in terms of efficiency of estimators, power of tests, and maintained assumptions regarding nuisance parameters;

- or, following the common practice of calibration, one may fix the value of parameters in $\theta_{\backslash 1}$ at pre-specified levels provided by other studies hoping that these calibrated values are not too far from the unknown ones and will not contaminate inference on θ_1. The validity of this practice has been extensively studied by Dridi, Guay and Renault (2007) who propose some encompassing tests for backtesting it.

In any case, both strategies will always maintain the assumption that, when $\theta_{\backslash 1}$ is fixed at its true unknown value $\theta_{\backslash 1}^0$, the remaining moment problem is well-behaved.

[6]Strictly speaking, Stock and Wright (2000) only consider the limit case with $\lambda = 1/2$.

Definition 2.4. *(Near-weak identification of a subvector)*
In the context of definition 2.1, with the block-diagonal structure (2.13) for the sequence of matrices M_T, θ_1 is near-weakly identified if the two following conditions are fulfilled.
(i) Definition 2.3 is fulfilled for the sequence of matrices M_{1T} in the context of the (infeasible) moment model

$$E\left[\phi_t(\theta_1, \theta^0_{\backslash 1})\right] \quad = \quad 0 \quad \textit{with} \quad \theta_1 \in \Theta(\theta^0_{\backslash 1}) = \left\{\theta_1 \in \mathbb{R}^{p_1}; (\theta_1, \theta^0_{\backslash 1}) \in \Theta\right\} .$$

(ii) For any GMM estimator $\hat{\theta}_T$ as in definition 2.2 and any sequence θ^*_T such that $\theta^*_{1T} = \hat{\theta}_{1T}$ and $\theta^*_{\backslash 1T} - \hat{\theta}_{\backslash 1T} = o_P(T^{-1/4})$, we have

$$\frac{\partial \overline{\phi}_T(\theta^*_T)}{\partial \theta'_{\backslash 1}} M_{\backslash 1T} = \mathcal{O}_p(1) \, .$$

When wondering whether a parameter vector that strictly nests θ_1 is near-weakly identified, let us recall that the key issue is to check that the convergence condition (2.10) holds for any sequence θ_T between the true value θ^0 and some GMM estimator $\hat{\theta}_T$, that is

$$\texttt{Plim}\left[\left(\frac{\partial \overline{\phi}_T(\theta_T)}{\partial \theta'} - \frac{\partial \overline{\phi}_T(\theta^0)}{\partial \theta'}\right) M_T\right] = 0 \, . \tag{2.14}$$

The reason why the maintained assumption $\lim_T (M_T/\sqrt{T}) = 0$ may not be sufficient to get (2.14) is that a rate of convergence for θ_T strictly slower than $\sqrt{T}$ may be unable to protect against the asymptotic blow-up of the sequence M_T. This issue will obviously be easier to control for, when the weakest parameters $\theta_{\backslash 1}$ will not be multiplied by the most explosive part of the sequence M_T, namely $M_{\backslash 1T}$. This explains why we consider first the most favorable circumstances of linearity with respect to $\theta_{\backslash 1}$.

- Case i): Moment conditions affine w.r.t. $\theta_{\backslash 1}$, $\phi_t(\theta) = a_t(\theta_1) + B_t(\theta_1)\theta_{\backslash 1}$.

Regarding condition (i), note that we have

$$\frac{\partial \overline{\phi}_T(\theta_T)}{\partial \theta'} M_T \quad = \quad \left[\begin{array}{cc} \frac{\partial \bar{\phi}_T(\theta_T)}{\partial \theta'_1} M_{1T} & \bar{B}_T(\theta_{1T}) M_{\backslash 1T} \end{array} \right] . \tag{2.15}$$

To get the second block of (2.15) consistent with (2.14) when the only assumption is that $\lim_T (M_{\backslash 1T}/\sqrt{T}) = 0$ θ_{1T} has to be $\sqrt{T}$-consistent, that is $M_{1T} = \mathbf{I}_{p_1}$. In this case, it is clear that a continuity assumption on $\frac{\partial \overline{a}_T(.)}{\partial \theta'_1}$ and $\frac{\partial \overline{B}_T(.)}{\partial \theta'_1}$ will be sufficient to deduce (2.14) from (2.15). Overall, up to some regularity conditions, we can conclude the following.

Golden rule for the test of identification strength in the linear case: If the moment conditions are affine w.r.t. a subvector $\theta_{\backslash 1}$, strong identification of the other components θ_1 ($M_{1T} = \mathbf{I}_{p_1}$) jointly with ($\lim_T (M_{\backslash 1T}/\sqrt{T}) = 0$) is sufficient to warrant near-weak identification of the whole vector θ. Note that when the moment conditions are affine w.r.t. the whole vector θ, no strong identification condition is required.

- Case ii): General (non-linear) moment conditions.

Losing linearity w.r.t. $\theta_{\backslash 1}$ implies that the second block of (2.15) is now $\left(\frac{\partial \bar{\phi}_T(\theta_T)}{\partial \theta'_{\backslash 1}}\right) M_{\backslash 1T}$, with $\left(\frac{\partial \bar{\phi}_T(\theta_T)}{\partial \theta'_{\backslash 1}}\right)$ that usually depends on the estimator of the weakest parameters $\theta_{\backslash 1T}$. Since these parameters are only consistent at a rate $\left\| M_{\backslash 1T} \right\| / \sqrt{T}$, a Taylor expansion of $\left(\frac{\partial \bar{\phi}_T(\theta_T)}{\partial \theta'_{\backslash 1}}\right)$ (assuming ϕ twice continuously differentiable) will allow us to prove (2.14) only if $\left\| M_{\backslash 1T} \right\|^2 / \sqrt{T}$ goes to zero when T goes to infinity. In other words, the condition to ensure that the poor identification of $\theta_{\backslash 1}$ does not impair near-weak identification of the whole vector θ is that $\left\| M_{\backslash 1T} \right\| = o(T^{1/4})$, or equivalently that the rate of convergence of all parameters in $\theta_{\backslash 1}$ (rate defined by the sequence of matrices $M_{\backslash 1T}/\sqrt{T}$) is more than $T^{1/4}$. As already emphasized by Antoine and Renault (2012), this condition is quite similar in spirit to Andrews' (1994) study of MINPIN estimators, or estimators defined as MINimizing a criterion function that might depend on a Preliminary Infinite dimensional Nuisance parameter estimator. Even without an infinite dimensional issue, we intuitively want to make sure that second-order terms in Taylor expansions (see the discussion above regarding a Taylor expansion of $\left(\frac{\partial \bar{\phi}_T(\theta_T)}{\partial \theta'_{\backslash 1}}\right)$) remain negligible in front of first-order terms. The fact that this condition is a byproduct of second-order Taylor expansions explains why no threshold like $T^{1/4}$ pops up in the linear case. In the non-linear case, the rule will then be as follows.

Golden rule for the test of identification strength in the non-linear case: When the subvector θ_1 is near-weakly identified, but the moment conditions are not affine w.r.t. the complementary subvector $\theta_{\backslash 1}$, considering any linear combination of $\theta_{\backslash 1}$ as unknown (in addition to θ_1) may impair global near-weak identification except if we assume that this linear combination is consistently estimated at a rate faster than $T^{1/4}$.

Following Antoine and Renault (2009), this latter property will be dubbed near-strong identification of the associated linear combination.

3 Testing identification strength

In this section, we are interested in assessing the identification strength of the structural parameter in order to detect weaker patterns of identification. Staiger and Stock (1997) propose a rule of thumb to detect weak instruments, whereas Stock and Yogo (2005) propose a formal characterization of the weakness of instruments based on the 2SLS bias as well as on the size of associated tests. Both Staiger and Stock (1997) and Stock and Yogo (2005) consider the null hypothesis that the instruments are weak, even though the parameters might be identified. Following these pioneer papers, we design two specifications tests which correspond respectively to the two golden rules stated in section 2. The null hypotheses will be designed in a way such that, failing to reject the null means that we have no sufficiently compelling evidence to trust an assumption of global near-weak identification. Accordingly, no standard asymptotic theory based on asymptotic normality is available and the researcher may resort to identification-robust procedures. Note also that both tests set the focus on the identification of subvector which is in contrast with existing procedures.
Even though they apply to two different settings (one with linearity w.r.t. the parameters under test, one without any linearity assumption at the cost of contemplating faster rates of convergence), the two testing strategies share a common structure: both amount to a conservative J-test questioning the rate of convergence of a given GMM estimator. This is the reason why we first build a general theoretical framework before discussing the feasibility of our tests in two more practically oriented sections.

3.1 Theoretical framework

Throughout section 3, null hypotheses under test are about the rate of convergence of a subset $\hat{\theta}_{\backslash 1T}$ of components of a given GMM estimator $\hat{\theta}_T$ defined according to definition 2.2. We maintain assumptions 2, 3, and 4, and in particular we know that

$$\sqrt{T} M_T^{-1} \left(\hat{\theta}_T - \theta^0 \right) = \mathcal{O}_P(1) \,.$$

Note however that we do not maintain assumption 1 of near-weak identification since it is precisely the focus of our interest. As done in section 2.3, we assume that only the subset θ_1 is near-weakly identified while the question is about the other parameters gathered in $\theta_{\backslash 1}$. In particular, the matrix M_T is block-diagonal and we do not know yet whether $\lim_T \left[M_{\backslash 1T}/\sqrt{T}\right] = 0$. We do not even know whether $\hat{\theta}_{\backslash 1T}$ is consistent.

To formulate a well-suited null hypothesis about the rate of convergence of $\hat{\theta}_{\backslash 1T}$, several remarks are in order.

(i) Following the practice that has been dominant since Staiger and Stock (1997), the null hypothesis sets the focus on the worse case scenario regarding the identification of $\theta_{\backslash 1}$, that is the rate of convergence of $\hat{\theta}_{\backslash 1T}$. Note however that our first extension w.r.t. the common practice is to set the focus on subvectors of θ. It should be clear that nothing prevents us from testing the identification of the whole parameter θ.

(ii) As stressed by the two golden rules of section 2, our worst case scenario regarding the rate of convergence a_T of $\hat{\theta}_{\backslash 1T}$ will be that, in the linear case, it is not even infinite whereas in the non-linear case it is slower than $T^{1/4}$.

(iii) As made explicit in the second golden rule, the worst case scenario of interest is actually that no linear combination of the parameters can be estimated at a satisfactory rate.

(iv) For a given GMM estimator $\hat{\theta}_T$ (and any given linear combination of $\hat{\theta}_{\backslash 1T}$), what really matters is not merely its rate of convergence but the rate of convergence of a well-suited subsequence. After all, a well-suited subsequence is able to properly identify the true unknown value of the linear combination of interest.

Therefore, for any real number $\nu \in [0, 1/2[$, we will generally consider null hypotheses of the following type:

> $H_0(\nu)$ *(No identification within* $\theta_{\backslash 1}$ *at rate faster than* ν*)*:
> For any subsequence of the estimator $\hat{\theta}_T$, for any deterministic sequence a_T such that $a_T/T^{\nu} \to \infty$, no non-zero linear combination of the subsequence $(\hat{\theta}_{\backslash 1T} - \theta^0_{\backslash 1})$ is $\mathcal{O}_P(1/a_T)$.

Note that, for sake of notational simplicity, we do not use an explicit notation like $\hat{\theta}_{m_T}$ for subsequences of $\hat{\theta}_T$. This abuse of notation will be maintained throughout. The key intuition for our proposed test of $H_0(\nu)$ comes from the following lemma, proved in the appendix.

Lemma 3.1. *(i) Under the null hypothesis $H_0(\nu)$, for any deterministic sequence a_T such that $a_T/T^\nu \to \infty$, we have*

$$\lim_T \left[\frac{\sqrt{T}}{a_T} M_{\backslash 1T}^{-1} \right] = 0 \,.$$

(ii) Under the alternative hypothesis to $H_0(\nu)$, under convenient regularity conditions, there exists a deterministic sequence a_T such that $a_T/T^\nu \to \infty$ and at least for a convenient subsequence

$$\lim_T \left\| \frac{\sqrt{T}}{a_T} M_{\backslash 1T}^{-1} \right\| = \infty \,.$$

As explained in the appendix, the required regularity condition amounts to the following application of Prohorov's theorem. By definition, under the alternative, we can find a deterministic sequence a_T with $a_T/T^\nu \to \infty$ such that, for some non-zero vector $\delta \in \mathbb{R}^{p_2}$, we have (for a well-suited subsequence)

$$a_T \delta' \left(\hat{\theta}_{\backslash 1T} - \theta^0_{\backslash 1} \right) = \mathcal{O}_P(1) \,,$$

that is,

$$a_T \gamma' (\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1}) = \mathcal{O}_P(1) \,,$$

for the non-zero vector $\gamma = R'\delta$. Then, since by our maintained assumption 2,

$$\sqrt{T} \Lambda_{\backslash 1T}^{-1} \left(\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1} \right) = \mathcal{O}_P(1) \,,$$

Prohorov's theorem tells us that $(\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1})$ (at least for a convenient subsequence) is endowed with an asymptotic distribution such that each component $\left(\hat{\eta}_{j,\backslash 1T} - \eta^0_{j,\backslash 1} \right)$ has a rate of convergence $\lambda_{j,\backslash 1T}/\sqrt{T}$ (with obvious notation for diagonal coefficients of $\Lambda_{\backslash 1T}$). Our regularity condition will amount to a non-degeneracy assumption about the joint limit distribution to ensure that $\gamma'(\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1})$ does not go to zero at a rate faster than the minimal rate, $\min_j \left[\lambda_{j,\backslash 1T}/\sqrt{T} \right]$. Otherwise, the proposed test would have no power against the alternative defined by the linear combination $\delta = R'^{-1}\gamma$.

Lemma 3.1 allows us to characterize the behavior of moment conditions computed at a conveniently distorted value of the GMM estimator $\hat{\theta}_T$. This distortion will depend on a deterministic sequence a_T and on a direction $\delta \in \mathbb{R}^{p_2}$. More precisely, we define the distorted estimator $\hat{\theta}_T^{a,\delta}$ as

$$\hat{\theta}_{1T}^{a,\delta} = \hat{\theta}_{1T} \quad \text{and} \quad \hat{\theta}_{\backslash 1T}^{a,\delta} = \hat{\theta}_{\backslash 1T} + \frac{\delta}{a_T} .$$

Then, the proposed test will be based on the comparison of norms of moment conditions computed as

$$J_T(\Omega) = T\overline{\phi}_T'(\hat{\theta}_T)\Omega_T\overline{\phi}_T(\hat{\theta}_T) \quad \text{and} \quad J_T^{a,\delta}(\Omega) = T\overline{\phi}_T'(\hat{\theta}_T^{a,\delta})\Omega_T\overline{\phi}_T(\hat{\theta}_T^{a,\delta}) .$$

where Ω_T is a sequence of symmetric matrices converging in probability towards a positive definite matrix Ω. Then, Lemma 3.1 allows us to show the following. As already explained through our two golden rules, the test of $H_0(\nu)$ will be especially relevant in the two following cases:

- Case i): Moment conditions affine w.r.t. θ and $\nu = 0$;
- Case ii): General (non-linear) moment conditions and $\nu = 1/4$.

Corollary 3.2. *(i) Under the null hypothesis $H_0(\nu)$, in case i) or ii) above, for any deterministic sequence a_T such that $a_T/T^{\nu} \to \infty$, we have for any $\delta \in \mathbb{R}^{p_2}$,*

$$\texttt{Plim}\left[J_T^{a,\delta}(\Omega) - J_T(\Omega)\right] = 0 .$$

(ii) Under the alternative hypothesis to $H_0(\nu)$, with convenient regularity conditions, there exists a deterministic sequence a_T such that $a_T/T^{\nu} \to \infty$ and a vector $\delta \in \mathbb{R}^{p_2}$ such that, at least for a convenient subsequence,

$$\texttt{Plim}\left[J_T^{a,\delta}(\Omega)\right] = \infty . \tag{3.1}$$

The convenient regularity conditions, made explicit in the appendix, are not really restrictive. They are implied in particular by the assumption that the moment conditions are affine w.r.t. $\theta_{\backslash 1}$. The key intuition is that when $\lim_T \left\| \frac{\sqrt{T}}{a_T} M_{\backslash 1T}^{-1} \right\| = \infty$ as in Lemma 3.1, we can be sure that $\lim_T \left\| \frac{\sqrt{T}}{a_T} M_{\backslash 1T}^{-1}\delta \right\| = \infty$ for generically all directions δ. Then, the result (3.1) follows by standard Taylor expansions (up to unlikely singularities of the Jacobian matrix introduced by non-linearities w.r.t. $\theta_{\backslash 1}$), knowing that $\sqrt{T}M_T^{-1}\left(\hat{\theta}_T - \theta^0\right) = \mathcal{O}_p(1)$.

3.2 Detecting near-weak identification in the non-linear case

As explained by our second golden rule, the general case of moment conditions that may not be affine with respect to the parameters under test $\theta_{\setminus 1}$ forces us to wonder whether some linear combinations of these parameters can be consistently estimated at a rate faster than $T^{1/4}$. In other words, we want to test the following null hypothesis (more precisely defined as $H_0(1/4)$ in section 3.1 above):

H_0 : No identification within $\theta_{\setminus 1}$ at rate faster than $T^{1/4}$.

As explained in the former subsection, we consider a well-suited distortion of a GMM estimator $\hat{\theta}_T$. For sake of expositional simplicity, we will assume throughout this subsection that $\hat{\theta}_T$ has been computed with an "efficient" weighting matrix, that is:

$$\hat{\theta}_T = \arg\min_{\theta} \left[T\overline{\phi}'_T(\theta) S_T^{-1} \overline{\phi}_T(\theta) \right] ,$$

where S_T stands for a consistent estimator of $S(\theta^0)$. In particular, such an estimator should be obtained from a first-step consistent estimator of θ^0. In other words, we implicitly maintain in this section that $\lim_T \left(M_T/\sqrt{T} \right) = 0$. As shown in the next subsection, this assumption can be relaxed in the linear case, at the price of a more involved approach. Extending this approach to the non-linear case should be straightforward and is not explicitly discussed.

For testing H_0, let us consider some deterministic sequence a_T such that $a_T/T^{1/4} \to \infty$. It will shortly become obvious that the slower the sequence $(a_T/T^{1/4})$ converges to infinity, the more powerful the resulting test will be; for instance, one may consider $a_T/T^{1/4} = \log(\log T)$, or an even slower sequence. As in the former subsection, this sequence a_T is used to build a distorted version $\hat{\theta}_T^{a,\delta}$ of the GMM estimator $\hat{\theta}_T$:

$$\hat{\theta}_{1T}^{a,\delta} = \hat{\theta}_{1T} \text{ and } \hat{\theta}_{\setminus 1T}^{a,\delta} = \hat{\theta}_{\setminus 1T} + \frac{\delta}{a_T} , \tag{3.2}$$

where δ is a given deterministic vector of size p_2. Our asymptotic conservative test for H_0 will be based on the corresponding distorted J-test statistic for overidentification

$$J_T^{a,\delta} = T\overline{\phi}'_T(\hat{\theta}_T^{a,\delta}) S_T^{-1} \overline{\phi}_T(\hat{\theta}_T^{a,\delta}) . \tag{3.3}$$

We can show the following result.

Theorem 3.3. *(Test of near-weak identification of $\theta_{\backslash 1}$ in the non-linear case)*
For an arbitrary choice of a deterministic sequence a_T such that $a_T/T^{1/4} \to \infty$ and of a vector $\delta \in \mathbb{R}^{p_2}$, we define the asymptotic test with critical region $W_T^{a,\delta}$,

$$W_T^{a,\delta} = \left\{ J_T^{a,\delta} > \chi^2_{1-\alpha}(K - p_1) \right\} ,$$

where $\chi^2_{1-\alpha}(K-p_1)$ is the $(1-\alpha)$-quantile of the chi-square distribution with $(K-p_1)$ degrees of freedom.
(i) Under assumptions 2 to 4, and assuming that θ_1 is near-weakly identified, the test $W_T^{a,\delta}$ is asymptotically conservative at level α for the null hypothesis H_0 of "no identification within $\theta_{\backslash 1}$ at rate faster than $T^{1/4}$".
(ii) The test $W_T^{a,\delta}$ is consistent against any alternative that makes the choice (a_T, δ) conformable to (3.1).

Of course, the consistency claim above is somewhat tautological. The important point is to remember that Lemma 3.1 and Corollary 3.2 have shown that, under the alternative hypothesis, we are likely to be successful in our choice of the pair (a_T, δ). The key intuition is that under the alternative $\left\| \sqrt{T} M_{\backslash 1T}^{-1} \right\|$ goes to infinity at a rate faster than $T^{1/4}$. Our main task is to pin down a rate a_T strictly between this rate and $T^{1/4}$. In finite samples, this bandwidth choice takes a data-based selection rule that will be described shortly. Let us first explain why the test cannot be oversized asymptotically. We have shown in the former subsection that, under the null,

$$J_T^{a,\delta} - J_T = o_P(1)$$

where

$$J_T = T\overline{\phi}'_T(\hat{\theta}_T) S_T^{-1} \overline{\phi}_T(\hat{\theta}_T)$$

is the standard J-test statistic for overidentification. By definition, we have

$$J_T \leq T\overline{\phi}'_T(\overline{\theta}_T) S_T^{-1} \overline{\phi}_T(\overline{\theta}_T) ,$$

where $\overline{\theta}_T$ is the (infeasible) GMM estimator computed when the components of $\theta_{\backslash 1}$ are fixed at their true (unknown) value $\theta^0_{\backslash 1}$. Under the maintained assumption that θ_1 is near-weakly identified, we know by Theorem 2.3 that $\left[T\overline{\phi}'_T(\overline{\theta}_T) S_T^{-1} \overline{\phi}_T(\overline{\theta}_T) \right]$ converges in distribution towards a chi-square distribution with $(K - p_1)$ degrees of freedom. Therefore, under the null,

$$\lim_T P(W_T^{a,\delta}) = \lim_T P\left(\left\{J_T > \chi^2_{1-\alpha}(K-p_1)\right\}\right) \leq \alpha\,,$$

and the test is asymptotically conservative at level α as announced.

As far as finite samples performance of the test $W_T^{a,\delta}$ is concerned, the key is for a given choice of the sequence a_T (such that $a_T/T^{1/4}$ converges slowly to infinity) to elicit a vector $\delta \in \mathbb{R}^{p_2}$ with a well-tuned length. We propose to select δ by subsampling. We consider all the subsamples of $\lfloor T^\nu \rfloor$ consecutive observations[7] (with ν given and $0 < \nu < 1$). For each such subsample s, we consider a grid of dimension p_2 that contains candidates for δ, say δ_m. For each such candidate, we consider the associated local-to-zero version of the estimator $\hat{\theta}_{T,s}^{a,\delta_m}$ and the associated test statistic $J_{T,s}^{a,\delta_m}$ defined respectively in (3.2) and (3.3),

$$\begin{aligned}
\hat{\theta}^{a,\delta_m}_{\lfloor T^\nu \rfloor,s} &= \hat{\theta}_{\lfloor T^\nu \rfloor,s} + \begin{pmatrix} \mathbf{0}_{p_1} \\ \delta_m / a_{\lfloor T^\nu \rfloor} \end{pmatrix}, \\
J^{a,\delta_m}_{\lfloor T^\nu \rfloor,s} &= \lfloor T^\nu \rfloor \overline{\phi}'_{\lfloor T^\nu \rfloor,s}(\hat{\theta}^{a,\delta_m}_{\lfloor T^\nu \rfloor,s}) S_T^{-1} \overline{\phi}_{\lfloor T^\nu \rfloor,s}(\hat{\theta}^{a,\delta_m}_{\lfloor T^\nu \rfloor,s})\,.
\end{aligned}$$

As a result, for each grid point δ_m, we obtain a cross-sectional distribution of the test statistic (3.3), say $(J^{a,\delta_m}_{\lfloor T^\nu \rfloor,s})_{s=1,\cdots,S}$. We can then extract the $(1-\alpha^*)$-quantile of the test statistic, for some user-chosen α^*. We select the perturbation vector δ_{m^*} associated with the $(1-\alpha^*)$-quantile the closest to the $(1-\alpha^*)$-quantile of the chi-square distribution with $(K - p_1)$ degrees of freedom. Note that $(1-\alpha^*)$ may, or may not correspond to the actual asymptotic size of the designed test. Regardless of the chosen $(1-\alpha^*)$ and associated perturbation vector δ_{m^*}, the asymptotic size of the test $(1-\alpha)$ is always controlled as shown above. In our Monte-Carlo experiments, we chose $\alpha = \alpha^*$ and our results were not too sensitive to this choice. See additional discussions for practical implementation in Appendix B where we also propose a data-based procedure to design a grid of candidate points δ_m.

[7] $\lfloor T^\nu \rfloor$ refers to the largest integer below T^ν. We consider consecutive observations to accommodate possible serial dependencies.

3.3 Testing weak identification in the linear case

As explained by our first golden rule, the case where moment conditions are affine with respect to the parameter θ allows us to wonder whether some linear combinations of the parameters under test $\theta_{\backslash 1}$ can be consistently estimated. In other words, we want to test the null $\theta_{\backslash 1}$ is only weakly identified (more precisely defined as $H_0(0)$ in section 3.1 above):

H_0: $\theta_{\backslash 1}$ is only weakly identified.

Our testing procedure in the linear case is somewhat similar to the procedure described in the previous section. More specifically, we consider a sequence a_T such that $a_T \to \infty$ and a deterministic vector δ to build a distorted version $\hat{\theta}_T^{a,\delta}$ of the GMM estimator $\hat{\theta}_T$ as in (3.2). Our asymptotic conservative test for H_0 is then based on the corresponding distorted J-test statistic $J_T^{a,\delta}$ as in (3.3). We refer the interested reader to section 3.2 and Appendix B. We can show the following result.

Theorem 3.4. *(Test of weak identification of $\theta_{\backslash 1}$)*
For an arbitrary choice of a deterministic sequence a_T such that $a_T \to \infty$ and of a vector $\delta \in \mathbb{R}^{p_2}$, we define the asymptotic test with critical region $W_T^{a,\delta}$

$$W_T^{a,\delta} = \left\{ J_T^{a,\delta} > \chi^2_{1-\alpha}(K - p_1) \right\}$$

where $\chi^2_{1-\alpha}(K-p_1)$ is the $(1-\alpha)$-quantile of the chi-square distribution with $(K-p_1)$ degrees of freedom.
(i) Under assumptions 2 to 4, and assuming that θ_1 is near-weakly identified, the test $W_T^{a,\delta}$ is asymptotically conservative at level α for the null hypothesis H_0 of "weak identification within $\theta_{\backslash 1}$".
(ii) The test $W_T^{a,\delta}$ is consistent against any alternative that makes the choice (a_T, δ) conformable to (3.1).

It is interesting to point out that in order to obtain a procedure that is less conservative, parameters known not to be weakly identified (e.g. the intercept) should not be included in $\theta_{\backslash 1}$ but rather in θ_1.
As highlighted in section 3.2, the above procedure crucially depends on a consistent estimator S_T of $S(\theta^0)$ to define the efficient GMM estimator $\hat{\theta}_T$; the existence of S_T is guaranteed

whenever $\lim_T \left(M_T/\sqrt{T} \right) = 0$. In this section, we relax this assumption. Hence, under H_0, there is no obvious (consistent) estimator of $S(\theta^0)$ since there is no first-step consistent estimator of θ^0. We propose the following testing procedure which is robust to inconsistent estimators of $S(\theta^0)$:
(i) build a confidence region for θ^0; we call it C_T and it is based on the test statistic of Stock and Wright ψ_T; If C_T is empty, the null is rejected;
(ii) if C_T is not empty, then we minimize the following test statistic

$$T\overline{\phi}_T'(\hat{\theta}_T^{a,\delta})S_T^{-1}(\theta)\overline{\phi}_T(\hat{\theta}_T^{a,\delta})$$

wrt to $\theta \in C_T$ and compare with the appropriate quantile.

We first introduce a few notations:
(i) GMM criterion function for some given positive definite matrix S_T^{-1}: $Q_T(\theta) = T\overline{\phi}_T'(\theta)S_T^{-1}\overline{\phi}_T(\theta)$.
- $\hat{\theta}_T$ is the feasible GMM estim. that minimizes the (usual) GMM criterion over the whole parameter vector θ.
- $\overline{\theta}_T$ is the infeasible GMM estim. that minimizes the (usual) GMM criterion but only wrt to subvector θ_1 when $\theta_{\backslash 1}$ is fixed at its true (unknown) value $\theta_{\backslash 1}^0$.
By definition:

$$Q_T(\hat{\theta}_T) \leq Q_T(\overline{\theta}_T)$$

and also

$$Q_T(\overline{\theta}_T) \xrightarrow{T} \chi^2(K - p_1)$$

whenever θ_1 is near-weakly identified and S_T is a consistent estimator of $S(\theta^0)$.

(ii) Similarly I define 1 more estimator related to the AR-type statistic of Stock and Wright (now optimization also involves the weighting matrix):
- ψ criterion of Stock and Wright: $\psi_T(\theta) = T\overline{\phi}_T'(\theta)S_T^{-1}(\theta)\overline{\phi}_T(\theta)$.
- $\hat{\hat{\theta}}_T$ is the feasible estim. that minimizes the (usual) ψ criterion over the whole parameter vector θ.

By definition:

$$\psi_T(\hat{\hat{\theta}}_T) \leq \psi_T(\theta^0)$$

and also

$$\psi_T(\theta^0) \overset{T}{\to} \chi^2(K)$$

I can use this result to build a confidence region for θ as follows:

$$C_T(1-\zeta) = \{\theta \in \Theta \,/\, \psi_T(\theta) \leq \chi^2_K(1-\zeta)\}, \ .$$

The asymptotic coverage of $C_T(1-\zeta)$ is $(1-\zeta)$, that is

$$P\left(\theta^0 \in C_T(1-\zeta)\right) \overset{T}{\to} 1-\zeta\,.$$

Since θ^0 has (asymptotic) probability $(1-\zeta)$ to be in C_T, we have

$$\inf_{\theta \in C_T(1-\zeta)} \left[T\overline{\phi}'_T(\hat{\theta}^{a,\delta}_T)S^{-1}_T(\theta)\overline{\phi}_T(\hat{\theta}^{a,\delta}_T)\right] \leq T\overline{\phi}'_T(\hat{\theta}^{a,\delta}_T)S^{-1}_T(\theta^0)\overline{\phi}_T(\hat{\theta}^{a,\delta}_T)\,,$$

with probability approaching $(1-\zeta^*) \geq (1-\zeta)$ since the above inequality may be true even if θ^0 does not belong to C_T.

From Corollary 3.2(i), the right-hand side of the previous inequality is such that, under H_0,

$$T\overline{\phi}'_T(\hat{\theta}^{a,\delta}_T)S^{-1}_T(\theta^0)\overline{\phi}_T(\hat{\theta}^{a,\delta}_T) = T\overline{\phi}'_T(\hat{\theta}_T)S^{-1}_T(\theta^0)\overline{\phi}_T(\hat{\theta}_T) + o_p(1)\,.$$

Since $S_T(\theta^0)$ is a consistent estimator of $S(\theta^0)$, we can use the result (i) above, that is:

$$T\overline{\phi}'_T(\hat{\theta}_T)S^{-1}_T(\theta^0)\overline{\phi}_T(\hat{\theta}_T) \leq T\overline{\phi}'_T(\overline{\theta}_T)S^{-1}_T(\theta^0)\overline{\phi}_T(\overline{\theta}_T) \overset{d}{\to} \chi^2(K-p_1)\,.$$

Hence, we have under H_0,

$$\inf_{\theta \in C_T(1-\zeta)} \left[T\overline{\phi}'_T(\hat{\theta}^{a,\delta}_T)S^{-1}_T(\theta)\overline{\phi}_T(\hat{\theta}^{a,\delta}_T)\right] \leq T\overline{\phi}'_T(\overline{\theta}_T)S^{-1}_T(\theta^0)\overline{\phi}_T(\overline{\theta}_T)\,,$$

with probability approaching $(1-\zeta^{**}) \geq (1-\zeta^*)$. This inequality can then be used to build a conservative testing procedure. More specifically, our test of H_0 is based on the following decision rule for some chosen $\alpha \in (0,1)$:

Reject H_0 if:

(i) $C_T(1-\zeta)$ is empty;

This happens with some non-zero probability ϵ.

or (ii) $\inf_{\theta \in C_T(1-\zeta)} \left[T \overline{\phi}'_T(\hat{\theta}_T^{a,\delta}) S_T^{-1}(\theta) \overline{\phi}_T(\hat{\theta}_T^{a,\delta}) \right] > \chi^2_{K-p_1}(1-\alpha)$.

This happens with probability $(1-\epsilon)\left[\zeta^{**} + (1-\zeta^*)(\alpha^*)\right]$ where $(1-\alpha^*) \geq (1-\alpha)$ because it is a conservative asymptotic test.

The probability of rejecting the null is then equal to:

$$\begin{aligned} & \epsilon + (1-\epsilon)\left[\zeta^{**} + (1-\zeta^{**})\alpha^*\right] \\ = & \epsilon + (1-\epsilon)\left[\zeta^{**} + \alpha^* - \zeta^{**}\alpha^*\right] \end{aligned}$$

When ϵ is small enough, the above probability is not too far from

$$\begin{aligned} & \zeta^{**} + \alpha^* - \zeta^{**}\alpha^* \\ = & \alpha^* + \zeta^{**}(1-\alpha^*) \\ \leq & \alpha + \zeta^{**}(1-\alpha^*) \\ \leq & \alpha + \zeta(1-\alpha^*) \end{aligned}$$

Note that the smaller ζ is, the larger the confidence set $C_T(1-\zeta)$ is, and the smaller the probability of $C_T(1-\zeta)$ to be empty (ie ϵ) is. As a result, by choosing ζ small enough, we are able to show that unconditional asymptotic size of the above test cannot exceed $(\alpha+\delta)$, where $\delta = \zeta(1-\alpha^*)$ is small.

This procedure is related to the projection method discussed in Chaudhuri and Zivot (2011). Note also that it does not rely at all on the linearity of moment conditions. Hence, extending this approach to the non-linear case of section 3.2 after relaxing the assumption $\lim_T (M_T/\sqrt{T}) = 0$ should be straightforward and is not explicitly discussed.

4 Monte-Carlo evidence

In this section, we use Monte-Carlo methods to illustrate the finite samples properties of the tests introduced in section 3. We consider a standard linear IV regression model with one intercept and one endogenous regressor, as well as a (non-linear) diffusion process with continuous record and increasing time span asymptotic.

4.1 Linear IV regression model

Consider the following standard linear IV regression model with one intercept and one endogenous regressor,

$$\begin{aligned} y_t &= \alpha_0 + Y_{1t}\beta_0 + h(X_t)\varepsilon_t\,, \\ Y_{1t} &= X_t'\Pi_x + U_i\,, \end{aligned} \tag{4.1}$$

where Y_{1t} is a univariate endogenous regressor, while X_t is a vector of L_x (exogenous) instrumental variables that follows a standard normal distribution. (ε_t, U_t) is normally distributed and independent of X_t. We set $\theta^0 = (\alpha_0\ \beta_0)' = (0\ 0)'$. We consider two versions of the model: a homoskedastic model with $h(x) = 1$ and a heteroskedastic model $h(x) = \sqrt{(1 + (e'x)^2)/(L_x + 1)}$ where e is the vector of ones of size L_x. In both, $(h(X_t)\varepsilon_t, U_t)$ has mean $\mathbf{0}$, unit unconditional variances, and unconditional correlation ρ. Π_x is proportional to the vector e and is related to the first stage R^2 by,

$$R_x^2 = \frac{\Pi_x'\Pi_x}{\Pi_x'\Pi_x + 1}\,.$$

It is worth pointing out that the intercept parameter is always strongly-identified, while the slope parameter is more or less weakly identified depending on the value of R_x^2.
In this experiment, we are interested in testing the strength of identification. We compare the performance of the following tests of weak identification: (i) the rule-thumb based on the first-stage F-statistic proposed by Staiger and Stock (1997); (ii) the test based on the 10%-bias of 2SLS proposed by Stock and Yogo[8] (2005); and two versions of the test proposed in section 3.3: (iii) the joint test on the whole parameter θ; (iv) the test on the subvector β. We consider the two specifications described above for sample size $T = 250$, (high) degree of endogeneity $\rho = 0.8$, and two values for Π_x such that $R_x^2 = 0.14$ and 0.01 respectively. We consider three instrumental variables, the constant and a bivariate X. Our results are

[8] This is the version of the test which is commonly used. We also report results for alternate versions of the test proposed by Stock and Yogo (2005) based on 5% bias, as well as 10% and 15% size distortion.

reported in Tables 1 to 4. In each case, we report the first four moments of the Monte-Carlo distribution of the standardized GMM estimator. The reader can then assess how far the Monte-Carlo distribution of the standardized GMM estimator is from its asymptotic approximation for which we expect mean 0, variance 1, skewness 0, and kurtosis 3. We also report the nominal 5% rejection frequencies of the tests described above. All simulations results are based on 5,000 replications.

Tables 1 and 2 report the performance of the above tests when R_x^2 is large (0.14) for a heteroskedastic and a homoskedastic model respectively. As expected, in both cases, the distribution of the GMM estimator of the intercept is well-approximated by the asymptotic one. However, the distribution of the estimator of the slope is slightly biased and skewed. This suggests that this estimator is not as strongly identified as the estimated intercept. Our joint test rejects the null hypothesis of weak identification with probability 0.14, whereas our test based only on the slope rejects with probability 0.05. Our joint test has some (limited) power to reject weak identification that comes from the strongly identified intercept parameter, whereas our test based only on the slope clearly indicates that the empirical evidence is not sufficient to reject weak identification. These results confirm the features highlighted in the Monte-Carlo distributions of the GMM estimators of the intercept and slope as described above. Competitive testing procedures lead to mixed results. The rule-of-thumb of Staiger and Stock rejects (global) weak identification with probability 0.50, whereas the different tests of Stock and Yogo reject with probabilities ranging between 0 to 0.60.

Tables 3 and 4 report the performance of the above tests when R_x^2 is small (0.01) for a heteroskedastic and a homoskedastic model respectively. As expected, the distribution of the GMM estimator of the intercept is still well-approximated by the asymptotic one, whereas the distribution of the estimator of the slope is quite biased and skewed, departing a lot from the asymptotic one. Our joint test rejects the null hypothesis of weak identification with probability around 0.20, whereas our test based only on the slope rejects with probability 0.08. Here again, our joint test has some power to reject weak identification that comes from the strongly identified intercept parameter. The rule-of-thumb of Staiger and Stock and the different tests of Stock and Yogo reject (global) weak identification with probability almost 0.

To conclude, our testing procedure is able to detect weak identification reliably by focusing on the specific component at stake, namely the slope parameter. Competing procedures can only hope to detect (global) weak identification which can be misleading as highlighted above.

4.2 Diffusion process with continuous record and increasing time span asymptotic

Consider the following continuous time Ornstein-Uhlenbeck process

$$dy_t = (\theta_0 - \theta_1 y_t)dt + \theta_2 dW_t \qquad \text{with} \qquad dW_t \overset{iid}{\sim} \mathcal{N}(0, dt)\,,$$

where $\theta_0/\theta_1 > 0$ represents the long run (unconditional) mean, $\theta_1 > 0$ captures the speed of the mean reversion, and $\theta_2 > 0$ gives the constant volatility of the process. It is well-known that its exact solution is the following discrete time AR(1) process

$$y_t = a + b y_{t-\Delta} + \sqrt{c}\epsilon_t\,, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)\,, \tag{4.2}$$

$$\text{with} \qquad a = \frac{\theta_0}{\theta_1}(1 - e^{-\theta_1 \Delta})\,, \qquad b = e^{-\theta_1 \Delta}\,, \qquad c = \theta_2^2 \left(\frac{1 - e^{-2\theta_1 \Delta}}{2\theta_1} \right)\,.$$

For simplicity, the parameters θ_0 and θ_2 are assumed to be known throughout, and are fixed at their true values in the structural model, while only the parameter θ_1 is estimated.
Suppose that n observations of (4.2) are available for $t = \Delta, \cdots, n\Delta$ with $T \equiv n\Delta$. Define the associated OLS estimators of the three parameters, a, b, and c, respectively $\hat{a}_{n,ols}$, $\hat{b}_{n,ols}$, and $\hat{c}_{n,ols}$. For fixed Δ, the usual asymptotic result for OLS estimators holds, and we have

$$\sqrt{n} \begin{bmatrix} \hat{a}_{n,ols} - a(\theta_1) \\ \hat{b}_{n,ols} - b(\theta_1) \\ \sqrt{\hat{c}_{n,ols}} - \sqrt{c(\theta_1)} \end{bmatrix} \overset{d}{\to} \mathcal{N}(0, \Sigma(\Delta)) \qquad \text{with} \qquad \Sigma(\Delta) = \begin{pmatrix} cE\left[(X_i X_i')\right]^{-1} & 0 \\ 0 & c/2 \end{pmatrix}\,,$$

where X_i' represents the i-th row of the matrix X. Our estimation procedure for θ_1 relies on the (overidentified) GMM estimation with three moment conditions,

$$\hat{\theta}_{1,n} = \arg\min_{\theta_1} [\phi(\theta_1)'\Omega_n \phi(\theta_1)]$$

$$\text{with} \quad \phi(\theta_1) = \frac{1}{\Delta}\begin{bmatrix} \hat{a}_{n,ols} - a(\theta_1) \\ \hat{b}_{n,ols} - b(\theta_1) \\ \sqrt{\hat{c}_{n,ols}} - \sqrt{c(\theta_1)} \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} \hat{a}_{n,ols} - \frac{\theta_0}{\theta_1}(1 - e^{-\theta_1 \Delta}) \\ \hat{b}_{n,ols} - e^{-\theta_1 \Delta} \\ \sqrt{\hat{c}_{n,ols}} - \sqrt{\theta_2^2 \left(\frac{1-e^{-2\theta_1 \Delta}}{2\theta_1}\right)} \end{bmatrix},$$

where Ω_n is a sequence of symmetric positive definite random matrices of size 3 converging towards a positive definite matrix Ω. In Appendix B.2, we show that each moment condition has a different identification strength controlled by Δ. More precisely, if we consider the three (just-identified) estimators obtained from the GMM estimation based on each moment condition separately, we get that, when $\Delta \to 0$ and $T \to \infty$:
- the estimator based on condition 2 converges at rate $\sqrt{T}$;
- the estimator based on condition 3 converges at rate $\sqrt{\Delta}\sqrt{T}$, with $\sqrt{\Delta}\sqrt{T} = o(\sqrt{T})$;
- the estimator based on condition 1 converges at rate $\Delta\sqrt{T}$, with $\Delta\sqrt{T} = o(\sqrt{\Delta}\sqrt{T})$.

Throughout, the following notations are used to distinguish the different estimators of θ_1 we consider:
- $\hat{\theta}_{all}$ refers to the (overidentified) GMM estimator based on the three moment conditions;
- $\hat{\theta}_{\backslash j}$ refers to the (overidentified) GMM estimator based on two moment conditions only, after condition j has been removed.

In this experiment, we are interested in testing the strength of identification. In our simple framework, we know that, asymptotically, the strongest moment condition dictates the rate of convergence of the associated estimator of θ_1.[9] In our experiment, we fix the time span T and vary the strength of identification by decreasing Δ (accordingly, n increases). Smaller values of Δ correspond to cases where the identification strength is weaker. The nominal size of the tests is 5%. Our results are displayed in Table 5. For each test, we provide the estimator of θ_1 being considered, as well as the associated Monte-Carlo rejection probability. Recall that rejection of the null hypothesis means that the estimator is sufficiently

[9] In general, this is not the case as discussed in Section 2. However, in our simple framework, there is only one parameter to identify, so it necessarily inherits the identification strength of the strongest moment condition available.

strongly identified for standard asymptotic results to hold. Further details regarding the implementation of this experiment are provided in Appendix B.2.

First, we consider the estimator $\hat{\theta}_{all}$ based on all moment conditions. As discussed above, this estimator is always strongly identified due to the moment condition 2. As a result, we expect our test to often be rejected regardless of the value of the parameter Δ. This is exactly what happens: the associated rejection probabilities are equal to 1 irrespective of the identification strength.

Second, we consider the estimator $\hat{\theta}_{\setminus 3}$. The presence of the moment condition 2 guarantees that this estimator is always strongly identified. And we also expect our test to often be rejected regardless of the value of the parameter Δ. The associated rejection probabilities are actually equal to 1 in all cases.

Finally, we consider the estimator based on moment conditions 1 and 3 only, $\hat{\theta}_{\setminus 2}$. As discussed above, this estimator is identified at rate $\sqrt{\Delta}\sqrt{T}$ due to the moment condition 3. As a result, we expect our test not to be rejected for sufficiently small values of Δ. The associated rejection probabilities are quite small in all cases, even for larger values of Δ. This suggests that the identifying power of (missing) condition 2 is stronger than the one of the two other (included) conditions.

To conclude, our test performs relatively well.

5 Conclusion

We have considered models defined by a set of moment restrictions that may be subject to weak identification. Recently, the strength of identification of the structural parameters has been reflected by the Jacobian of the moment conditions, and our first contribution was to unify several characterizations of identification previously given in the literature. Accordingly, we have defined near-weak identification, and we have also shown that it is key to deliver standard asymptotic normality of GMM estimators, albeit at rates of convergence slower than usual for different linear combinations of such estimator.

In this setup, we have proposed two tests to assess the identification strength of the parameters. First, we have proposed a test to detect weak identification in linear settings. Second, we have proposed a test to detect parameters for which no standard asymptotic theory is available. Both testing procedures relied on a conservative overidentification test computed at a properly distorted GMM estimator. We have also highlighted how subsampling can easily be used in practice to get such appropriate distortions. Both tests are straightforward to apply and we have discussed why we expect such simple tests to have good power properties.

Finally, we have illustrated the finite samples performance of our tests through Monte-Carlo simulations. The linear IV regression model and a (persistent) AR(1) model calibrated to interest rate data were considered. In both cases, we have shown that our tests are well-behaved compared to contenders, both in terms of size and power.

To conclude, given the simplicity of the above tests and their good power properties, we believe that practitioners may benefit from using them.

References

[1] D.W.K. Andrews, *Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity*, Econometrica **62** (1994), 43–72.

[2] B. Antoine and E. Renault, *Efficient GMM with Nearly-weak Instruments*, The Econometrics Journal **12** (2009), 135–171.

[3] ———, *Efficient Inference with Poor Instruments: a General Framework*, ch. 2 in Handbook of Empirical Economics and Finance, pp. 29–70, edited by A. Ullah and D.E. Giles, Taylor & Francis, 2010.

[4] ———, *Efficient Minimum Distance Estimation with Multiple Rates of Convergence*, forthcoming Journal of Econometrics (2012).

[5] M. Caner, *Testing, Estimation in GMM and CUE with Nearly-Weak Identification*, Econometric Reviews **29** (2010), 330–363.

[6] S. Chaudhuri and E. Zivot, *A new method of projection-based inference in GMM with weakly identified nuisance parameters*, Journal of Econometrics **164** (2011), 239–251.

[7] R. Dridi, A. Guay, and E. Renault, *Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models*, Journal of Econometrics **136** (2007), 397–430.

[8] J.-M. Dufour, *Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models*, Econometrica **65** (1997), 1365–1388.

[9] J.-M. Dufour and P. Valéry, *Wald-type tests when rank conditions fail: a smooth regularization approach*, Working paper (2011).

[10] J. Hahn and J. Hausman, *A new Specification Test for the Validity of Instrumental Variables*, Econometrica **70** (2002), no. 1, 163–190.

[11] ———, *Weak Instruments: Diagnosis and Cures in Empirical Econometrics*, American Economic Review **93** (2003), 118–125.

[12] J. Hahn and G. Kuersteiner, *Discontinuities of Weak Instruments limiting Distributions*, Economics Letters **75** (2002), 325–331.

[13] C. Han and P.C.B. Phillips, *GMM with Many Moment Conditions*, Econometrica **74** (2006), no. 1, 147–192.

[14] C. Hansen, J. Hausman, and W. Newey, *Estimation with Many Instrumental Variables*, Journal of Business and Economic Statistics **26** (2008), 398–422.

[15] L.P. Hansen, *Large Sample Properties of Generalized Method of Moments Estimators*, Econometrica **50** (1982), no. 4, 1029–1054.

[16] R.A. Horn and C.R. Johnson, *Matrix Analysis*, 1985.

[17] F. Kleibergen, *Testing Parameters in GMM without assuming that they are identified*, Econometrica **73** (2005), 1103–1123.

[18] F. Kleibergen and S. Mavroeidis, *Inference on subsets of parameters in GMM without assuming identification*, Working paper, Brown University (2009).

[19] W.K. Newey and K.D. West, *A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*, Econometrica **55** (1987), 703–708.

[20] W.K. Newey and F. Windmeijer, *GMM with Many Weak Moment Conditions*, Econometrica **77** (2009), 687–719.

[21] Sargan, *The Estimation of Economic Relationships Using Instrumental Variables*, Econometrica **26** (1958), 393–415.

[22] D. Staiger and J. Stock, *Instrumental Variables Regression with Weak instruments*, Econometrica **65** (1997), 557–586.

[23] J.H. Stock and J.H. Wright, *GMM with Weak Identification*, Econometrica **68** (2000), no. 5, 1055–1096.

[24] J.H. Stock and M. Yogo, *Asymptotic Distributions of Instrumental Variables Statistics With Many Instruments*, ch. 8 in Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, pp. 109–120, edited by D.W.K. Andrews and J.H. Stock, Cambridge University Press, Cambridge, U.K., 2005.

[25] J. Wright, *Detecting Lack of Identification in GMM*, Econometric Theory **19** (2003), 322–330.

A Proofs of the main results

Notations:

- For any vector v with element $(v_i)_{1\leq i\leq H}$, we define: $\|v\|^2 = \sum_{i=1}^H v_i^2$.
- $[M]_{k.}$ denotes the k-th row of matrix M.

Proof of Theorems 2.1 and 2.2 *(Asymptotic normality of GMM estimator)*:
A mean-value expansion of the moment conditions around θ^0 for $\tilde{\theta}_T$ between $\hat{\theta}_T$ and θ^0 gives

$$\overline{\phi}_T(\hat{\theta}_T) = \overline{\phi}_T(\theta^0) + \frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'}(\hat{\theta}_T - \theta^0)\,. \tag{A.1}$$

Combined with the first-order conditions,

$$\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\overline{\phi}_T(\hat{\theta}_T) = 0\,,$$

this yields to

$$\begin{aligned} & \frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\overline{\phi}_T(\theta^0) + \frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'}(\hat{\theta}_T - \theta^0) = 0 \\ \Leftrightarrow\quad & M'_T\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\sqrt{T}\overline{\phi}_T(\theta^0) + M'_T\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'}M_T M_T^{-1}(\hat{\theta}_T - \theta^0) = 0\,. \end{aligned} \tag{A.2}$$

Under the near-weak identification assumption 1, we have

$$\begin{aligned} & \frac{\partial\overline{\phi}_T(\hat{\theta}_T)}{\partial\theta'}M_T \xrightarrow{P} \Gamma(\theta^0) \qquad \text{and} \qquad \frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'}M_T \xrightarrow{P} \Gamma(\theta^0) \\ \Rightarrow\quad & M'_T\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'}M_T \xrightarrow{P} \Gamma'(\theta^0)\Omega\Gamma(\theta^0) \\ \Rightarrow\quad & M'_T\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\frac{\partial\overline{\phi}_T(\hat{\theta}_T)}{\partial\theta'}M_T \qquad \text{is invertible for } T \text{ large enough.} \end{aligned}$$

Combined with (A.2), we get

$$M_T^{-1}\sqrt{T}(\hat{\theta}_T - \theta^0) = -\left[M'_T\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'}M_T\right]^{-1} M'_T\frac{\partial\overline{\phi}'_T(\hat{\theta}_T)}{\partial\theta}\Omega_T\sqrt{T}\overline{\phi}_T(\theta^0)\,.$$

The CLT assumption 3 and a standard argument for optimality allow to conclude. ■

Justification of assumption 4 *(Structure of matrix M_T)*:
A singular value decomposition of the matrix M_T allows us to write:

$$M_T = W_T \Lambda_T V_T',$$

where Λ_T is a diagonal matrix with diagonal coefficients equal to the square-roots of the eigenvalues of the matrices $M_T M_T'$ and $M_T' M_T$ according to the diagonalization formulas:

$$M_T M_T' = W_T \Lambda_T^2 W_T' \qquad \text{and} \qquad M_T' M_T = V_T \Lambda_T^2 V_T',$$

where W_T (respectively V_T) is an orthogonal matrix of eigenvectors of $M_T M_T'$ (respectively $M_T' M_T$).
We then justify assumption 4 by making clear that, insofar as one is ready to maintain the assumption of near-weak identification, that is the properties of the sequence of matrices M_T as listed in definition 2.3, it does not restrict much the generality to assume, in addition, the existence of a fixed nonsingular matrix R such that $M_T = R\Lambda_T$, with Λ_T as defined above.
The acronym "w.l.o.g." used below stands for "without loss of generality".

1st step: Up to considering only a subsequence, we can assume w.l.o.g. that $M_T = W_T \Lambda_T$.
Definition 2.3 stipulates that the sequence M_T must fulfill two sets of conditions: (i) $M_T/\sqrt{T} \xrightarrow{T} 0$; (ii) for a set of random matrices J_T (for simplicity, the dependence on a choice of an estimator of θ is not made explicit here), $J_T M_T$ should converge towards a full-column rank matrix Γ.
Thus, we want to show that, up to considering only a subsequence, the sequence of matrices $M_T = W_T \Lambda_T V_T'$ can be replaced by $M_T^* = W_T \Lambda_T = M_T V_T$ without modifying the aforementioned conditions.
It is well known that the group of real orthogonal matrices is compact (see Horn and Johnson (1985, p 71)). Thus, this set is bounded and in particular:

$$\lim_T \frac{M_T}{\sqrt{T}} = 0 \;\Rightarrow\; \lim_T \frac{M_T V_T}{\sqrt{T}} = 0,$$

and:

$$\lim_T [J_T M_T - \Gamma] = 0 \;\Rightarrow\; \lim_T [J_T M_T - \Gamma] V_T = 0\,.$$

In other words, we have the two conditions:

$$\lim_T \frac{M_T^*}{\sqrt{T}} = 0 \qquad \text{and} \qquad \lim_T [J_T M_T^* - \Gamma V_T] = 0\,.$$

Moreover, since the sequence (V_T) takes its values in the compact set of real orthogonal matrices, it exists a subsequence converging towards some orthogonal matrix V. In other words, up to considering some subsequence (for simplicity, such subsequence is not accounted for in our notations), we have the required properties for the sequence M_T^*:

$$\lim_T \frac{M_T^*}{\sqrt{T}} = 0 \qquad \text{and} \qquad \lim_T J_T M_T^* = \Gamma^*\,, \tag{A.3}$$

with $\Gamma^* = \Gamma V$ full-column rank as Γ.

2nd step: Up to considering only a subsequence, we can assume w.l.o.g. up to a minor regularity condition that $M_T = W\Lambda_T$.
From the sequence $M_T^* = W_T \Lambda_T$ defined in the first step, we use (again) the argument that the sequence (W_T) takes its values in the compact set of real orthogonal matrices to note that it exists a subsequence converging towards some orthogonal matrix W.
Then, considering the corresponding subsequence (to simplify notations, the subsequence is explicitly taken into account) $M_T^{**} = W\Lambda_T$, we want to show that it also fulfills (A.3), that is akin to say that:

$$\lim_T \frac{(Id - WW_T')M_T^*}{\sqrt{T}} = 0\,, \tag{A.4}$$

and

$$\lim_T J_T (Id - WW_T') M_T^* = 0\,. \tag{A.5}$$

Since the sequence $(Id - WW_T')$ is bounded (as the sequence of orthogonal matrices (WW_T')) we deduce immediately (A.4) from the first condition of (A.3).
For the same reason, we deduce easily from the second condition of (A.3) as well as the fact that $\lim_T W_T{}' = W^{-1}$ that:

$$\lim_T J_T M_T^*(Id - WW_T{}') = 0\,. \tag{A.6}$$

Therefore, we only need to maintain a minor regularity condition to make sure that (A.6) implies (A.5). In order to underpin such a regularity condition, it would take a thorough analysis of the asymptotic behavior of eigenspaces involving random matrices like J_T (see Dufour and Valery (2011)). This is beyond the scope of this paper. ■

Proof of Theorem 2.3 *(J-test)*:
Using (A.1) and (A.2), we get:

$$\begin{aligned} \sqrt{T}\overline{\phi}_T(\hat{\theta}_T) &= \sqrt{T}\overline{\phi}_T(\theta^0) - \frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'} M_T \left[M_T' \frac{\partial\overline{\phi}_T'(\hat{\theta}_T)}{\partial\theta} S_T^{-1} \frac{\partial\overline{\phi}_T(\tilde{\theta}_T)}{\partial\theta'} M_T \right]^{-1} \\ &\quad \times M_T' \frac{\partial\overline{\phi}_T'(\hat{\theta}_T)}{\partial\theta} S_T^{-1} \sqrt{T}\overline{\phi}_T(\theta^0) \\ \Rightarrow \quad TQ_T(\hat{\theta}_T) &= \left[\sqrt{T}\overline{\phi}_T(\theta^0)\right]' S_T'^{-1/2} \left[\mathbf{I}_K - P_X\right] S_T^{-1/2} \left[\sqrt{T}\overline{\phi}_T(\theta^0)\right] + o_P(1) \end{aligned}$$

with $S_T^{-1} = S_T'^{-1/2} S_T^{-1}$ and $P_X = X(X'X)^{-1}X'$ for $X = S_T^{-1/2} \frac{\partial\overline{\phi}_T(\hat{\theta}_T)}{\partial\theta'} M_T$. And we get the expected result. ■

Proof of Lemma 3.1:
(i) Assume that we can find a deterministic sequence a_T with $a_T/T^\nu \to \infty$ such that the sequence of matrices $\frac{\sqrt{T}}{a_T} M_{\backslash 1T}^{-1}$ does not converge to zero.
Then, there exists a vector $\delta \in \mathbb{R}^{p_2}$ such that $\sqrt{T} M_{\backslash 1T}^{-1} \frac{\delta}{a_T}$ does not converge to zero. Assume, for expositional simplicity, that the first coefficient b_T of this vectorial sequence does not converge to zero. Then, up to eliciting a well-chosen subsequence, we can claim that for some $\varepsilon > 0$, we have for all T,

$$|b_T| > \varepsilon\,.$$

However, we know that:

$$\sqrt{T} M_T^{-1} \left(\hat{\theta}_T - \theta^0 \right) = O_P(1) \, ,$$

and, in particular, with obvious notations

$$\sqrt{T} M_{\backslash 1T}^{-1} \left(\hat{\theta}_{\backslash 1T} - \theta^0_{\backslash 1} \right) = \sqrt{T} \Lambda_{\backslash 1T}^{-1} R_{\backslash 1}^{-1} \left(\hat{\theta}_{\backslash 1T} - \theta^0_{\backslash 1} \right) = \mathcal{O}_P(1) \, .$$

Note that $R_{\backslash 1}^{-1}\theta = \eta_{\backslash 1}$ where $\eta = R^{-1}\theta$ is the new vector of parameters (after rotation of the parameter space) defined just after assumption 4. Thus we have

$$\sqrt{T} \Lambda_{\backslash 1T}^{-1} \left(\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1} \right) = \mathcal{O}_P(1) \, ,$$

and, in particular, focusing on first (diagonal) coefficient $\lambda_{1,\backslash 1T}$ of $\Lambda_{\backslash 1T}$ and first coefficient $\hat{\eta}_{1,\backslash 1T}$ of $\hat{\eta}_{\backslash 1T}$, we have:

$$\frac{\sqrt{T}}{\lambda_{1,\backslash 1T}} (\hat{\eta}_{1,\backslash 1T} - \eta^0_{1,\backslash 1}) = \mathcal{O}_P(1) \, .$$

However, since b_T has been defined as the first coefficient of

$$\sqrt{T} M_{\backslash 1T}^{-1} \frac{\delta}{a_T} = \sqrt{T} \Lambda_{\backslash 1T}^{-1} R_{\backslash 1}^{-1} \frac{\delta}{a_T} \, ,$$

it can be written

$$b_T = \frac{\sqrt{T}}{\lambda_{1,\backslash 1T}} \frac{\delta_1}{a_T} \, ,$$

where δ_1 stands for the first coefficient of $R_{\backslash 1}^{-1}\delta$. Note that $\delta_1 \neq 0$ (since $|b_T| > \varepsilon$) and we deduce from a comparison of the two above formulas that

$$\frac{b_T}{\delta_1} a_T (\hat{\eta}_{1,\backslash 1T} - \eta^0_{1,\backslash 1}) = \mathcal{O}_P(1) \, .$$

Since $|b_T| > \varepsilon$ for all T (or at least a subsequence), this implies that, at least along a subsequence,

$$a_T (\hat{\eta}_{1,\backslash 1T} - \eta^0_{1,\backslash 1}) = \mathcal{O}_P(1) \, .$$

Therefore, the null hypothesis $H_0(\nu)$ must be violated since $a_T/T^{\nu} \to \infty$ and $(\hat{\eta}_{1,\backslash 1T} - \eta^0_{1,\backslash 1})$ has been built as a linear combination of $\left(\hat{\theta}_{\backslash 1T} - \theta^0_{\backslash 1} \right)$.

(ii) Under the alternative, we can find a deterministic sequence b_T with $b_T/T^\nu \to \infty$ such that for some non-zero vector $\delta \in \mathbb{R}^{p_2}$ we have (for a well-suited subsequence):

$$b_T \delta' \left(\hat{\theta}_{\backslash 1T} - \theta^0_{\backslash 1}\right) = \mathcal{O}_P(1)\,.$$

Then:

$$\delta' \left(\hat{\theta}_{\backslash 1T} - \theta^0_{\backslash 1}\right) = \gamma'(\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1}) = \mathcal{O}_P(1/b_T)\,,$$

for some non zero vector $\gamma = R'\delta$. Let us consider another deterministic sequence a_T with $a_T/T^\nu \to \infty$ but $a_T/b_T \to 0$. Then:

$$\gamma' a_T(\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1}) = o_P(1)\,.$$

Since we maintain the assumption that, at least for a convenient subsequence, $\gamma'(\hat{\eta}_{\backslash 1T} - \eta^0_{\backslash 1})$ does not go to zero at a rate faster than $\min_j \left[\lambda_{j\backslash 1T})/\sqrt{T}\right]$, we are able to conclude that at least one diagonal coefficient of $a_T \frac{\Lambda_{\backslash 1T}}{\sqrt{T}}$ goes to zero.
Therefore, since

$$\frac{\sqrt{T}}{a_T} M^{-1}_{\backslash 1T} = \frac{\sqrt{T}}{a_T} \Lambda^{-1}_{\backslash 1T} R^{-1}_{\backslash 1}\,,$$

at least one line of this matrix is such that the sum of the absolute coefficients goes to infinity. In other words, the norm $\|.\|_\infty$ of this matrix (maximum row sum norm, see Horn and Johnson (1985) p295) goes to infinity. Since, for the spectral matrix norm $\|.\|$ we are using in this paper, we have $\left\|M^{-1}_{\backslash 1T}\right\| \geq \sqrt{p_2}\left\|M^{-1}_{\backslash 1T}\right\|_\infty$ (see Horn and Johnson (1985) p314), we can conclude that, at least for a convenient subsequence,

$$\lim_T \left\| \frac{\sqrt{T}}{a_T} M^{-1}_{\backslash 1T} \right\| = \infty$$

■

Proof of Corollary 3.2:
(i) From Lemma 3.1(i), we get:

$$\lim_T \left[\frac{\sqrt{T}}{a_T} M^{-1}_{\backslash 1T}\delta\right] = 0\,. \tag{A.7}$$

Moreover, a mean-value expansion of the moment conditions gives:

$$\sqrt{T}\overline{\phi}_T(\hat{\theta}_T^{a,\delta}) = \sqrt{T}\overline{\phi}_T(\hat{\theta}_T) + \sqrt{T}\frac{\partial\overline{\phi}_T}{\partial\theta'}(\tilde{\theta}_T)(\hat{\theta}_T^{a,\delta} - \hat{\theta}_T)\,,$$

where, with the standard abuse of notation, we have defined component by component some $\tilde{\theta}_T$ between $\hat{\theta}_T^{a,\delta}$ and $\hat{\theta}_T$. Then, by definition of $\hat{\theta}_T^{a,\delta}$,

$$\begin{aligned}\sqrt{T}\overline{\phi}_T(\hat{\theta}_T^{a,\delta}) &= \sqrt{T}\overline{\phi}_T(\hat{\theta}_T) + \sqrt{T}\frac{\partial\bar{\phi}_T}{\partial\theta'_{\backslash 1}}(\tilde{\theta}_T)\frac{\delta}{a_T} \\ &= \sqrt{T}\overline{\phi}_T(\hat{\theta}_T) + \frac{\partial\overline{\phi}_T}{\partial\theta'_{\backslash 1}}(\tilde{\theta}_T)M_{\backslash 1T}\frac{\sqrt{T}}{a_T}M_{\backslash 1T}^{-1}\delta\,.\end{aligned} \tag{A.8}$$

It is worth realizing that in both cases i) and ii), we know that

$$\frac{\partial\overline{\phi}_T}{\partial\theta'_{\backslash 1}}(\tilde{\theta}_T)M_{\backslash 1T} = \mathcal{O}_P(1) \tag{A.9}$$

In case i), it is implied by definition 2.1 since $\frac{\partial\bar{\phi}_T}{\partial\theta'_{\backslash 1}}(\tilde{\theta}_T) = \frac{\partial\bar{\phi}_T}{\partial\theta'_{\backslash 1}}(\theta^0)$.
In case ii), it is implied by definition 2.4 since

$$\tilde{\theta}_{1T} = \hat{\theta}_{1T} \quad \text{and} \quad \tilde{\theta}_{\backslash 1T} - \hat{\theta}_{\backslash 1T} = \mathcal{O}(\delta/a_T) = o(T^{-1/4})\,.$$

Then, from (A.7), (A.8) and (A.9), we deduce

$$\sqrt{T}\overline{\phi}_T(\hat{\theta}_T^{a,\delta}) = \sqrt{T}\overline{\phi}_T(\hat{\theta}_T) + o_P(1)\,,$$

and the required result as an immediate consequence.
(ii) Since from Lemma 3.1, we have, under convenient regularity conditions,

$$\lim_T \left\|\frac{\sqrt{T}}{a_T}M_{\backslash 1T}^{-1}\right\| = \infty\,,$$

we have for most vectors $\delta \in \mathbb{R}^{p_2}$

$$\lim_T \left\|\frac{\sqrt{T}}{a_T}M_{\backslash 1T}^{-1}\delta\right\| = \infty\,. \tag{A.10}$$

Only vectors δ in the orthogonal space of the relevant eigenspace would not fulfill this condition. Then, using the expansion (A.8), we expect the vector $\sqrt{T}\overline{\phi}_T(\hat{\theta}_T^{a,\delta})$ to "blow-up" like the vector

$$z_T \equiv \frac{\partial \overline{\phi}_T(\tilde{\theta}_T)}{\partial \theta'_{\backslash 1}} M_{\backslash 1T} \frac{\sqrt{T}}{a_T} M_{\backslash 1T}^{-1} \delta \,.$$

z_T must blow-up since, by definition 2.1, $\left[\frac{\partial \overline{\phi}_T(\theta^0)}{\partial \theta'_{\backslash 1}} M_{\backslash 1T}\right]$ is asymptotically full-column rank. If $\left[\frac{\partial \overline{\phi}_T(\tilde{\theta}_T)}{\partial \theta'_{\backslash 1}} M_{\backslash 1T}\right]$ is different from $\left[\frac{\partial \overline{\phi}_T(\theta^0)}{\partial \theta'_{\backslash 1}} M_{\backslash 1T}\right]$ (due to some non-linearity w.r.t. $\theta_{\backslash 1}$), it would take some perverse asymptotic singularity to erase the blow-up in (A.10). Note that insofar as the vector $\sqrt{T}\overline{\phi}_T(\hat{\theta}_T^{a,\delta})$ blows up, we can be sure that $\texttt{Plim}\left[J_T^{a,\delta}(\Omega)\right] = \infty$ since, for T sufficiently large,

$$J_T^{a,\delta}(\Omega) \geq \texttt{Mineg}(\Omega_T) \left\|\sqrt{T}\overline{\phi}_T(\hat{\theta}_T^{a,\delta})\right\|^2 \geq \texttt{Mineg}(\Omega)\frac{1}{2} \left\|\sqrt{T}\overline{\phi}_T(\hat{\theta}_T^{a,\delta})\right\|^2 ,$$

with probability one asymptotically, where $\texttt{Mineg}(A)$ is the smallest eigenvalue of a matrix A and $\texttt{Mineg}(\Omega) > 0$ by positive definiteness. ∎

B Monte-Carlo study

B.1 Choice of the perturbation for the tests of identification strength

We now describe the automatic data-driven procedure that selects the perturbation vector δ highlighted in section 3. Our procedure has two steps: first, we design a grid that collects candidate points for the perturbation vector; second, we select a specific point in the grid.
- Step 1: design of the grid of candidate points for the perturbation vector.
For some user-chosen ν_1, we consider all subsamples of $\lfloor T^{\nu_1} \rfloor$ consecutive observations. For each such subsample, we calculate the associated GMM estimator. As a result, we obtain a cross-sectional distribution of the parameter vector θ. We can then extract the minimum and maximum for each component and create a grid of candidate points for the perturbation. The fineness of the grid is user-chosen: in our experiments, we used 10 points for each component of θ. Our results were not too sensitive to this choice.
Note: there are other ways to obtain a meaningful grid of candidate points for the perturbation vector that may be less computer-intensive.

- Step 2: selection of the perturbation vector.

For each perturbation vector, say δ_m, we consider all the subsamples of $\lfloor T^\nu \rfloor$ consecutive observations for some user-chosen ν. For each such subsample, say s, we calculate the associated local-to-zero version of the estimator

$$\hat{\theta}^{a,\delta_m}_{\lfloor T^\nu \rfloor, s} = \hat{\theta}_{\lfloor T^\nu \rfloor, s} + \begin{pmatrix} \mathbf{0}_{p_1} \\ \delta_m / [\log(\log(\lfloor T^\nu \rfloor) \times \lfloor T^\nu \rfloor^{1/4}] \end{pmatrix},$$

and the associated test statistic

$$J^{a,\delta_m}_{\lfloor T^\nu \rfloor, s} = \lfloor T^\nu \rfloor \overline{\phi}'_{\lfloor T^\nu \rfloor, s}(\hat{\theta}^{a,\delta_m}_{\lfloor T^\nu \rfloor, s}) S_T^{-1} \overline{\phi}_{\lfloor T^\nu \rfloor, s}(\hat{\theta}^{a,\delta_m}_{\lfloor T^\nu \rfloor, s}).$$

As a result, for each perturbation vector δ_m, we obtain a cross-sectional distribution of the test statistic J^{a,δ_m}_T. We can then extract the $(1-\alpha^*)$-quantile of ξ, for some user-chosen α^*. We select the perturbation vector associated with the $(1-\alpha^*)$-quantile the closest to the $(1-\alpha^*)$-quantile of the chi-square distribution with $(K-p_1)$ degrees of freedom. Note that $(1-\alpha^*)$ may, or may not correspond to the actual size of the designed test. Regardless of the chosen $(1-\alpha^*)$, the size of the test $(1-\alpha)$ is always controlled as we have shown above. In our experiments we used $\alpha^* = \alpha$. Our results were not too sensitive to this choice.

B.2 Diffusion process with continuous record and increasing time span asymptotic

• **Asymptotic distribution of OLS estimators for fixed Δ:**

Define X the $(n,2)$-matrix of regressors, Y and U the $(n,1)$-vector of regressand and errors respectively as follows:

$$X = \begin{bmatrix} 1 & y_0 \\ \vdots & \vdots \\ 1 & y_{(n-1)\Delta} \end{bmatrix} \quad , \quad Y = \begin{pmatrix} y_\Delta \\ \vdots \\ y_{n\Delta} \end{pmatrix} \quad \text{and} \quad U = \begin{pmatrix} u_\Delta \\ \vdots \\ u_{n\Delta} \end{pmatrix} = \sqrt{c} \begin{pmatrix} \epsilon_\Delta \\ \vdots \\ \epsilon_{n\Delta} \end{pmatrix}.$$

The OLS estimators are:

$$\begin{pmatrix} \hat{a}_{n,ols} \\ \hat{b}_{n,ols} \end{pmatrix} = (X'X)^{-1} X'Y$$

$$\hat{c}_{n,ols} = \frac{\hat{u}'\hat{u}}{n-2} \quad \text{with} \quad \hat{u} = Y - X[\hat{a}_{n,ols}\ \hat{b}_{n,ols}].$$

For fixed Δ, the usual asymptotic result for OLS estimators holds:

$$\sqrt{n}\begin{bmatrix} \hat{a}_{n,ols} - a(\theta_1) \\ \hat{b}_{n,ols} - b(\theta_1) \\ \sqrt{\hat{c}_{n,ols}} - \sqrt{c(\theta_1)} \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma(\Delta)) \qquad \text{with} \qquad \Sigma(\Delta) = \begin{pmatrix} cE[(X_i X_i'^{-1} & 0 \\ 0 & \frac{c}{2} \end{pmatrix}.$$

We use $\hat{\Sigma}$ to estimate $\Sigma(\Delta)$ where

$$\hat{\Sigma} = \begin{pmatrix} \hat{c}_{n,ols} \times n(X'X)^{-1} & 0 \\ 0 & \hat{c}_{n,ols}/2 \end{pmatrix}.$$

• **Identification strength of the three moment conditions:**

We now study the asymptotic properties of the three estimators of θ_1 and show that each estimator converges at a different rate.

(1) $\hat{\theta}_{1,b}$ denotes the estimator of θ_1 based on the moment condition 2. There is a one-to-one relationship between θ_1 and b. Its mean-value expansion gives:

$$\hat{\theta}_{1,b} - \theta_1^0 = -\frac{1}{\Delta b}(\hat{b}_{n,ols} - b) \;\Rightarrow\; \sqrt{n}(\hat{\theta}_{1,b} - \theta_1^0) = -\frac{1}{\Delta b}\sqrt{n}(\hat{b}_{n,ols} - b)$$
$$\Rightarrow\; \mathtt{Var}(\sqrt{T}(\hat{\theta}_{1,b} - \theta_1^0)) = \frac{1}{\Delta^2 b^2}\mathtt{Var}(\sqrt{n}(\hat{b}_{n,ols} - b)).$$

Recall that

$$\mathtt{Var}(\sqrt{n}(\hat{b}_{n,ols} - b)) \equiv \Sigma_b(\Delta) = \frac{\sigma^2}{\mathrm{Var}(y_t)} = 1 - b^2 = 1 - e^{-2\theta_1\Delta}.$$

In addition, when Δ is small enough, we have:

$$1 - e^{-2\theta_1\Delta} \sim 2\theta_1\Delta \;\Rightarrow\; \mathtt{Var}(\sqrt{T}(\hat{\theta}_{1,b} - \theta_1^0)) \sim \frac{2\theta_1}{\Delta b^2}.$$

As a result, $\mathtt{Var}(\sqrt{T}(\hat{\theta}_{1,b} - \theta_1^0))$ is finite when Δ is small enough, and we conclude that the rate of convergence of $\hat{\theta}_{1,b}$ is $\sqrt{T}$.

(2) $\hat{\theta}_{1,a}$ denotes the estimator of θ_1 based on moment condition 1. Similarly to the previous estimator, a mean-value expansion leads to

$$\hat{a}_{n,ols} - a \;=\; \theta_0\left[-\frac{1}{\theta_1^2}(1 - e^{-\theta_1\Delta}) + \frac{1}{\theta_1}\Delta e^{-\theta_1\Delta}\right](\hat{\theta}_{1,a} - \theta_1^0),$$

where

$$\left[-\frac{1}{\theta_1^2}(1-e^{-\theta_1\Delta})+\frac{1}{\theta_1}\Delta e^{-\theta_1\Delta}\right]\sim-\frac{\Delta^2}{2}\quad\text{when }\Delta\text{ is small enough.}$$

Note also that

$$\texttt{Var}(\sqrt{n}(\hat{a}_{n,ols}-a))=\Sigma_a(\Delta)=\sigma_2^2=\theta_2^2\left(\frac{1-e^{-2\theta_1\Delta}}{2\theta_1}\right)\sim\theta_2^2\Delta\quad\text{when }\Delta\text{ is small enough.}$$

As a result, $\texttt{Var}(\Delta\sqrt{T}(\hat{\theta}_{1,a}-\theta_1^0))$ is finite when Δ is small enough, and we conclude that the rate of convergence of $\hat{\theta}_{1,a}$ is $\Delta\sqrt{T}$ which is slower than $\sqrt{T}$.

(3) $\hat{\theta}_{1,c}$ denotes the estimator of θ_1 based on moment condition 3. Similarly to the previous estimators, a mean-value expansion leads to:

$$\hat{c}_{n,ols}-c\;=\;\frac{\theta_2^2}{2}\left[-\frac{1}{\theta_1^2}(1-e^{-2\theta_1\Delta})+\frac{2}{\theta_1}\Delta e^{-2\theta_1\Delta}\right](\hat{\theta}_{1,c}-\theta_1^0)\,,$$

where

$$\left[-\frac{1}{\theta_1^2}(1-e^{-2\theta_1\Delta})+\frac{2}{\theta_1}\Delta e^{-2\theta_1\Delta}\right]\sim-2\Delta^2\quad\text{when }\Delta\text{ is small enough.}$$

Note also that:

$$\texttt{Var}(\sqrt{n}(\hat{c}_{n,ols}-c))\Sigma_c(\Delta)=2\sigma_2^4=2\theta_2^4\left(\frac{1-e^{-2\theta_1\Delta}}{2\theta_1}\right)^2\sim2\theta_2^4\Delta^2\quad\text{when }\Delta\text{ is small enough.}$$

As a result, $\texttt{Var}(\sqrt{\Delta}\sqrt{T}(\hat{\theta}_{1,c}-\theta_1^0))$ is finite when Δ is small enough, and we conclude that the rate of convergence of $\hat{\theta}_{1,c}$ is $\sqrt{\Delta}\sqrt{T}$ which is slower than $\sqrt{T}$ but faster than $\Delta\sqrt{T}$.

(4) Identification strengths as a function of Δ:

Strong	$\Delta=1$
Near-strong	$\frac{1}{T^{1/2}}<<\Delta<<1$
Near-weak	$\frac{1}{T}<<\Delta<<\frac{1}{T^{1/2}}$
Weak	$\Delta=\frac{1}{T}$

B.3 Tables of results

	Distribution of the standardized GMM estimator			
	Mean	Variance	Skewness	Kurtosis
Intercept	0.0078	0.9874	0.0054	2.7716
Slope	0.2994	1.1657	0.7161	3.3948
	Staiger and Stock	Stock and Yogo	Antoine and Renault	
	(joint)	(joint - bias 10%)	(joint)	(slope only)
Rej. frequencies	0.4974	0.5884	0.1446	0.0516
	Other versions of Stock and Yogo			
	(bias 5%)	(bias 10%)	(size 10%)	(size 15%)
Rej. frequencies	0.1680	0.5884	0.0044	0.2346

Table 1: Testing weak identification in the linear model with heteroskedasticity. We provide the first four moments of the Monte-Carlo distribution of the standardized GMM estimator to assess the accuracy of the asymptotic approximation. We also provide rejection probabilities associated with Staiger and Stock rule-of-thumb, Stock and Yogo test based on 2SLS 10%-bias, and two versions of our test, joint and on the subvector of the slope. We also include other versions of Stock and Yogo test based on the bias and size. The parameters are $T = 250$, $M = 5,000$, $\rho = 0.8$, $R_x^2 = 0.14$, $\theta^0 = (0\ 0)'$ and we use 3 IV.

	Distribution of the standardized GMM estimator			
	Mean	Variance	Skewness	Kurtosis
Intercept	0.0046	0.9979	0.0073	2.8108
Slope	0.2987	1.1260	0.7342	3.4558
	Staiger and Stock	Stock and Yogo	Antoine and Renault	
	(joint)	(joint - bias 10%)	(joint)	(slope only)
Rej. frequencies	0.4974	0.5884	0.1366	0.0526
	Other versions of Stock and Yogo			
	(bias 5%)	(bias 10%)	(size 10%)	(size 15%)
Rej. frequencies	0.1680	0.5884	0.0044	0.2346

Table 2: Testing weak identification in the linear model with heteroskedasticity. We provide the first four moments of the Monte-Carlo distribution of the standardized GMM estimator to assess the accuracy of the asymptotic approximation. We also provide rejection probabilities associated with Staiger and Stock rule-of-thumb, Stock and Yogo test based on 2SLS 10%-bias, and two versions of our test, joint and on the subvector of the slope. We also include other versions of Stock and Yogo test based on the bias and size. The parameters are $T = 250$, $M = 5,000$, $\rho = 0.8$, $R_x^2 = 0.14$, $\theta^0 = (0\ 0)'$ and we use 3 IV.

	Distribution of the standardized GMM estimator			
	Mean	Variance	Skewness	Kurtosis
Intercept	0.0067	0.6548	0.0232	3.2929
Slope	1.0370	1.4095	0.7681	3.1806
	Staiger and Stock (joint)	Stock and Yogo (joint - bias 10%)	Antoine and Renault (joint)	(slope only)
Rej. frequencies	0.0002	0.0004	0.1998	0.0838
	Other versions of Stock and Yogo (bias 5%)	(bias 10%)	(size 10%)	(size 15%)
Rej. frequencies	0	0.0004	0	0

Table 3: Testing weak identification in the linear model with heteroskedasticity. We provide the first four moments of the Monte-Carlo distribution of the standardized GMM estimator to assess the accuracy of the asymptotic approximation. We also provide rejection probabilities associated with Staiger and Stock rule-of-thumb, Stock and Yogo test based on 2SLS 10%-bias, and two versions of our test, joint and on the subvector of the slope. We also include other versions of Stock and Yogo test based on the bias and size. The parameters are $T = 250$, $M = 5,000$, $\rho = 0.8$, $R_x^2 = 0.01$, $\theta^0 = (0\ 0)'$ and we use 3 IV.

	Distribution of the standardized GMM estimator			
	Mean	Variance	Skewness	Kurtosis
Intercept	0.0037	0.6957	0.0207	3.1171
Slope	1.0893	1.4867	0.8017	3.2778
	Staiger and Stock (joint)	Stock and Yogo (joint - bias 10%)	Antoine and Renault (joint)	(slope only)
Rej. frequencies	0.0002	0.0004	0.2198	0.0956
	Other versions of Stock and Yogo			
	(bias 5%)	(bias 10%)	(size 10%)	(size 15%)
Rej. frequencies	0	0.0004	0	0

Table 4: Testing weak identification in the linear model with homoskedasticity. We provide the first four moments of the Monte-Carlo distribution of the standardized GMM estimator to assess the accuracy of the asymptotic approximation. We also provide rejection probabilities associated with Staiger and Stock rule-of-thumb, Stock and Yogo test based on 2SLS 10%-bias, and two versions of our test, joint and on the subvector of the slope. We also include other versions of Stock and Yogo test based on the bias and size. The parameters are $T = 250$, $M = 5,000$, $\rho = 0.8$, $R_x^2 = 0.01$, $\theta^0 = (0\ 0)'$ and we use 3 IV.

Identification strength Δ	GMM estimator $\hat{\theta}_{all}$	$\hat{\theta}_{\backslash 3}$	$\hat{\theta}_{\backslash 2}$
Strong (with $\Delta = 1$)	1	1	0.025
Mildly weak (with $\Delta = 0.342$)	1	1	0.022
Medium weak (with $\Delta = 0.108$)	1	1	0.019
Very weak (with $\Delta = 0.054$)	1	1	0.018

Table 5: Testing identification strengths in the non-linear model. We provide the rejection probabilities for different identification strengths. Smaller values of Δ correspond to cases where the identification strength is weaker. Each test is characterized by the GMM estimator being considered either $\hat{\theta}_{all}$ based on the 3 moment conditions, $\hat{\theta}_{\backslash 3}$ based on conditions 1 and 2, or $\hat{\theta}_{\backslash 2}$ based on conditions 1 and 3. $T = 100$, $M = 1000$, $\theta^0 = (0.125\ 0.75\ 0.006)'$.