Econometric Theory, **0**, 2023, 1–51.
doi:10.1017/S0266466623000221

ON GMM INFERENCE: PARTIAL IDENTIFICATION, IDENTIFICATION STRENGTH, AND NONSTANDARD ASYMPTOTICS

Donald S. Poskitt
Monash University

This paper analyses aspects of generalized method of moments (GMM) inference in moment equality models in settings where standard regularity conditions may break down. Explicit analytic formulations for the asymptotic distributions of estimable functions of the GMM estimator and statistics based on the GMM criterion function are derived under relatively mild assumptions. The moment Jacobian is allowed to be rank deficient, so first order identification may fail, the values of the Jacobian singular values are not constrained, thereby allowing for varying levels of identification strength, the long-run variance of the moment conditions can be singular, and the GMM criterion function weighting matrix may also be chosen sub-optimally. The large-sample properties are derived without imposing a specific structure on the functional form of the moment conditions. Closed-form expressions for the distributions are presented that can be evaluated using standard software without recourse to bootstrap or simulation methods. The practical operation of the results is illustrated via examples involving instrumental variables estimation of a structural equation with endogenous regressors and a common CH features model.

1. INTRODUCTION

The generalized method of moments (GMM) approach to econometric modeling has several well-known advantages, not least of which is that many economic and financial models can be cast into the GMM framework. There is considerable evidence from experimental investigations using simulation designs of empirical relevance in economics and finance, however, that asymptotic theory can provide poor approximations to the sampling distributions of GMM estimators and test statistics. Examples of this discrepancy date back to the work of Tauchen (1986)

I am grateful to two anonymous referees for valuable comments which helped to improve the paper. I am indebted to the Co-Editor (Patrik Guggenberger) for insightful and constructive criticism and for helpful suggestions on the content and presentation of the paper. I am also grateful to the Editor (Peter C.B. Phillips) for correcting errors, directing my attention to related extant literature, and editorial assistance. This research has been supported by the Australian Research Council (ARC) Discovery Grant DP120102344. Address correspondence to Donald S. Poskitt, Department of Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia; e-mail: Donald.Poskitt@monash.edu.au

and Hansen, Heaton, and Yaron (1996), and Podivinsky (1999) reviews a range of early evidence showing that the sampling distributions of GMM estimators can be skewed and heavy-tailed, and test statistics can exhibit substantial size distortions. A more recent literature has sought to explain the shortcomings of standard GMM asymptotics by reference to the concept of "weakness," following ideas developed in the context of instrumental variable (IV) estimation in linear equations models. The motivation behind the arguments used in the IV literature stems from the observation that standard large-sample approximations work poorly when the IVs are weak, and the alternate paradigm chosen to explain this problem in linear IV models is to allow reduced form-regression coefficients to approach zero as the sample size n increases.[1] There is an extensive literature on weak instruments (see, inter alios, Stock, Wright, and Yogo, 2002; Hahn and Hausman, 2003; Andrews and Stock, 2007), and given that IV estimation is perhaps the leading special case of GMM, following a similar paradigm in investigations of GMM is natural.

In linear IV models, the reduced-form parameters that are the source of weak identification do not appear in the structural equations of interest, and in their analysis of weak identification in GMM, Stock and Wright (2000) similarly consider population moment functions in which the IVs do not appear and place conditions on these moment functions that imply that the IVs are weakly identifying. Guggenberger and Smith (2005) derive the asymptotic distribution of the generalized empirical likelihood (GEL) estimator and present GEL test statistics that are invariant to the strength of the IVs for models specified by nonlinear moment restrictions where identification is characterized as in Stock and Wright (2000). Han and Phillips (2006) generalize the GMM limit theory of Stock and Wright (2000) by considering moment conditions that can be functions of endogenous and exogenous variables and IVs, and by allowing the number of moment conditions to be large, while at the same time permitting the moment conditions to be weak. Newey and Windmeijer (2009) further extend the work of Stock and Wright (2000), Chao and Swanson (2005), and Han and Phillips (2006) by allowing for differing convergence rates in their many weak moment asymptotics. Antoine and Renault (2009) also examine GMM estimation with instruments of varying strength, with an emphasis on asymptotic efficiency, and Caner (2010) presents a similar generalization of Stock and Wright (2000) by allowing for instruments of varying strength. Dovonon and Renault (2013) generalize Stock and Wright (2000) by providing a general asymptotic theory for GMM when the rank condition of the Jacobian matrix fails, but the moment conditions identify the true parameter value via second-order identification (identification through the second derivative of the moment conditions).

Many of the features observed in investigations of weak moment GMM asymptotics parallel those seen in the analysis of weak IVs, and these in turn parallel those obtained with partially identified linear models. (See Poskitt and Skeels, 2013,

[1] See Phillips (2006) for an interesting historical perspective on the development of weak instruments dating back as far as the 1970s.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

for a brief historical account of the latter connection.) To enquire as to what features emerge from a consideration of partially identified models with mixed levels of moment condition identification strength, and to ask if these explain the shortcomings previously observed in applications of GMM, therefore seem reasonable questions to raise. The present paper is designed to address these questions by investigating the properties of GMM when the model is unidentified or, more correctly, partially identified, and by examining the extent to which previously observed features are reflected in GMM asymptotics with different levels of moment identification strength.

Kleibergen (2005) has considered aspects of GMM inference without the necessity to assume that the parameters are identified. Kleibergen's focus is on hypothesis testing, and by (i) replacing standard GMM necessary conditions for identification with an assumption that the moment conditions and their Jacobian are asymptotically jointly normal, and (ii) structuring the test statistics as quadratic forms in the corresponding conditional Gaussian random variables, he is able to derive conventional chi-squared asymptotic distributions and less conventional chi-squared mixtures for his modified GMM test statistics. Andrews and Guggenberger (2017) examine nonlinear conditional likelihood ratio (CLR) tests that depend on a rank statistic that measures the rank of the expected Jacobian. Andrews and Guggenberger (2017) follow the approach of Kleibergen using a weighted orthogonalized version of the sample Jacobian. Lee and Liao (2018) use the approach of Kleibergen to revisit the analysis of Dovonon and Renault (2013), and they tackle the issue of identification failure by examining the behavior of Hansen's $\mathfrak{J}$-test (Hansen, 1982) based on the moment conditions and a zero Jacobian of known form. Andrews and Guggenberger (2019) construct quasi CLR and Anderson–Rubin test statistics that are robust to singularities using estimated stochastic linear combinations of the moments that depend on the rank and singular values of the Jacobian. In all these papers, the authors suggest that the critical values of the nonstandard distributions derived for their test statistics be obtained using bootstrap procedures or simulation methods.

In a contribution that steps outside the confines of the previous literature, Phillips (2016) examines a form of identification failure due to regressor asymptotic degeneracy. He develops a limit theory for regression estimates and test statistics that shows that near-collinearity in stationary and trend stationary regressors can produce serious failures for conventional inference. These include inconsistency, the absence of an invariance principle, and mixed asymptotic normality for appropriately standardised estimates. In spite of coefficient inconsistency and a breakdown of standard central limit theory, near-singular regression designs of the type considered in Phillips are not disastrous. They lead to random functional characterizations of the limit theory in terms of series representations that result in size distortions in linear regression models, and for IV regression estimates in structural equation models, in test procedures, such as the overidentification test of Sargan (1958), being conservative if implemented using standard chi-squared critical values.

A different line of investigation to that followed in the aforementioned references considers moment inequality models. The parameters in such models are partially identified, and the technical apparatus developed by Chernozhukov, Hong, and Tamer (2007) and Menzel (2014) to investigate the estimation of identified sets in moment inequality models can be exploited in moment equality models for (as observed by the Co-Editor) a zero expected moment is equivalent to the expectation of the moment and the expectation of its negative both being nonnegative.[2] Likewise, a moment inequality can be transformed into a moment equality by the introduction of a nonnegative slack parameter. It seems reasonable to conjecture that the techniques developed in this paper can be used to allow for the violation of the necessary condition for identification that the introduction of such slack parameters would entail, and that the nonstandard distribution theory for estimable functions and test statistics developed in this paper can be adapted to apply to moment inequality models. The details of such an analysis are beyond the scope of the current paper, however, and will be pursued elsewhere.

The current paper examines similar issues to those considered in the extant literature, but directs its attention to an analysis of the properties of standard GMM estimators and GMM test statistics and diagnostic devices in moment equality models. This paper therefore follows different avenues of investigation from those taken in the preceding literature. Here, we impose no identification assumptions of any type other than that the practitioner has employed a sufficient number of linearly independent moments to meet the necessary condition for identification, and that the expected value of the moment conditions equals zero at the true parameter point. We do not impose any conditions on the functional form of the moment conditions, and we allow the long-run variance of the moment conditions to be singular. Singularity of the moment condition long-run variance implies that the conventional choice for the GMM criterion function efficient weighting matrix is not available, but the weighting matrix can be chosen sub-optimally without affecting the results presented here other than a loss in efficiency. The moment Jacobian is assumed to have constant rank in a neighborhood of the true parameter vector but is allowed to be rank deficient, so first-order lack of identification may hold. No constraints are imposed on the values of the nonzero Jacobian singular values, thereby allowing for varying levels of moment condition identification strength. Under relatively mild assumptions on the large-sample convergence properties of the moment conditions and the Jacobian, explicit analytic formulations for the nonstandard asymptotic distributions of estimable functions of the GMM estimator, and GMM test statistics, are derived. A significant contribution of this paper therefore lies in the generality which the asymptotic results and associated inferential procedures that are derived are applicable.

[2]Chernozhukov et al. (2007) and Menzel (2014) provide results on the consistency and coverage probability of set estimators defined as lower contour sets of GMM-type criterion functions. These ideas are exploited below to establish set convergence in the current setting.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

To set the scene, consider a situation where we have data $\mathbf{w}_t$, $t = 1,\ldots,n$, on a v element random variable where the data generating process (DGP) comes from a class of statistical models $\mathcal{M}$ characterized by a sample space $(\mathbb{R}^{vn}, \mathfrak{B}^{vn})$ with an associated sequence of probability measures $P_{\theta,n}(\cdot)$ that are dependent on a parameter $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_p)' \in \Theta \subset \mathbb{R}^p$. We suppose that the DGP gives rise to a set of moment conditions $E[\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta})]$ where $\boldsymbol{\mu} : \mathbb{R}^v \times \mathbb{R}^p \longmapsto \mathbb{R}^k$ and $\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta})$ is a $k \times 1$ vector-valued measurable function of the process $\mathbf{w}_t$. For each DGP in $\mathcal{M}$, we assume that there exists a parameter value $\boldsymbol{\theta}_0 \in \Theta$ for which the moment conditions $E[\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta}_0)] = \mathbf{0}$ hold. Since in this setting the moment conditions are the only vehicle available with which to characterise features of the model and investigate the data, any modeling assumptions and estimation and inference techniques must be defined and conducted via these conditions.

In the following section, Section 2, the GMM criterion function $Q_n(\boldsymbol{\theta})$ and GMM estimator $\widehat{\boldsymbol{\theta}}_n$ are defined, and the core assumptions adopted throughout the paper are presented. A brief discussion of some GMM basics and other developments in the extant literature that form a background to the discussion presented in this paper is also presented. The paper then proceeds as follows.

In Section 3, the first results presented in this paper on partially identified GMM models are established. The treatment of partially identified GMM models introduces substantial complications into the analysis of the properties of $Q_n(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\theta}}_n$. The concept of a quasi-true parameter set Θ_{0n} is introduced in Section 3.1, and it is shown that Θ_{0n} generates a GMM $Q_n(\boldsymbol{\theta})$ observational equivalence class and that $\widehat{\boldsymbol{\theta}}_n$ converges to a quasi-true parameter value $\boldsymbol{\theta}_{0n} \in \Theta_{0n}$ in the sense that the GMM solution set $\widehat{\Theta}_n = \{\boldsymbol{\theta} \in \Theta : Q_n(\boldsymbol{\theta}) = \min_{\theta\in\Theta} Q_n(\boldsymbol{\theta})\}$ and Θ_{0n} are Hausdorff–Kuratowski convergent as $n \to \infty$. Section 3.2 shows that although $\widehat{\boldsymbol{\theta}}_n$ does not converge in a conventional manner, under a scenario involving a drifting sequence of quasi-true parameters akin to a Pitman sequence, a class of estimable functions can be defined that have limiting normal distributions. The second contribution of this paper is thus to show that a consequence of working with partially identified models is that certain functions of the parameters are estimable, and the GMM estimators of these estimable functions will be consistent and have meaningful limit distributions. It is demonstrated how hypotheses tests and confidence regions can be constructed using the asymptotic normality of estimable functions as a basis for the inference. Categories of moment condition identification strength are presented in Section 3.2, and the effects of identification strength on the limit distribution of estimable functions are explored. A further contribution of this paper is thus to show that in the presence of weak identification, estimable functions will converge to nonstandard convolutions of singular inverted Wishart covariance matrix Gaussian mixtures.

Section 4 illustrates the ideas and concepts discussed in Sections 2 and 3 in the context of two examples. Section 4.1 considers IV estimation of a simultaneous equation system comprising a single structural equation with reduced-form equations for the endogenous regressors and provides experimental illustrations of the results on estimable functions presented in Section 3.2. Section 4.2

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

demonstrates how the ideas and concepts introduced in the paper relate to the common conditional heteroskedastic (CH) features model of Engle and Kozicki (1993). For this particular nonlinear model, it is known that nullity of the moment Jacobian at any common feature means that it is not possible to build a $\sqrt{n}$ consistent estimate of $\boldsymbol{\theta}_0$ (Dovonon and Renault, 2013; Lee and Liao, 2018). Consequently, the focus of interest turns to the nonstandard asymptotic properties of $Q_n(\boldsymbol{\theta})$ induced by the nonstandard asymptotic behavior of $\widehat{\boldsymbol{\theta}}_n$.

A fourth contribution of this paper is therefore to derive the limiting distribution of Hansen's $\mathfrak{J}$-test statistic of overidentification (Hansen, 1982), and associated statistics based on $Q_n(\boldsymbol{\theta})$, when the standard theory does not apply due to, for example, singularity in the moment long-run variance matrix or rank deficiency in the moment Jacobian. The new limiting distributions can be expressed as uniformly convergent series expansions in generalized Laguerre polynomials and are presented in Section 5. A key feature of these distributions is that closed-form expressions for their cumulative distribution functions are available. The distributions can therefore be implemented numerically using standard software without recourse to bootstrap or simulation methods.

Section 6 extends the illustrations presented in Section 4 and investigates the properties of Hansen's $\mathfrak{J}$-test statistic and associated statistics based on $Q_n(\boldsymbol{\theta})$ in the context of both the IV simultaneous equations model, Section 6.1, and the common CH features model, Section 6.2. The paper concludes in Section 7 with a brief summary and comment on the implications of the results for the practitioner. The Supplementary Material of this article provides further technical material, including regularity conditions, limit theory, and an illustration of algebraic and numerical consequences of partial identification.

2. ASSUMPTIONS AND BACKGROUND

Suppose that the data $\mathbf{w}_t$, $t = 1,\ldots,n$, are generated by a DGP from $\mathcal{M}$ characterized by the parameter value $\boldsymbol{\theta}_0 \in \Theta$. Then $\boldsymbol{\theta}_0$ is globally identified by the moment conditions if

$$\frac{1}{n}\sum_{t=1}^{n} E[\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta})] = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \mathbf{0} \iff \boldsymbol{\theta} = \boldsymbol{\theta}_0 . \tag{2.1}$$

A direct consequence of the condition in (2.1) and the implicit function theorem is that a necessary condition for the global identification of $\boldsymbol{\theta}_0$ is that $k \geq p$, namely, the model is "just-identified" or "over-identified." Since the researcher is at liberty to specify any set of moment conditions for which they believe $\boldsymbol{\theta}_0$ is identified, and checking if $k \geq p$ simply involves counting the number of moments and the number of parameters, we will assume throughout that $k \geq p$, although the case where $k < p$ can be handled with relatively minor adjustments. We will not impose condition (2.1) in the remainder of the paper and identification of the true parameter $\boldsymbol{\theta}_0$ is not assumed.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

For a "just-identified" or "over-identified" model, the GMM estimator is obtained by minimizing the quadratic form

$$Q_n(\boldsymbol{\theta}) = \mathbf{m}_n(\mathbf{w},\boldsymbol{\theta})'\mathbf{W}_n\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta}), \quad \textbf{(2.2)}$$

where the sample moment is

$$\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta}),$$

and $\mathbf{W}_n$ is a $k \times k$ symmetric and bounded weighting matrix, $\mathbf{W}_n = \mathbf{W}_n'$ and $0 < \|\mathbf{W}_n\| < \infty$ with probability one where $\|\mathbf{W}_n\|^2 = \mathrm{tr}(\mathbf{W}_n\mathbf{W}_n')$. By definition,

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\theta\in\Theta} Q_n(\boldsymbol{\theta}), \quad \textbf{(2.3)}$$

and supposing that the moment conditions are continuously differentiable in $\boldsymbol{\theta}$, the first order condition for a minimum of the criterion $Q_n(\boldsymbol{\theta})$ at $\widehat{\boldsymbol{\theta}}_n$ is

$$\partial Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n} = 2\,\mathbf{D}_n(\mathbf{w},\boldsymbol{\theta})'\mathbf{W}_n\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n} = \mathbf{0}, \quad \textbf{(2.4)}$$

where

$$\mathbf{D}_n(\mathbf{w},\boldsymbol{\theta}) = \frac{\partial\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$$

is the $k \times p$ sample Jacobian matrix.

Assumption 2.1. The parameter space Θ is a compact subset of $\mathcal{R}^p$, the moment conditions $\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta})$ are continuously differentiable in $\boldsymbol{\theta}$, and there exists a $\delta > 0$ such that the Jacobian matrix

$$\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial E[\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}$$

has constant rank $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} = q_n$ for all $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta) = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta}_0 - \boldsymbol{\theta}\| < \delta\}$, the open neighborhood in Θ with center $\boldsymbol{\theta}_0$ and radius $\delta > 0$.

From a practical viewpoint, the presumption that the parameter space is closed and bounded is unfortunate—it can be relaxed in some situations by using appropriate convexity conditions (Hansen, 1982)—but it is maintained here for ease of exposition and conformity with much of the existing literature. Continuous differentiability of the moment conditions $\boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\theta})$ is readily checked. Compactness and continuity ensure that there exists a $\widehat{\boldsymbol{\theta}}_n$ that satisfies (2.3) and solves (2.4), but uniqueness is not guaranteed. The solution set $\widehat{\Theta}_n = \{\boldsymbol{\theta} \in \Theta : Q_n(\boldsymbol{\theta}) = \min_{\theta\in\Theta} Q_n(\boldsymbol{\theta})\}$ will be either a singleton or a set of points, an issue that we will return to in Section 3. When $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} = q_n \leq p$ for all $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta)$, the Jacobian matrix is said to be regular. If the Jacobian matrix $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ is regular and rank deficient, then the classical condition for (local) identification that the Jacobian has full-column rank p will be violated. We present this result formally in the following lemma.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

LEMMA 2.1. *Suppose that Assumption 2.1 holds. Then* $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} = q_n < p$ *if and only if for every neighborhood* $N(\boldsymbol{\theta}_0;\delta')$, $\delta' < \delta$, *there exists a point* $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta')$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, *such that* $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)$. *Moreover, when* $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} = q_n < p$, *then* $\{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta) : \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)\}$ *is a compact, connected subset of* Θ *of dimension* $p - q_n$.

The proof of Lemma 2.1, which is modeled on the proof of Rothenberg (1971, Thm. 1, p. 579), is given in the Supplementary Material. Lemma 2.1 implies that $\boldsymbol{\theta}_0$ is (locally) identified if and only if $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ has full-column rank, but checking if $\boldsymbol{\theta}_0$ is identified can be difficult if the structure of $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ is complicated. Identification is often therefore simply assumed, although a number of studies have shown that imposing identification with some models may be too strong as an assumption, and that conventional inference procedures break down when it fails (see Stock et al., 2002, for example, and for further discussion and examples of identification and the lack thereof (Hall, 2005, Chap. 3.1)).

For any fixed k and p with $k \geq p$, the singular value decomposition of $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)$ is given by

$$\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0) = \bar{\mathbf{U}}_{0n}\bar{\mathbf{S}}_{0n}\bar{\mathbf{V}}_{0n}, \tag{2.5}$$

where the p columns of $\bar{\mathbf{U}}_{0n}$, $(k \times p)$, are an orthonormal p-frame of $\mathcal{O}(k)$, the Stiefel manifold in $\mathbb{R}^k$, so $\bar{\mathbf{U}}_{0n}'\bar{\mathbf{U}}_{0n} = \mathbf{I}_p$, $\bar{\mathbf{S}}_{0n} = \mathrm{diag}\{\bar{s}_{0n,1},\ldots,\bar{s}_{0n,p}\}$, and $\bar{\mathbf{V}}_{0n}$, $(p \times p)$, belongs to the orthogonal group $\mathcal{O}(p)$. Without loss of generality, the singular values will be assumed to be listed in nonincreasing order, $\bar{s}_{0n,1} \geq \bar{s}_{0n,2} \geq \cdots \geq \bar{s}_{0n,q_n} > \bar{s}_{0n,q_n+1} = \cdots = \bar{s}_{0n,p} = 0$, where $q_n = r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\}$, and $0 < \bar{s}_{0n,q_n} \leq \cdots \leq \bar{s}_{0n,1} < \infty$. The value of q_n will be called the identification rank, with $1 \leq q_n \leq p$. When all the singular values are zero and $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)$ is null, we will set $q_n = 0$. The strength of the identification will depend on the magnitude of the singular values, and borrowing from the categories and nomenclature of Andrews and Guggenberger (2017, 2019):

- (i) The identification strength of the moment conditions will be said to be strong or semi-strong when $lim_{n\to\infty} n^{\frac{1}{2}}\bar{s}_{0n,p} = \infty$. Strong identification occurs when $lim_{n\to\infty} n^{\frac{1}{2}}\bar{s}_{0n,p} = \infty$ and $lim_{n\to\infty}\bar{s}_{0n,p} > 0$, and semi-strong identification occurs when $lim_{n\to\infty} n^{\frac{1}{2}}\bar{s}_{0n,p} = \infty$ and $lim_{n\to\infty}\bar{s}_{0n,p} = 0$.
- (ii) When $lim_{n\to\infty} n^{\frac{1}{2}}\bar{s}_{0n,1} < \infty$, the identification strength of the moment conditions will be said to be weak.
- (iii) When $lim_{n\to\infty} n^{\frac{1}{2}}\bar{s}_{0n,1} = \infty$ and $lim_{n\to\infty} n^{\frac{1}{2}}\bar{s}_{0n,p} < \infty$, we will refer to the moment condition identification strength as being mixed.

From the singular value decomposition in equation (2.5), the cth column of $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)$ is given by

$$\bar{\boldsymbol{\delta}}_{0n,c} = \bar{\mathbf{U}}_{0n}\bar{\mathbf{S}}_{0n}\bar{\mathbf{v}}_{0n,c}, \quad c = 1,\ldots,p,$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

where $\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta}_0) = [\bar{\boldsymbol{\delta}}_{0n,1}, \ldots, \bar{\boldsymbol{\delta}}_{0n,p}]$ and $\bar{\mathbf{V}}_{0n} = [\bar{\mathbf{v}}_{0n,1}, \ldots, \bar{\mathbf{v}}_{0n,p}]$. Since each column of the Jacobian characterizes the behavior of the corresponding parameter estimate, and $\|\bar{\boldsymbol{\delta}}_{0n,c}\| = (\sum_{r=1}^{p} \bar{s}_{0n,r}^2 \bar{v}_{0n,rc}^2)^{\frac{1}{2}}$, it follows that $\widehat{\boldsymbol{\theta}}_n$ will inherit its identification strength from that of the moment conditions, and that it can do so in various different ways.

When the identification strength of the moment conditions is strong or semi-strong, the singular values share a common divergent lower bound, and the identification strength of $\widehat{\boldsymbol{\theta}}_n$ will also be strong or semi-strong since $\|\bar{\boldsymbol{\delta}}_{0n,c}\| \geq \bar{s}_{0n,p}(\sum_{r=1}^{p} \bar{v}_{0n,rc}^2)^{\frac{1}{2}}$. On the other hand, when the moment conditions are weakly identifying, they share a common limiting upper bound, and the identification strength of $\widehat{\boldsymbol{\theta}}_n$ will also be weak (as in Staiger and Stock, 1997, for example) because $\|\bar{\boldsymbol{\delta}}_{0n,c}\| \leq \bar{s}_{0n,1}(\sum_{r=1}^{p} \bar{v}_{0n,rc}^2)^{\frac{1}{2}}$. Mixed moment identification strength is referred to in Andrews and Guggenberger (2017) as "weak identification in a nonstandard sense." When the identification rank $q_n < p$, for example, $\bar{s}_{0n,q_n+1} = \cdots = \bar{s}_{0n,p} = 0$, and if $lim_{n\to\infty} n^{\frac{1}{2}} \bar{s}_{0n,q_n} = \infty$, then the identification strength will be strong or semi-strong since $\|\bar{\boldsymbol{\delta}}_{0n,c}\| = (\sum_{r=1}^{q_n} \bar{s}_{0n,r}^2 \bar{v}_{0n,rc}^2)^{\frac{1}{2}}$. In such a case, the identification is weak in a nonstandard sense because it is a mixture of identification failure with strong or semi-strong partial identification. More generally, whenever the singular values have conflicting asymptotic properties, the identification strength will be mixed, with individual components exhibiting different levels of identification strength ranging from weak through to semi-strong and strong. These cases include the combination of weak and strong parameter identification considered in Stock and Wright (2000). Further discussion of the impact of these different identification strengths is given in what follows, where it is shown that GMM statistics have nonstandard asymptotic distributions due to identification failure when $\boldsymbol{\theta}_0$ is only partially identified or due to identification deficiency when the moment conditions are weakly identifying.

Assumption 2.2. The empirical moments $\mathbf{m}_n(\mathbf{w}, \boldsymbol{\theta})$ converge in probability to their ensemble counterparts $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}$, more precisely,

$$\sup_{\theta \in \Theta} \|\mathbf{m}_n(\mathbf{w}, \boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\| \xrightarrow{p} 0.$$

Furthermore, the moment Jacobian matrix $\mathbf{D}_n(\mathbf{w}, \boldsymbol{\theta})$ converges uniformly to its theoretical counterpart $\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta})$, that is,

$$\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \left(\boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta}) - E[\boldsymbol{\mu}(\mathbf{w}_t, \boldsymbol{\theta})] \right)}{\partial \boldsymbol{\theta}'} \right\| \xrightarrow{p} 0.$$

Assumption 2.2 is a so-called "high-level' assumption that accommodates various types of data. Such assumptions are commonly adopted in GMM since they have the advantage of making the results applicable to a wide range of special cases. Their disadvantage is that their plausibility must be checked on a case-by-case basis. Primitive regularity conditions that in turn imply Assumption 2.2 can of

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

course be specified, an example of such conditions is given in the Supplementary Material.

Assumption 2.3. The weighting matrix $\mathbf{W}_n$, which can be a function of the data, is positive semi-definite with probability one and $\|\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n\| \xrightarrow{p} 0$, where $0 < \|\bar{\boldsymbol{\Omega}}_n\| < \infty$ and $\bar{\boldsymbol{\Omega}}_n = \bar{\boldsymbol{\Omega}}_n'$ is such that

$$r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\} = \{r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\} = q_n \leq p \leq r\{\bar{\boldsymbol{\Omega}}_n\} \leq k.$$

Assumption 2.3 implies that lack of identification is not due to $\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta})$ incorporating an insufficient number of linearly independent moments, but otherwise imposes no further identification conditions. Assumption 2.3 is nonstandard in that it is more conventional to assume that $\mathbf{W}_n$ and $\bar{\boldsymbol{\Omega}}_n$ are nonsingular, in which case it is well known that when the model is identified choosing $\mathbf{W}_n$ such that $\bar{\boldsymbol{\Omega}}_n$ equals the inverse of the long-run variance of the moment conditions is optimal (see Hall, 2005, Thm. 3.4, for example[3]). If the long-run variance, $\bar{\boldsymbol{\Sigma}}_{0n}$ say, is singular, the conventional choice for $\mathbf{W}_n$ as the regular inverse of a consistent estimate of $\bar{\boldsymbol{\Sigma}}_{0n}$ is unavailable. It is obvious that $\mathbf{W}_n$ still determines the relative weight given to the components of $\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta})$, however, and the intuition that $\bar{\boldsymbol{\Omega}}_n$ should be inversely proportional to the long-run variance still stands. It is shown below that when $\bar{\boldsymbol{\Sigma}}_{0n}$ is singular, an efficient (minimum variance) GMM estimator is obtained when $\mathbf{W}_n$ is chosen such that $\|(\bar{\boldsymbol{\Omega}}_n - \bar{\boldsymbol{\Sigma}}_{0n}^+)\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\| \to 0$ where $\bar{\boldsymbol{\Sigma}}_{0n}^+$ denotes the Moore–Penrose generalized-inverse (g-inverse). It is well known that discrepancies between the GMM weighting matrix limiting value and the optimal choice are a possible source of poor performance for standard GMM asymptotics (see, for example, Pagan and Robertson, 1997; Windmeijer, 2005, for discussions of such issues and further references). As well as accommodating singularity in the moment long-run variance—and rank deficiency in the Jacobian—the alternative asymptotic theory developed in this paper for $Q_n(\boldsymbol{\theta})$ allows for the possibility that $\bar{\boldsymbol{\Omega}}_n$ may not equal $\bar{\boldsymbol{\Sigma}}_{0n}^+$, and whatever is the nature of the relationship of $\bar{\boldsymbol{\Omega}}_n$ to $\bar{\boldsymbol{\Sigma}}_{0n}^+$, the new limiting distributions will adapt so as to yield the correct asymptotic distribution.

Newey (1985) assumes that $\bar{\boldsymbol{\Omega}}_n$ is positive semi-definite, but requires that $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)$ has full rank, $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\} = p$, which implies that $r\{\bar{\boldsymbol{\Omega}}_n\} \geq p$ and that the model is identified since

$$r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\} \leq \min\{r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\}, r\{\bar{\boldsymbol{\Omega}}_n\}\}$$

and $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\} \leq p$. Kleibergen (2005) allows for the possibility that the long-run variance matrix in his assumed joint asymptotic normal distribution of (in the notation of this paper) $\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta})$ and $\mathbf{D}_n(\mathbf{w},\boldsymbol{\theta})$ is singular, but the singularity can

[3]Hall assumes that $\mathbf{W}_n$ is positive semi-definite but that $\bar{\boldsymbol{\Omega}}_n$ is non-singular (Hall, 2005, Assum. 3.7). But if $\mathbf{W}_n$ converges to $\bar{\boldsymbol{\Omega}}_n$, then it can be shown that the rank of $\mathbf{W}_n$ is bounded below by that of $\bar{\boldsymbol{\Omega}}_n$ for all n sufficiently large (see Puri, Russell, and Mathew, 1984, Lem. 3).

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

only come from relationships between the moments and the Jacobian because he presupposes that the moment long-run variance is positive definite. In their extension of Kleibergen's approach, Andrews and Guggenberger (2017) assume that $\bar{\mathbf{\Omega}}_n$ is positive definite and Lee and Liao (2018) suppose that $\mathbf{W}_n$ equals the regular inverse of a consistent estimate of a positive definite moment long-run variance. In their analysis of moment inequality models, Chernozhukov et al. (2007) and Menzel (2014) suppose that the weighting matrix in their GMM-type criterion functions is positive definite. As pointed out in Andrews and Guggenberger (2019, Sect. 4), the condition that $\mathbf{W}_n$ and $\bar{\mathbf{\Omega}}_n$ be nonsingular is not inconsequential because in a number of models, lack of identification and singularity of the moment long-run variance run hand in hand.[4] Andrews and Guggenberger (2019) therefore allow the long-run variance matrix of the moment conditions to be singular, and in their analysis of identification- and singularity-robust test statistics, they place no restrictions on the rank of $\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta}_0)$. As such, the work of Andrews and Guggenberger (2019) is (to the best of this author's knowledge) the only direct precursor to that of the current paper.

Let $\theta_{i(1)},\ldots,\theta_{i(p)}$ denote a permutation of $\theta_1,\ldots,\theta_p$. The following result arises as a consequence of the function inversion theorem and the implicit function theorem. See Marsden (1974, Thm. 5, pp. 236–237) for an equivalent statement together with a proof.

LEMMA 2.2. *Suppose that Assumption 2.1 holds, that $\boldsymbol{\theta}_0$ lies in the interior of Θ, and that $r\{\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta})\} = q_n < p$, for all $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta)$. Then there exist an open set $U \subset \Theta$ and an open set $V \subset \Theta$ with $\boldsymbol{\theta}_0 \in V$, and a continuously differentiable function $\mathbf{g} : U \to V$ with a continuously differentiable inverse $\mathbf{g}^{-1} : V \to U$, such that the composition $\bar{\boldsymbol{\mu}} \circ \mathbf{g}(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\mathbf{g}(\boldsymbol{\theta})) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta}))$, where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_{q_n})' = (\theta_{i(1)},\ldots,\theta_{i(q_n)})'$ and $(\theta_{i(q_n+1)},\ldots,\theta_{i(p)})' = \boldsymbol{\alpha}(\boldsymbol{\beta}) = (\alpha_1(\boldsymbol{\beta}),\ldots,\alpha_{p-q_n}(\boldsymbol{\beta}))'$.*

Lemma 2.2 implies that there exist a permutation matrix $\mathbf{P}$ and a parameter vector $\mathbf{P}\boldsymbol{\theta} = (\theta_{i(1)},\ldots,\theta_{i(p)})'$ such that the subvector $\boldsymbol{\beta}_0$ where $\mathbf{P}\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0',\boldsymbol{\alpha}(\boldsymbol{\beta}_0)')$ is identified from the compound moment condition $\boldsymbol{\mu}(\mathbf{w}_t,\mathbf{g}^{-1}(\boldsymbol{\theta})) = \boldsymbol{\mu}(\mathbf{w}_t,\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta}))$. To see this, note that by an application of the chain rule,

$$\frac{\partial \bar{\boldsymbol{\mu}}_n(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}'} = \bar{\mathbf{\Delta}}_n(\boldsymbol{\theta})\mathbf{P}'\begin{bmatrix}\mathbf{I}\\ \frac{\partial \boldsymbol{\alpha}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\end{bmatrix},$$

and since the first q_n columns of $\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta})\mathbf{P}'$ have rank q_n, the Jacobian matrix has full-column rank q_n for all $(\boldsymbol{\beta}',\boldsymbol{\alpha}(\boldsymbol{\beta})')' = \mathbf{g}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta)$. Thus, Lemma 2.2 indicates that if $\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta})$ is rank deficient, the model is still partially identified in the sense that certain components of $\boldsymbol{\theta}$ can be identified, the number of parameters that can be identified being determined by the rank of $\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta})$. The partially identified

[4] Singularity of the long-run variance occurs in likelihood -based models in which submodels of interest are nested within encompassing models containing nuisance parameters that do not appear in the submodel, for example, and Andrews and Guggenberger (2019) list seven examples of models where such singularities occur.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

parameter $\boldsymbol{\beta}$ will satisfy Assumption 2.1, and a GMM estimator $\widehat{\boldsymbol{\beta}}_n$ obtained by replacing $\mathbf{m}_n(\mathbf{w},\boldsymbol{\theta})$ in (2.2) by $\mathbf{m}_n(\mathbf{w},\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta}))$ would, under the previously stated conditions, converge to $\boldsymbol{\beta}_0$. In general, however, the practitioner will not be able to construct $\widehat{\boldsymbol{\beta}}_n$ because the function $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{P}\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}(\boldsymbol{\beta})')$ will be unknown and the estimator will be unfeasible.

Analogous properties are observed in Stock and Wright (2000), wherein it is assumed that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_{\beta,n}(\boldsymbol{\beta}) + n^{-\frac{1}{2}}\bar{\boldsymbol{\mu}}_{\alpha,n}(\boldsymbol{\alpha})$ where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ and $\partial\bar{\boldsymbol{\mu}}_{\beta,n}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$ has full-column rank, implying that $\boldsymbol{\beta}$ is well identified but $\boldsymbol{\alpha}$ is only weakly identified or unidentified (Stock and Wright, 2000, Assum. C, Sect. 2.3, pp. 1060–1062). The imposition of such a specialised additive structure entails assigning the weakly and strongly identified directions in the parameter space, and as is shown in Andrews and Guggenberger (2017), it is unlikely to be a straightforward task to either verify or refute the existence of such structure in a broad range of familiar econometric models.[5] In this paper, we will ascertain the consequences of employing the GMM estimator calculated from $Q_n(\boldsymbol{\theta})$ while allowing the data to be generated from a DGP that may be partially identified and that can contain different levels of moment identification strength. We will do this without imposing any special structure on the expected moment conditions or their derivatives.

3. IDENTIFICATION AND ESTIMATION

3.1. Criterion and Estimator Convergence

The link between lack of identification and the estimation of $\boldsymbol{\theta}$ via $Q_n(\boldsymbol{\theta})$ follows from the following result.

THEOREM 3.1. *Suppose that Assumptions 2.1–2.3 hold and that* $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} < p$, *for all* $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta) = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta}_0 - \boldsymbol{\theta}\| < \delta\}$, $\delta > 0$. *Then there exists a* $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0;\delta)$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, *such that, for every* $\eta > 0$, *there exists an* $n_\eta < \infty$ *such that*

$$Pr[|Q_n(\boldsymbol{\theta}_0) - Q_n(\boldsymbol{\theta})| < \eta] > 1 - \eta,$$

for all $n > n_\eta \geq 1$.

A basic requirement for $\widehat{\boldsymbol{\theta}}_n$ to be a consistent estimator of $\boldsymbol{\theta}_0$ is that $Q_n(\boldsymbol{\theta})$ can distinguish $\boldsymbol{\theta}_0$ from alternative values. Theorem 3.1 indicates that when n is sufficiently large and $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ has reduced rank, $Q_n(\boldsymbol{\theta})$ will change by an arbitrarily small amount between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ with an arbitrarily large probability, and for almost all samples, it will not be possible for $Q_n(\boldsymbol{\theta})$ to distinguish between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$. This amounts to a form of GMM $Q_n(\boldsymbol{\theta})$ observational equivalence.

A consequence of Theorem 3.1 is that $\widehat{\boldsymbol{\theta}}_n$ will not converge in a conventional manner. To analyze this further, recall that the DGP is characterized by a parameter

[5]For a discussion of the role that Assumption C of Stock and Wright (2000) has played in the literature, see Andrews and Guggenberger (2017, Sect. 2, pp. 1052–1055)

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

value $\boldsymbol{\theta}_0 \in \Theta$ such that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0) = \mathbf{0}$. If $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ is regular with reduced rank $r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} < p$, then from Lemma 2.1, we know that $\boldsymbol{\theta}_0$ is (locally) unidentified and there exists a compact, connected subset of Θ such that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0) = \mathbf{0}$. We can think of this latter set as a collection of quasi-true parameters, motivating the following definition.

DEFINITION. *When $\boldsymbol{\theta}_0$ is unidentified, the subset of the parameter space given by*

$$\Theta_{0n} = \{\boldsymbol{\theta} \in \Theta : \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \mathbf{0}\}$$

will be referred to as the quasi-true parameter space.

It is perhaps worth noting that whereas $\boldsymbol{\theta}_0$ is uniquely determined via the DGP as the parameter value for which the moment conditions $E[\boldsymbol{\mu}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$ hold, and is identified if (2.1) holds, Θ_{0n} delineates those values $\boldsymbol{\theta}$ for which $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0) = \mathbf{0}$ and (2.1) is violated, and in general, the quasi-true parameter space will, as with $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$, depend on n.

LEMMA 3.1. *Let $\bar{Q}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$. If Assumptions 2.1–2.3 hold, then $|Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta})| \xrightarrow{p} 0$ uniformly in $\boldsymbol{\theta}$.*

Now, observe that $\bar{Q}_n(\boldsymbol{\theta}) = \bar{Q}_n(\boldsymbol{\theta}_0) = 0$ for all $\boldsymbol{\theta} \in \Theta_{0n}$. Combining the result in Lemma 3.1 with the non-uniqueness implicit in Theorem 3.1, we can anticipate that under suitable regularity conditions, the GMM solution set $\widehat{\Theta}_n = \{\boldsymbol{\theta} \in \Theta : Q_n(\boldsymbol{\theta}) = \min_{\theta \in \Theta} Q_n(\boldsymbol{\theta})\}$ will converge to the quasi-true parameter space.

To characterize the behavior of the GMM estimator when the Jacobian is rank deficient and the model is only partially identified, we will invoke the concept of Hausdorff–Kuratowski set convergence. The distance between a point $\boldsymbol{\theta}$ in Θ and a subset A of Θ is given by

$$d(\boldsymbol{\theta}; A) = \begin{cases} \inf_{\boldsymbol{\theta}_A \in A}\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_A\|\}, & \text{for} \quad A \neq \emptyset, \\ \infty, & \text{for} \quad A = \emptyset \end{cases}$$

and the neighborhood $N(A; \epsilon) = \{\boldsymbol{\theta} : d(\boldsymbol{\theta}; A) < \epsilon\}$. The Hausdorff distance between subsets A and B of Θ is defined as

$$d_H(A; B) = \max\{D(A; B); D(B; A)\},$$

where

$$\begin{aligned} D(A; B) &= \inf\{\epsilon > 0 : A \subseteq N(B; \epsilon)\} \\ &= \sup_{\theta \in A} d(\boldsymbol{\theta}; B), \end{aligned}$$

unless both A and B are empty, in which case $d_H(A; B) = 0$. For any sequence of compact subsets $A_n \subseteq \Theta$, the Kuratowski inferior and superior limits of A_n as $n \to \infty$ are defined as

$$\liminf_{n \to \infty} A_n = \left\{\boldsymbol{\theta} \in \Theta : \limsup_{n \to \infty} d(\boldsymbol{\theta}, A_n) = 0\right\}$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

and

$$\limsup_{n\to\infty} A_n = \left\{\boldsymbol{\theta} \in \Theta : \liminf_{n\to\infty} d(\boldsymbol{\theta}, A_n) = 0\right\},$$

respectively. If the Kuratowski inferior and superior limits agree, i.e., they are the same subset of Θ, then their common value is called the Kuratowski limit of the sets A_n as $n \to \infty$. For compact metric spaces, Kuratowski convergence coincides with convergence of the Hausdorff metric (Ambrisio and Tilli, 2004). We will therefore say that the sequence A_n is Hausdorff–Kuratowski convergent to a closed subset A of Θ if and only if $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = A$, or equivalently $\lim_{n\to\infty} d_H(A_n; A) = 0$. Hausdorff–Kuratowski convergence of two sequences, $\lim_{n\to\infty} d_H(A_n; B_n) = 0$, is defined similarly *mutatis mutandis*.

THEOREM 3.2. *Suppose that Assumptions 2.1–2.3 hold and that* $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) \in \mathbb{R}^k \backslash \{\mathbf{x} : \bar{\boldsymbol{\Omega}}_n \mathbf{x} = 0\}$, *for all* $\boldsymbol{\theta} \in \Theta \backslash \Theta_{0n}$. *Then the sequence of GMM solution sets* $\widehat{\Theta}_n, \widehat{\Theta}_{n+1}, \widehat{\Theta}_{n+2}, \ldots$ *associated with* $\mathbf{W}_n$ *is Hausdorff–Kuratowski convergent to* Θ_{0n} *with probability approaching one, that is, for all* $\epsilon > 0$, *there exists an* $n_\epsilon \geq 1$ *such that* $\Pr[d_H(\widehat{\Theta}_n; \Theta_{0n}) \leq \epsilon] > 1 - \epsilon$, *for all* $n > n_\epsilon$.

Imposition of the condition that $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$ does not belong to the null space of $\bar{\boldsymbol{\Omega}}_n$ when $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) \neq \mathbf{0}$ is required in Theorem 3.2 since otherwise it would be possible for there to exist a parameter sequence $\boldsymbol{\theta}^n$ with $\boldsymbol{\theta}^n \in \Theta \backslash \Theta_{0n}$ such that $Q_n(\boldsymbol{\theta}^n)$ converges to zero as $n \to \infty$. Lemma 3.2 in the following section indicates that an optimal choice of weight matrix is obtained when $\bar{\boldsymbol{\Omega}}_n = \bar{\boldsymbol{\Sigma}}_{0n}^+$, and this choice removes the possibility that $\bar{\boldsymbol{\Omega}}_n \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \mathbf{0}$ for $\boldsymbol{\theta} \in \Theta \backslash \Theta_{0n}$ as the g-inverse eliminates redundant moments in $\bar{Q}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})' \bar{\boldsymbol{\Omega}}_n \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$ and $\bar{Q}_n(\boldsymbol{\theta}) = 0$ if and only if $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) = \mathbf{0}$. Hausdorff distance has, of course, been employed elsewhere to study convergence properties in situations where a set rather than a point is the focus of interest. Recognizing that $\widehat{\Theta}_n$ corresponds to the minimal contour level set of $Q_n(\boldsymbol{\theta})$, an appeal to Theorem 3.1 of Chernozhukov et al. (2007) can be made to show that $d_H(\widehat{\Theta}_n; \Theta_{0n}) = O_p(n^{-\frac{1}{2}})$ for any $q_n = r\{\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} \neq 0$. For completeness, a direct proof of the weaker result presented in Theorem 3.2 is provided in the Supplementary Material.

3.2. Estimable Functions and Asymptotic Normality

When the model is identified, it is well known that $\widehat{\boldsymbol{\theta}}_n$ is consistent for the true parameter value $\boldsymbol{\theta}_0$ and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges in distribution (see Hall, 2005, Thm. 3.2, for example). In order to examine the behavior of $\widehat{\boldsymbol{\theta}}_n$ when the model need not be identified, we will derive the asymptotic properties of (functions of) $(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ where, for any $\widehat{\boldsymbol{\theta}}_n \in \widehat{\Theta}_n$, we will set

$$\boldsymbol{\theta}_{0n} = \arg\min_{\theta \in \Theta_{0n}} \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|, \tag{3.1}$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

the projection of $\widehat{\boldsymbol{\theta}}_n$ on to Θ_{0n}.[6] Since $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}\| = d(\widehat{\boldsymbol{\theta}}_n; \Theta_{0n}) \leq D(\widehat{\Theta}_n, \Theta_{0n})$, it follows from Theorem 3.2 that $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}\| < \epsilon$ and we can see that the large-sample behavior of $\widehat{\boldsymbol{\theta}}_n$ is being derived under a scenario akin to that employed when using a Pitman sequence to analyze the properties of an estimator, only here $\boldsymbol{\theta}_{0n}$ is drifting around in the quasi-true parameter set Θ_{0n} such that $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_{0n}$ form a matching pair and $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}\|$ approaches zero as $\widehat{\boldsymbol{\theta}}_n$ enters $N(\Theta_{0n}, \epsilon)$.

Results on estimable functions will be stated and derived by first considering the following combinations of identification rank and identification strength as previously defined in Section 2; (i) identification rank $q_n = p$ and identification strength is strong or semi-strong, and (ii) identification rank $q_n < p$ and identification strength is mixed with $\lim_{n\to\infty} n^{\frac{1}{2}} \bar{s}_{0n,q_n} \to \infty$ and $n^{\frac{1}{2}} \|\bar{\boldsymbol{\delta}}_{0n,c}\| = (\sum_{r=1}^{q_n} \bar{s}_{0n,r}^2 \bar{v}_{0n,rc}^2)^{\frac{1}{2}} \to \infty, c = 1, \ldots, p$. Subsequently, the results will be extended to cover cases where the identification strength can be weak. This way of proceeding allows us to separate strict identification failure where $q_n < p$ from the less rigid notion of identification deficiency inherent in the concept of weakness.

Rewrite the first-order condition in (2.4) that defines $\widehat{\boldsymbol{\theta}}_n$ as an implicit function of the data as

$$\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0},$$

wherein we abbreviate $\mathbf{D}_n(\mathbf{w}, \boldsymbol{\theta})$ to $\mathbf{D}_n(\boldsymbol{\theta})$, and $\mathbf{m}_n(\mathbf{w}, \boldsymbol{\theta})$ to $\mathbf{m}_n(\boldsymbol{\theta})$. Expanding the moment $\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n)$ about $\mathbf{m}_n(\boldsymbol{\theta}_{0n})$ gives

$$\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \mathbf{D}_n(\boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}), \tag{3.2}$$

where $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_{0n}\| \leq \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}\|$, and substituting this into the first-order condition, we have

$$\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \mathbf{D}_n(\boldsymbol{\theta}^*) \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \sqrt{n} \mathbf{m}_n(\boldsymbol{\theta}_{0n}). \tag{3.3}$$

The general solution to (3.3) can be expressed as

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -\mathbf{G}_n \mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \sqrt{n} \mathbf{m}_n(\boldsymbol{\theta}_{0n}) + (\mathbf{H}_n - \mathbf{I})\mathbf{z}, \tag{3.4}$$

where $\mathbf{G}_n = (\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \mathbf{D}_n(\boldsymbol{\theta}^*))^+$, $\mathbf{H}_n = (\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \mathbf{D}_n(\boldsymbol{\theta}^*))^+ (\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \mathbf{D}_n(\boldsymbol{\theta}^*))$ and $\mathbf{z}$ is arbitrary (Rao and Mitra, 1971, Thm. 2.3.1(b)).

Theorem 3.2 implies that $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}\| \xrightarrow{p} 0$, and from the assumed convergence of $\mathbf{D}_n(\boldsymbol{\theta})$ to $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$, it follows that $\|\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\| \xrightarrow{p} 0$ and $\|\mathbf{D}_n(\boldsymbol{\theta}^*) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\| \xrightarrow{p} 0$. From Slutsky's theorem, we can therefore conclude that

$$\|\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n - \bar{\boldsymbol{\Delta}}_{0n}' \bar{\boldsymbol{\Omega}}_n\| \xrightarrow{p} 0 \text{ and } \|\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)' \mathbf{W}_n \mathbf{D}_n(\boldsymbol{\theta}^*) - \bar{\boldsymbol{\Delta}}_{0n}' \bar{\boldsymbol{\Omega}}_n \bar{\boldsymbol{\Delta}}_{0n}\| \xrightarrow{p} 0, \tag{3.5}$$

[6]That $\boldsymbol{\theta}_{0n}$ is the projection of $\widehat{\boldsymbol{\theta}}_n$ on to Θ_{0n} is established in the Supplementary Material.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

where $\bar{\mathbf{\Delta}}_{0n} = \bar{\mathbf{\Delta}}_n(\boldsymbol{\theta}_{0n})$, and since (under suitable regularity) $\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_0)$ is asymptotically normal, the general solution in (3.4) can be reexpressed as

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -\bar{\mathbf{\Gamma}}_{0n}\bar{\mathbf{\Delta}}_{0n}^{'}\bar{\mathbf{\Omega}}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + (\bar{\mathbf{H}}_{0n} - \mathbf{I})\mathbf{z} + o_p(1), \tag{3.6}$$

where $\bar{\mathbf{\Gamma}}_{0n} = (\bar{\mathbf{\Delta}}_{0n}^{'}\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Delta}}_{0n})^+$ and $\bar{\mathbf{H}}_{0n} = (\bar{\mathbf{\Delta}}_{0n}^{'}\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Delta}}_{0n})^+(\bar{\mathbf{\Delta}}_{0n}^{'}\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Delta}}_{0n})$.

When the model is identified, $r\{\bar{\mathbf{\Delta}}_{0n}\} = p$, $\boldsymbol{\theta}_{0n} = \boldsymbol{\theta}_0$, and the equation system (3.6) has a unique solution for $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ which converges in distribution because $\bar{\mathbf{\Delta}}_{0n}^{'}\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Delta}}_{0n}$ is invertible and $\bar{\mathbf{H}}_{0n} = \mathbf{I}$. When $r\{\bar{\mathbf{\Delta}}_{0n}\} = q_n < p$, however, $\bar{\mathbf{\Delta}}_{0n}^{'}\bar{\mathbf{\Omega}}_n\bar{\mathbf{\Delta}}_{0n}$ is singular and many solutions to (3.6) exist. A unique solution could be generated by combining (3.6) with the constraint $\mathbf{P}\boldsymbol{\theta} - (\boldsymbol{\beta}', \boldsymbol{\alpha}(\boldsymbol{\beta})') = \mathbf{0}$ so as to extract the re-scaled deviation $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{0n})$ for the parameter $\boldsymbol{\beta}$. But as previously observed, this is not feasible. In the absence of such parameter constraints, alternative conditions that confine the infinite number of possible solutions to (3.6) to a feasible finite set are required.

By Rao and Mitra (1971, Thm. 2.3.1(c)), the linear combination $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ is unique for any vector $\mathbf{q} = (q_1, \ldots, q_p)'$ if (and only if) $\mathbf{q}'\bar{\mathbf{H}}_{0n} = \mathbf{q}'$. This condition is satisfied by any vector of the form $\mathbf{q}' = \mathbf{z}'\bar{\mathbf{H}}_{0n}$ because $\mathbf{q}'\bar{\mathbf{H}}_{0n} = \mathbf{z}'\bar{\mathbf{H}}_{0n}^2 = \mathbf{z}'\bar{\mathbf{H}}_{0n} = \mathbf{q}'$ since $\bar{\mathbf{H}}_{0n}$ is idempotent. Thus, by Theorem 3.2, for an arbitrary vector $\mathbf{z}$, the linear combination $\mathbf{q}'\widehat{\boldsymbol{\theta}}_n$, where $\mathbf{q}' = \mathbf{z}'\bar{\mathbf{H}}_{0n}$ will be consistent for the same linear combination $\mathbf{q}'\boldsymbol{\theta}_{0n}$ of the parameter $\boldsymbol{\theta}_{0n} \in \Theta_{0n}$, and as will be shown below, for such linear combinations $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ is asymptotically normal. Henceforth we will therefore label a linear combination $\mathbf{q}_{0n}'\boldsymbol{\theta}$ where the vector $\mathbf{q}_{0n} = \mathbf{z}'\bar{\mathbf{H}}_{0n}$ an *estimable function*.

Borrowing notation from the empirical process literature, let

$$\mathbb{G}_n(\boldsymbol{\theta}) = n^{\frac{1}{2}}\{\mathbb{P}_n - \mathrm{P}_n\}\boldsymbol{\mu}(\boldsymbol{\theta}) = n^{\frac{1}{2}}\{\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\}$$

define an empirical process $\mathbb{G}_n(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta$ where, for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, the covariance kernel

$$\bar{\mathbf{\Sigma}}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = nE[(\mathbf{m}_n(\boldsymbol{\theta}_1) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_1)(\mathbf{m}_n(\boldsymbol{\theta}_2) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_2)'].$$

Now, suppose that the re-centered and re-scaled sample moments satisfy the following functional CLT (FCLT) assumption, wherein the symbol $\Rightarrow$ denotes weak convergence of random functions on Θ with respect to the supremum norm, and $\bar{\mathbf{\Sigma}}_n^{\frac{1}{2}}(\boldsymbol{\theta})$ denotes the positive semi-definite symmetric square root of $\bar{\mathbf{\Sigma}}_n(\boldsymbol{\theta})$ where $\bar{\mathbf{\Sigma}}_n(\boldsymbol{\theta}) = \bar{\mathbf{\Sigma}}_n(\boldsymbol{\theta}, \boldsymbol{\theta})$.

Assumption 3.1. The empirical process $\mathbb{G}_n(\boldsymbol{\theta}) = \bar{\mathbf{\Sigma}}_n^{\frac{1}{2}}(\boldsymbol{\theta})\mathbb{B}_n(\boldsymbol{\theta})$, where $\mathbb{B}_n(\boldsymbol{\theta}) \Rightarrow \mathbb{B}(\boldsymbol{\theta})$, a zero-mean Gaussian stochastic process indexed by $\boldsymbol{\theta} \in \Theta$ with bounded continuous sample paths and an identity covariance kernel.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

Conditions under which the FCLT presented in Assumption 3.1 will hold are discussed in the Supplementary Material.

THEOREM 3.3. *Let* $\mathbf{q}_{0n}'\boldsymbol{\theta}$ *be an estimable function where, for any* $\mathbf{z} \neq \mathbf{0}$, $\mathbf{q}_{0n}' = \mathbf{z}'\bar{\mathbf{H}}_{0n}$, *and suppose that Assumptions 2.1–2.3 and 3.1 hold. Then*

$$\sqrt{n}\mathbf{q}_{0n}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -\mathbf{z}'\bar{\boldsymbol{\Gamma}}_{0n}(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n)\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\mathbb{B}_n(\boldsymbol{\theta}_{0n}) + o_p(1),$$

where $\bar{\boldsymbol{\Sigma}}_{0n} = \bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}_{0n})$ *and* $\mathbb{B}_n(\boldsymbol{\theta}_{0n}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$.

Employing the Cramér–Wold device and Slutsky's theorem in conjunction with Theorem 3.3, we can conclude that the distribution of $\sqrt{n}\mathbf{q}_{0n}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ will be approximately normal with zero mean and variance

$$\mathbf{z}'\bar{\boldsymbol{\Gamma}}_{0n}(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}\bar{\boldsymbol{\Omega}}_n'\bar{\boldsymbol{\Delta}}_{0n})\bar{\boldsymbol{\Gamma}}_{0n}\mathbf{z}. \quad \mathbf{(3.7)}$$

Hence, we can conclude that $\mathbf{q}_{0n}'\widehat{\boldsymbol{\theta}}_n$ is a consistent and asymptotically normal (CAN) estimator of the estimable function $\mathbf{q}_{0n}'\boldsymbol{\theta}$.

The previous development derived the asymptotic distribution of the estimable function $\mathbf{q}_{0n}'\boldsymbol{\theta}$ by expanding the moment conditions, following Hansen (1982). An alternative approach is to follow the route taken by Newey and Windmeijer (2009) and to expand the first-order condition. The latter is a standard approach in the analysis of extremum estimators, and that it remains valid when the model is only partially identified follows from the following companion theorem to Theorem 3.3.

THEOREM 3.4. *Suppose that Assumptions 2.1–2.3 and 3.1 hold and that the moment conditions* $\boldsymbol{\mu}_t(\mathbf{w}_t, \boldsymbol{\theta})$ *are twice continuously differentiable in* $\boldsymbol{\theta}$. *Let* $\mathbf{Q}_n^{(2)}(\boldsymbol{\theta})$ *denote the Hessian matrix* $\partial^2 Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$, *and set* $\mathbf{Q}_{0n}^{(2)} = \mathbf{Q}_n^{(2)}(\boldsymbol{\theta}_{0n})$ *and* $\mathbf{q}_{0n}' = \mathbf{z}'\mathbf{Q}_{0n}^{(2+)}\mathbf{Q}_{0n}^{(2)}$ *where* $\mathbf{Q}_{0n}^{(2+)} = (\mathbf{Q}_{0n}^{(2)})^+$. *Then*

$$\sqrt{n}\mathbf{q}_{0n}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -2\mathbf{z}'\mathbf{Q}_{0n}^{(2+)}(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n)\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\mathbb{B}_n(\boldsymbol{\theta}_{0n}) + o_p(1),$$

where $\bar{\boldsymbol{\Sigma}}_{0n} = \bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta}_{0n})$ *and* $\mathbb{B}_n(\boldsymbol{\theta}_{0n}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$.

When the model is identified, expanding the first-order condition leads to a *sandwich* form for the GMM variance estimator that is consistent under standard and (as is shown in Newey and Windmeijer, 2009) many weak moment asymptotics. An immediate consequence of Theorem 3.4 is that such an estimator remains valid when the model is partially identified with $(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n})^+$ replaced by $2\mathbf{Q}_{0n}^{(2+)}$ in the asymptotic variance formula.

LEMMA 3.2. *If* $r\{\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n'\bar{\boldsymbol{\Delta}}_{0n}\} = r\{\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n}\} = q_n$, *then the asymptotic variance of* $\sqrt{n}\mathbf{q}_{0n}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ *where* $\mathbf{q}_{0n}' = \mathbf{z}'\bar{\mathbf{H}}_{0n}$, $\mathbf{z} \neq \mathbf{0}$, *is bounded below by*

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

$\mathbf{z}'(\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n})^{+}\mathbf{z}$, *and a sufficient condition for the estimator associated with* $\mathbf{W}_n$ *to be asymptotically efficient is that* $\|(\bar{\boldsymbol{\Omega}}_n - \bar{\boldsymbol{\Sigma}}_{0n}^{+})\bar{\boldsymbol{\Delta}}_{0n}\| \to 0$.

Lemma 3.3 indicates that choosing $\mathbf{W}_n$ such that $\|\mathbf{W}_n - \bar{\boldsymbol{\Sigma}}_{0n}^{+}\| \xrightarrow{p} 0$ is optimal, for then

$$\|\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n} - \bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n}\| \leq \left(\|(\bar{\boldsymbol{\Omega}}_n - \mathbf{W}_n)\| + \|\mathbf{W}_n - \bar{\boldsymbol{\Sigma}}_{0n}^{+}\|\right)\cdot\|\bar{\boldsymbol{\Delta}}_{0n}\|,$$

implying that $\|\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n} - \bar{\boldsymbol{\Sigma}}_{0n}^{+}\bar{\boldsymbol{\Delta}}_{0n}\| \xrightarrow{p} 0$. One such choice for a two-step GMM estimator would be $\mathbf{W}_n = \bar{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\theta}}_{n1})^{+}$, where $\widehat{\boldsymbol{\theta}}_{n1}$ is a first-step estimator obtained using a weight matrix $\mathbf{W}_n$ that is independent of $\boldsymbol{\theta}$—and typically not dependent on the data, $\mathbf{W}_n = \mathbf{I}$ for example. Theorem 3.2 indicates that there will exist a $\boldsymbol{\theta}_{0n}$ such that $\|\bar{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\theta}}_{n1}) - \bar{\boldsymbol{\Sigma}}_{0n}\| \xrightarrow{p} 0$, and if $r\{\bar{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\theta}}_{n1})\} = r\{\bar{\boldsymbol{\Sigma}}_{0n}\}$, then $\|\bar{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\theta}}_{n1})^{+} - \bar{\boldsymbol{\Sigma}}_{0n}^{+}\| \xrightarrow{p} 0$. In general, however, explicit use of $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta})$ will be either (i) not feasible because derivation of the long-run variance matrix is intractable, or (ii) too difficult because the structure of $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta})$ is too complicated. The commonly adopted solution is to replace $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\theta})$ by a consistent estimate. If $\boldsymbol{\mu}_t(\mathbf{w}_t,\boldsymbol{\theta})$, $t = 1,\ldots,n$ are independent and identically distributed, for example, $n(\sum_{t=1}^{n}\boldsymbol{\mu}(\mathbf{w}_t,\widehat{\boldsymbol{\theta}}_{n1})\boldsymbol{\mu}(\mathbf{w}_t,\widehat{\boldsymbol{\theta}}_{n1})')^{+}$ can serve for $\mathbf{W}_n$. For further details on estimating the long-run variance matrix, see Hall (2005, Chap. 3.5). Henceforth we will refer to a CAN estimator $\mathbf{q}_{0n}'\widehat{\boldsymbol{\theta}}_n$ of $\mathbf{q}_{0n}'\boldsymbol{\theta}_{0n}$ that satisfies the efficiency bound as a consistent and efficient asymptotically normal (CEAN) estimator.

Suppose that $\mathbf{W}_n$ has been chosen optimally. Then, for any given $\mathbf{z}$, Theorem 3.3 can be implemented using $\mathbf{z}'(\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n))^{+}\mathbf{z}$ to estimate the variance in (3.7) since by Theorem 3.2 $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}\| = o_p(1)$. Similarly, $\mathbf{q}_{0n}' = \mathbf{z}'\bar{\mathbf{H}}_{0n}$ can be estimated by $\mathbf{z}'\widehat{\mathbf{H}}_{0n}$ where $\widehat{\mathbf{H}}_{0n} = (\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n))^{+}(\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n))$. Unknown ensemble averages are thereby replaced by sample statistics, circumventing the fact that in practice $\boldsymbol{\theta}_{0n}$ will not be known. The applied researcher will want to conduct inference on the true parameter $\boldsymbol{\theta}_0$ of course. Whenever $\boldsymbol{\theta}_{0n} \in N(\boldsymbol{\theta}_0;\delta)$, then $\|\bar{\mathbf{H}}_{0n}(\boldsymbol{\theta}_{0n} - \boldsymbol{\theta}_0)\| \leq \delta$, and if $\delta = o(n^{-\frac{1}{2}})$, we can conclude that $\sqrt{n}\mathbf{z}'\bar{\mathbf{H}}_{0n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = \sqrt{n}\mathbf{z}'\bar{\mathbf{H}}_{0n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o(1)$. Inference on $\boldsymbol{\theta}_0$ can then be conducted by constructing hypotheses tests or confidence regions based on the asymptotic normality of $\sqrt{n}\mathbf{z}'\widehat{\mathbf{H}}_{0n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ as given in Theorem 3.3.

The distributional properties presented in Theorems 3.3 and 3.4 presuppose that the identification strength of identified and partially identified parameters is strong. In general, however, the behavior of estimable functions will depend on the level of individual parameter identification strength. To delineate the range of parameter identification strength that may occur, we will follow the existing literature and characterize parameter identification strength by reference to a matrix-valued normalizing sequence and set $\bar{\mathbf{J}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\mathbf{N}_n$, where $\mathbf{N}_n = \mathrm{diag}\{n^{\lambda_1},\ldots,n^{\lambda_p}\}$, $0 \leq \lambda_c \leq 0.5$, $c = 1,\ldots,p$. Let $\bar{\mathbf{J}}_{0n} = \bar{\mathbf{J}}_n(\boldsymbol{\theta}_{0n}) = [\bar{\mathbf{j}}_{0n,1},\ldots,\bar{\mathbf{j}}_{0n,p}]$. The identification strength of the cth component of $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{n1},\ldots,\widehat{\theta}_{np})'$ will be reflected in the value of $\lim_{n\to\infty} n^{-\lambda_c}\|\bar{\mathbf{j}}_{0n,c}\|$, $c = 1,\ldots,p$, and adopting commonly employed labels we

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

will divide $\boldsymbol{\theta}$ into strongly, moderately, and weakly identified parameter subsets according to whether $\lambda_c = 0$, $0 < \lambda_c < 0.5$, or $\lambda_c = 0.5$, respectively. These more conventional parameter identification strength labels translate into the moment condition identification strength categories via the relationship

$$\lim_{n\to\infty} n^{\frac{1}{2}} \|\bar{\boldsymbol{\delta}}_{0n,c}\| = \lim_{n\to\infty} n^{\frac{1}{2}} \left(\sum_{r=1}^{q_n} \bar{s}^2_{0n,r} \bar{v}^2_{0n,rc} \right)^{\frac{1}{2}} = \lim_{n\to\infty} \|\bar{\mathbf{j}}_{0n,c}\| n^{\frac{1}{2}-\lambda_c},$$

from which it follows that

$$\lim_{n\to\infty} n^{\lambda_c} \bar{s}_{0n,q_n} \left(\sum_{r=1}^{q_n} \bar{v}^2_{0n,rc} \right)^{\frac{1}{2}} \le \lim_{n\to\infty} \|\bar{\mathbf{j}}_{0n,c}\| \le \lim_{n\to\infty} n^{\lambda_c} \bar{s}_{0n,1} \left(\sum_{r=1}^{q_n} \bar{v}^2_{0n,rc} \right)^{\frac{1}{2}}$$

and the parameter identification strength labels need not match the moment condition identification strength categories. For example, if while the moment condition identification strength is mixed $\lim_{n\to\infty} n^{\lambda_c} \bar{s}_{0n,1} < \infty$ when $0 \le \lambda_c < 0.5$, the parameter identification strength may appear to be strong, rather than moderate, a case called "joint weak identification" in Andrews and Guggenberger (2017).

In order to examine the effects of variations in parameter identification strength, we will add the following assumption.

Assumption 3.2. Let

$$\mathbb{F}_n(\boldsymbol{\theta}) = n^{\frac{1}{2}} \{\mathbb{P}_n - \mathrm{P}_n\} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = n^{\frac{1}{2}} \{\mathbf{D}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} : \boldsymbol{\theta} \in \Theta$$

denote the empirical process with covariance kernel

$$\bar{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = nE[\mathrm{vec}\{\mathbf{D}_n(\boldsymbol{\theta}_1) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_1)\}\mathrm{vec}\{\mathbf{D}_n(\boldsymbol{\theta}_2) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_2)\}'].$$

Then $\mathrm{vec}\{\mathbb{F}_n(\boldsymbol{\theta})\} = \bar{\boldsymbol{\Lambda}}_n^{\frac{1}{2}}(\boldsymbol{\theta})\mathbb{W}_n(\boldsymbol{\theta})$ where $\bar{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}, \boldsymbol{\theta})$ and $\mathbb{W}_n(\boldsymbol{\theta}) \Rightarrow \mathbb{W}(\boldsymbol{\theta})$, a zero-mean Gaussian stochastic process indexed by $\boldsymbol{\theta} \in \Theta$ with bounded continuous sample paths and an identity covariance kernel. Furthermore, the joint distribution of $\mathrm{vec}\{\mathbb{F}_n(\boldsymbol{\theta})\}$ and $\mathbb{G}_n(\boldsymbol{\theta})$ is given by

$$\begin{bmatrix} \mathbb{G}_n(\boldsymbol{\theta}) \\ \mathrm{vec}\{\mathbb{F}_n(\boldsymbol{\theta})\} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \bar{\mathbf{C}}(\boldsymbol{\theta})\bar{\boldsymbol{\Sigma}}_n^{+}(\boldsymbol{\theta}) & \mathbf{I} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\Sigma}}_n^{\frac{1}{2}}(\boldsymbol{\theta})\mathbb{B}_n(\boldsymbol{\theta}) \\ \bar{\boldsymbol{\Lambda}}_n^{\frac{1}{2}}(\boldsymbol{\theta})\mathbb{W}_n(\boldsymbol{\theta}) \end{bmatrix},$$

wherein $\mathbb{W}_n(\boldsymbol{\theta})$ is asymptotically independent of $\mathbb{B}_n(\boldsymbol{\theta})$.

Suppose that there exists a permutation of the parameter vector $\mathbf{P}\boldsymbol{\theta} = (\theta_{c(1)}, \ldots, \theta_{c(p)})'$ such that $\bar{\mathbf{J}}(\boldsymbol{\theta})\mathbf{P} = \bar{\boldsymbol{\Delta}}(\boldsymbol{\theta})\mathbf{P}\mathbf{M}_n$ where $\mathbf{M}_n = \mathbf{P}'\mathbf{N}_n\mathbf{P} = \mathrm{diag}\{n^{\lambda_{c(1)}}, \ldots, n^{\lambda_{c(p)}}\}$ where $\lambda_{c(i)} = 0.0$, for $i = 1, \ldots, p_s$, $0 < \lambda_{c(i)} < 0.5$, for $i = p_s + 1, \ldots, p_s + p_m$, and $\lambda_{c(i)} = 0.5$, for $i = p - p_w + 1, \ldots, p$, $p_s + p_m + p_w = p$. The consequences for estimable functions are summarized in the following theorem.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

THEOREM 3.5. *Suppose that Assumptions 2.1–2.3, 3.1, and 3.2 hold. Then, for any* $\mathbf{z} \neq \mathbf{0}$ *and permuted parameter vector* $\mathbf{P}\boldsymbol{\theta}$*, we have* $\mathbf{q}_{0n}'\mathbf{P}\boldsymbol{\theta}$ *is an estimable function whenever* $\mathbf{q}_{0n}' = \mathbf{z}'\mathbf{H}_{0n}$ *where* $\mathbf{H}_{0n} = (\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})^+(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})$*, and*

$$\sqrt{n}\mathbf{q}_{0n}'\mathbf{M}_n^{-1}\mathbf{P}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -\mathbf{z}'(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})^+(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n)\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\mathbb{B}_n(\boldsymbol{\theta}_{0n}) + o_p(1),$$

where $\mathbf{K}_{0n} = \bar{\mathbf{J}}_{0n}\mathbf{P} + \mathbf{F}_{0n}\mathbf{P}\mathcal{S}_w$*,* $\mathcal{S}_w = diag(\overbrace{0,\ldots,0}^{p_s+p_m}, \overbrace{1,\ldots,1}^{p_w})$*, and* $(\mathbf{P}'\otimes\mathbf{I})vec\{\mathbf{F}_{0n}\} = (\mathbf{P}'\otimes\mathbf{I})\bar{\boldsymbol{\Lambda}}_{0n}^{\frac{1}{2}}\mathbb{W}_n(\boldsymbol{\theta}_{0n})$ *where* $\bar{\boldsymbol{\Lambda}}_{0n} = \bar{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}_{0n})$ *and* $\mathbb{W}_n(\boldsymbol{\theta}_{0n}) \Rightarrow \mathcal{N}(\mathbf{0},\mathbf{I}_{kp})$*.*

Theorem 3.5 is analogous to Theorem 1 of Stock and Wright (2000), and Theorem 2 of Guggenberger and Smith (2005), wherein the parameter vector is shown to converge in distribution under (Stock and Wright, 2000, Assum. C) to a functional of a weakly convergent sequence. Theorem 3.5 is derived without recourse to Stock and Wright (2000, Assum. C), and it relates to the convergence in probability of estimable functions to a random variable comprised of nonstandard convolutions of functions of singular inverted Wishart covariance matrix Gaussian mixtures.[7]

From the identity $\sqrt{n}\mathbf{q}_{0n}'\mathbf{M}_n^{-1}\mathbf{P}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) \equiv \sum_{i=1}^{p} n^{(0.5-\lambda_{c(i)})}q_{on,c(i)}(\widehat{\theta}_{nc(i)} - \theta_{0n,c(i)})$, it is apparent that if the parameter identification strength is strong overall ($p_s = p$), then the previous limit theory in Theorem 3.3 will apply with a standard $n^{\frac{1}{2}}$-convergence rate. If the parameter identification strength is only moderate to strong ($p_m + p_s = p$), consistency and asymptotic normality will still hold, but the convergence rate of the components $n^{(0.5-\lambda_{c(i)})}q_{c(i)}(\widehat{\theta}_{nc(i)} - \theta_{0n,c(i)})$, $i = p_s+1,\ldots,p$, could be quite slow relative to the standard $\sqrt{n}$ rate if $\lambda_{c(i)}$ is close to 0.5. These convergence rates may be amplified or attenuated via the influence of $\bar{\mathbf{J}}_{0n}\mathbf{P}$. When the parameter identification strength is strong and $\|\bar{\mathbf{J}}_{0n}\| = O(n^{-\kappa})$, for example, the $n^{\frac{1}{2}}$-convergence rate will be deflated if the moment identification strength is semi-strong and $0 < \kappa < \frac{1}{2}$. If the parameter identification strength is partly weak, however ($\lambda_{c(i)} = 0.5$, for $i = p - p_w + 1,\ldots,p$ where $1 \leq p_w < p$), Theorem 3.5 indicates that uncertainty will be retained in the limit by way of estimable functions containing a random component with a non-degenerate non-normal asymptotic distribution. Thus, in a partially identified model with mixed levels of identification strength, GMM estimates of estimable functions will exhibit different large-sample behavior according to the dictates of the identification strength attributable to individual parameter estimates. Since the latter will be unknown, this does not augur well for their application in the construction of test statistics or confidence regions.

[7]The singular inverted Wishart distribution is defined as the distribution of the Moore–Penrose inverse of a singular Wishart distributed matrix (see Bodnar, Mazur, and Podgorski, 2016; Srivastava, 2003 for details). Theorem 3.5 provides a generalization of Corollary 3.1(b) to Theorem 3.1 of Choi and Phillips (1992) where the asymptotic distribution of the IV estimator is shown to be a non-degenerate covariance matrix mixture (see also Staiger and Stock (1997, Sect. 2B, p. 562))). The asymptotic results presented in Choi and Phillips (1992) derive from the central limit theory given in Phillips (1989), which also delivers the limit theory used in Staiger and Stock (1997) .

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

4. ILLUSTRATIONS I

4.1. The Linear Equations Model

Here, we will illustrate the previous results via what might be regarded as a leading case of GMM; namely the classical linear equations model with structural equation

$$y_t = \mathbf{x}_t'\boldsymbol{\theta} + u_t, \quad t = 1, \ldots, n, \tag{4.1}$$

and reduced form

$$\mathbf{x}_t = \boldsymbol{\Pi}\boldsymbol{\xi}_t + \mathbf{v}_t, \quad t = 1, \ldots, n, \tag{4.2}$$

where $\mathbf{x}_t' = (x_{1t}, \ldots, x_{pt})$ is a vector of endogenous regressors, $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_p)$, $\boldsymbol{\xi}_t' = (\xi_{1t}, \ldots, \xi_{kt})$ is a vector of IVs that is uncorrelated with $\boldsymbol{\nu}_t' = (u_t, \mathbf{v}_t')$, and the reduced-form parameter $\boldsymbol{\Pi}$ is $p \times k$, $k \geq p$. It will be assumed that $n^{-1}\sum_{t=1}^n \boldsymbol{\nu}_t\boldsymbol{\nu}_t' \xrightarrow{p} \mathbf{V} = E[\boldsymbol{\nu}_t\boldsymbol{\nu}_t']$, $n^{-1}\sum_{t=1}^n \boldsymbol{\xi}_t\boldsymbol{\xi}_t' \xrightarrow{p} \boldsymbol{\Xi} = E[\boldsymbol{\xi}_t\boldsymbol{\xi}_t']$, and that $n^{-\frac{1}{2}}\sum_{t=1}^n \boldsymbol{\xi}_t \otimes \boldsymbol{\nu}_t' \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi} \otimes \mathbf{V})$ where $\boldsymbol{\Xi} \otimes \mathbf{V}$ is positive definite. For the purposes of this simulation, exercise $\boldsymbol{\xi}_t \otimes \boldsymbol{\nu}_t'$, $t = 1, \ldots, n = 250$, were generated as independent and identically distributed $\mathcal{N}(\mathbf{0}, \boldsymbol{\Xi} \otimes \mathbf{V})$ random variables with $\boldsymbol{\Xi} = \mathbf{I}_k$ and

$$\mathbf{V} = \begin{bmatrix} 1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \mathbf{T} \end{bmatrix},$$

where the vector $\boldsymbol{\rho} = (\rho, \ldots, \rho)'$ and $\mathbf{T}$ is a pth-order Toeplitz matrix with the first row $(1, \rho, \rho^2, \ldots, \rho^{p-1})$. A mild degree of endogeneity was induced by setting $\rho = 0.5$. The true structural equation coefficient was set equal to the p element sum vector, $\boldsymbol{\theta}_0 = (1, \ldots, 1)'$.

The above simultaneous equations IV model maps into the previous GMM notation via $\mathbf{w}_t' = (y_t, \mathbf{x}_t', \boldsymbol{\xi}_t')$, $\boldsymbol{\mu}(\mathbf{w}_t, \boldsymbol{\theta}) = \boldsymbol{\xi}_t(y_t - \mathbf{x}_t'\boldsymbol{\theta})$ and $\partial\boldsymbol{\mu}(\mathbf{w}_t, \boldsymbol{\theta})/\partial\boldsymbol{\theta}' = -\boldsymbol{\xi}_t\mathbf{x}_t'$. Writing $y_t = \boldsymbol{\xi}_t'\boldsymbol{\pi} + (u_t + \mathbf{v}_t'\boldsymbol{\theta})$ for the reduced form for y_t leads to the expression $\boldsymbol{\Xi}(\boldsymbol{\pi} - \boldsymbol{\Pi}'\boldsymbol{\theta})$ for $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$. The long-run variance $\bar{\boldsymbol{\Sigma}}_{0n} = \sigma_u^2\boldsymbol{\Xi}$ where $\sigma_u^2 = E[u_t^2]$ and the optimal weighting matrix corresponds to $\bar{\boldsymbol{\Omega}}_n = n\sigma_u^2(\sum_{t=1}^n \boldsymbol{\xi}_t\boldsymbol{\xi}_t')^+$. The Jacobian $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}) = -\boldsymbol{\Xi}\boldsymbol{\Pi}'$. From the equality $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0) = -\boldsymbol{\Xi}\boldsymbol{\Pi}'(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, it follows that $\Theta_{0n} = \{\boldsymbol{\theta} : \boldsymbol{\Xi}\boldsymbol{\Pi}'(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \mathbf{0}\}$ and hence that $\bar{\mathbf{H}}_{0n}\boldsymbol{\theta}_{0n} = \bar{\mathbf{H}}_{0n}(\bar{\mathbf{H}}_{0n}\boldsymbol{\theta}_0)$ where $\bar{\mathbf{H}}_{0n} = \mathbf{H} = (\boldsymbol{\Pi}\boldsymbol{\Xi}\boldsymbol{\Pi}')^+(\boldsymbol{\Pi}\boldsymbol{\Xi}\boldsymbol{\Pi}')$. From the results presented in Section 4, we also know that $\mathbf{q}'\widehat{\boldsymbol{\theta}}_n$ will be a CEAN estimator of $\mathbf{q}'\boldsymbol{\theta}_{0n} = \mathbf{z}'\mathbf{H}\boldsymbol{\theta}_{0n} = \mathbf{z}'\mathbf{H}\boldsymbol{\theta}_0 = \mathbf{q}'\boldsymbol{\theta}_0$ for any vector of the form $\mathbf{q}' = \mathbf{z}'\mathbf{H}$ where $\mathbf{z}$ is arbitrary.

For this model, the Jacobian $-\boldsymbol{\Xi}\boldsymbol{\Pi}'$ is independent of $\boldsymbol{\theta}$, and since by assumption $\boldsymbol{\Xi}$ is positive definite, the identification of $\boldsymbol{\theta}$ depends solely upon the reduced-form coefficient. From an experimental perspective, this represents a useful simplification since it allows the degree of partial identification and the moment condition identification strength to be controlled directly via manipulations of the singular values of $\boldsymbol{\Pi}$. The singular value decomposition $\boldsymbol{\Pi}' = \mathbf{USV}$ was constructed by drawing the columns of $\mathbf{U}$ $(k \times p)$ and $\mathbf{V}$ $(p \times p)$ randomly from the uniform distributions in $\mathcal{O}(k)$ and $\mathcal{O}(p)$, respectively, and fixing the singular values via

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

the specification $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_q, 0, \ldots, 0)$ so that the identification rank equals $r\{\boldsymbol{\Pi}\} = q \leq p$. The individual parameter identification strength, which is here equivalent to instrument strength, was controlled by setting $\boldsymbol{\Pi}' = \bar{\mathbf{J}}\mathbf{N}_n^{-1}$ where $\bar{\mathbf{J}} = \mathbf{USV}$ and $\mathbf{N}_n$ is chosen to give the desired combination of strong, moderate, and weak identification with $\mathbf{P} = \mathbf{I}$.

For the classical linear equations model, it is well known that the efficient two-step GMM estimator reduces to the two-stage least squares (TSLS) estimator. In the current setting, the TSLS solution set is given by those parameter values that solve the second-stage normal equations,

$$\widehat{\Theta}_n = \left\{\boldsymbol{\theta} : (\mathbf{X}'\mathbb{P}_\xi\mathbf{X})\boldsymbol{\theta} = \mathbf{X}'\mathbb{P}_\xi\mathbf{y}\right\},$$

where $\mathbf{X}' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $\mathbf{y} = (y_1, \ldots, y_n)'$, and the projection matrix $\mathbb{P}_\xi = \mathbb{X}(\mathbb{X}'\mathbb{X})^+\mathbb{X}'$, where $\mathbb{X}' = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)$. A typical element of $\widehat{\Theta}_n$ is given by

$$\widehat{\boldsymbol{\theta}}_n = (\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+\mathbf{X}'\mathbb{P}_\xi\mathbf{y} + ((\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+(\mathbf{X}'\mathbb{P}_\xi\mathbf{X}) - \mathbf{I})\mathbf{z}, \tag{4.3}$$

where $\mathbf{z}$ is arbitrary. Substituting $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}_0 + \mathbf{u}$ into (4.3) and taking limits as $n \to \infty$ yields the result that $d_H(\widehat{\Theta}_n; \{\boldsymbol{\theta} : \boldsymbol{\theta} = \mathbf{H}\boldsymbol{\theta}_0 + (\mathbf{H} - \mathbf{I})\mathbf{z}\}) \xrightarrow{p} 0$, and hence that $d_H(\widehat{\Theta}_n; \Theta_{0n}) \xrightarrow{p} 0$ since $\{\boldsymbol{\theta} : \boldsymbol{\theta} = \mathbf{H}\boldsymbol{\theta}_0 + (\mathbf{H} - \mathbf{I})\mathbf{z}\} = \Theta_{0n}$ because $\boldsymbol{\Xi}\,\boldsymbol{\Pi}'(\mathbf{H} - \mathbf{I}) = \mathbf{0}$ (*cf.* Theorem 3.2). Substituting $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}_0 + \mathbf{u}$ into (4.3) and multiplying through on the left-hand and right-hand sides by $\mathbf{z}'(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})$ gives

$$\mathbf{z}'(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathbf{z}'(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+\mathbf{X}'\mathbb{P}_\xi\mathbf{u}. \tag{4.4}$$

Equation (4.4) forms a TSLS counterpart to Theorem 3.3, from which we know that the asymptotic distribution of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ where $\mathbf{q}' = \mathbf{z}'\mathbf{H}$ will be normal with zero mean and variance $\sigma_u^2\mathbf{z}'(\boldsymbol{\Pi}\,\boldsymbol{\Xi}\,\boldsymbol{\Pi}')^+\mathbf{z}$. To implement this result, in practice, the obvious estimate of $\mathbf{H}$ can be obtained by inserting $n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})$ in place of $(\boldsymbol{\Pi}\,\boldsymbol{\Xi}\,\boldsymbol{\Pi}')$ to give $\widehat{\mathbf{H}} = (\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})$, similarly, the variance can be estimated by replacing $(\boldsymbol{\Pi}\,\boldsymbol{\Xi}\,\boldsymbol{\Pi}')$ by $n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})$ and the residual variance σ_u^2 by the estimate constructed from the TSLS residuals as $\widehat{\sigma}_u^2 = (n-p)^{-1}\sum_{t=1}^n(y_t - \mathbf{x}_t'\widehat{\boldsymbol{\theta}}_n)^2$. The quasi-true parameter $\boldsymbol{\theta}_{0n}$ is given here, as indicated in the theoretical development in Section 3.2, by $\widehat{\mathbf{H}}\boldsymbol{\theta}_0$.

Figure 1 graphs the simulated distribution function and the normal probability plot of the theoretical quantity $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ when $\mathbf{q}' = \mathbf{z}'\mathbf{H}$ where $\mathbf{z}' = (1, \ldots, 1)$ and $se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n) = \sigma_u\sqrt{\mathbf{z}'(\boldsymbol{\Pi}\,\boldsymbol{\Xi}\,\boldsymbol{\Pi}')^+\mathbf{z}}$, and the corresponding empirical quantity $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ when $\mathbf{q}' = \mathbf{z}'\widehat{\mathbf{H}}$ and $se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n) = \widehat{\sigma}_u\sqrt{n\mathbf{z}'(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+\mathbf{z}}$ was used for the estimated standard error. The results depicted in Figure 1 were obtained when $p = 5$ and $k = 9$, using $s_i = 5.0(0.8)^{(i-1)}$, $i = 1, \ldots, p$, and $R = 5{,}000$ replications.

Perusal of Figure 1 indicates that in settings where the model is identified, the level of endogeneity is moderate, the degree of overidentification is quite large, and the instruments are strong, a sample size of $n = 250$ is sufficient for the

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

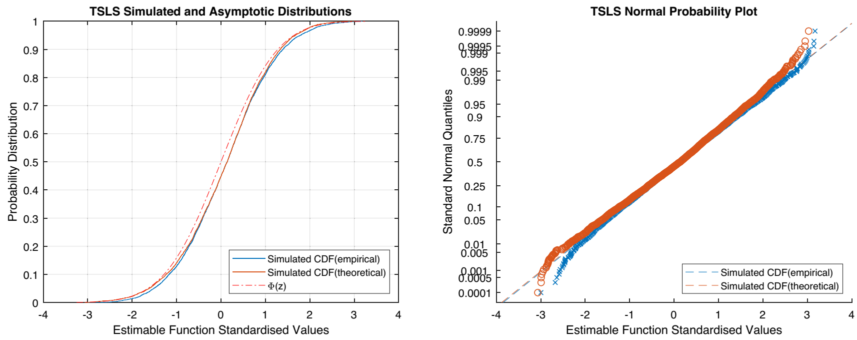


FIGURE 1. Distribution of TSLS estimable function estimates: number of endogenous regressors $p = 5$, number of instruments $k = 9$, sample size $n = 250$. Identified model: identification rank $q = p = 5$.

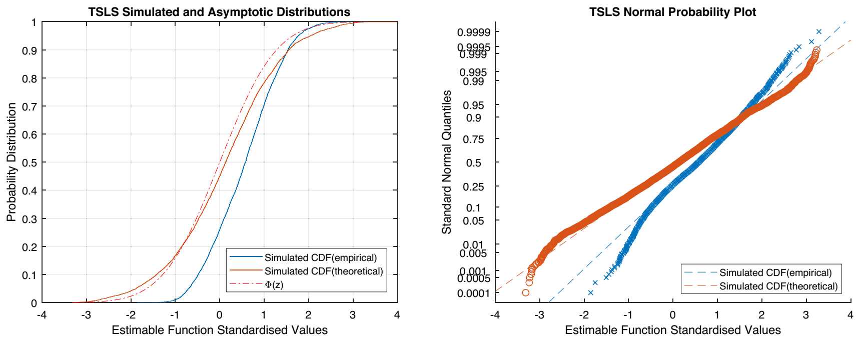
finite-sample distribution of the theoretical and empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ to be virtually identical, with both following the dictates of Theorem 3.3 reasonably well. The distribution and normal probability plots in Figure 1 indicate that in this case the quantiles of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ are well approximated by the quantiles of a standard normal distribution $\zeta_{(1-\alpha)}$ for $0.01 \leq \alpha \leq 0.99$. Such results are not new, of course, but serve here to provide a basis for comparison.

The results depicted in Figure 2 were obtained using the same experimental design as Figure 1 save that $s_i = 5.0(0.8)^{(i-1)}$, $i = 1,\ldots,q$, $s_i = 0$, $i = q+1,\ldots,p$, $q = p-2 = 3$, so the DGP is only partially identified. The detrimental impact of lack of identification upon the sampling distribution of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ can be clearly seen in Figure 2. Figure 2a shows that the quantiles of the $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ theoretical values remain well approximated by $\zeta_{(1-\alpha)}$ for $0.025 \leq \alpha \leq 0.975$, but the empirical $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ quantile values are no longer approximated by those of a standard normal distribution.[8] The ability of estimable functions to ameliorate the adverse effects of identification failure can be gleaned, nevertheless, from an examination of Figure 2b. Figure 2b plots the simulated distribution function and the normal probability plot of the theoretical and empirical quantities $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ when the non-estimable value $\mathbf{q} = \mathbf{z}$ is used in place of the estimable values employed in Figure 2a.

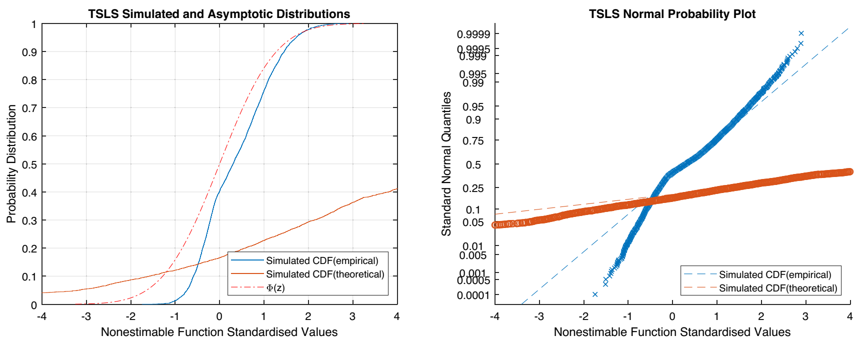
The extreme tails of the theoretical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ observed in Figure 2b accord with the findings of Phillips (1989) where it is shown that the TSLS estimator of an unidentified parameter converges in law to a non-degenerate distribution and that the uncertainty that results from lack of identification persists

[8]Results obtained using a sandwich estimate for the standard error, namely $se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n) = (n\mathbf{z}'(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+ (\mathbf{X}'\mathbb{X}(\mathbb{X}'\mathbb{X})^+(\mathbb{X}'\widehat{\mathbf{D}}_u\mathbb{X})(\mathbb{X}'\mathbb{X})^+\mathbb{X}'\mathbf{X})(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})^+\mathbf{z})^{\frac{1}{2}}$ where $\widehat{\mathbf{D}}_u$ denotes the diagonal matrix $\mathrm{diag}((y_1 - \mathbf{x}_1'\widehat{\boldsymbol{\theta}}_n)^2,\ldots,(y_n - \mathbf{x}_n'\widehat{\boldsymbol{\theta}}_n)^2)$, were qualitatively the same.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press



(a) TSLS estimable function estimates



(b) TSLS non-estimable function “estimates”

FIGURE 2. Distribution of TSLS estimable and non-estimable function values in panels (a) and (b) respectively: number of endogenous regressors $p = 5$, number of instruments $k = 9$, sample size $n = 250$. Partially identified model: identification rank $q = p - 2 = 3$.

in the limit in the form of a covariance matrix normal mixture. In the simulations reported here, 4,046 out of 5,000 replications gave a theoretical value of $|\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)|$ in excess of 3.25, for a standard Cauchy random variable $\Pr(|C| \geq 3.25) = 0.81$. The use of estimable functions has clearly corrected for the extreme behavior of the theoretical values, as seen in a comparison of the properties of the theoretical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ observed in Figure 2a and Figure 2b. Interestingly enough, and by way of contrast, the empirical $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ quantile values do not appear to have altered much between the two scenarios depicted in Figure 2a and Figure 2b the distribution being similarly leptokurtic and skewed in both cases. The differences in behavior between the theoretical and empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ arise in part because $n^{-1}(\mathbf{X}'\mathbb{P}_{\xi}\mathbf{X})$ is frequently deemed to be nonsingular, a consequence of which is that $\widehat{\mathbf{H}} = \mathbf{I}$ and $\mathbf{q}' = \mathbf{z}'\widehat{\mathbf{H}} = \mathbf{z}'$, so the empirical estimable function values collapse to the non-estimable function values. A more detailed explanation of the causes of this phenomenon is presented in the Supplementary Material.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

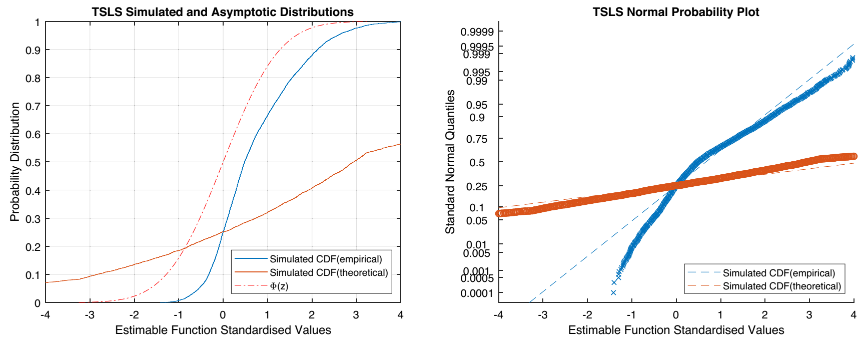


FIGURE 3. Distribution of TSLS estimable function values: number of endogenous regressors $p = 5$, number of instruments $k = 9$, sample size $n = 250$. Identified model: identification rank $q = p = 5$ and identification strength mixed, $p_s = 3$ and $p_w = 2$.

We will now extend the previous illustrations and assess how mixed levels of identification strength will influence the previously observed behavior.

Figure 3 is based on the same experimental design as employed in Figure 1, save that in Figure 3 the DGP is identified and the strength of the instruments is mixed, that is, the identification rank $q = p = 5$ and the instruments contain both strong $p_s = 3$ and weak $p_w = 2$ components. Figure 3 plots the theoretical and empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ when $\mathbf{q} = \mathbf{z}$, which is estimable because the DGP is identified and $\mathbf{H} = \mathbf{I}$.

The only difference between the DGP giving rise to Figure 3 and the DGP that gave rise to Figure 1 is that in Figure 3 the strength of the instruments is mixed, $p_s = 3$ and $p_w = 2$, whereas in Figure 1, the instruments were uniformly strong, $p_s = p = 5$. The upshot of this difference is clearly seen in the contrast between Figure 1, wherein the behavior of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ was in close accord with asymptotic normality, and Figure 3, wherein the theoretical and empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ exhibit significant bias, with a distribution that is positively skewed relative to the normal distribution. Similar effects to those seen in Figure 3 were first documented in Staiger and Stock (1997), where the nomenclature “local-to-zero asymptotics” was coined to describe the weak instrument case $\boldsymbol{\Pi}' = \bar{\mathbf{J}}n^{-\frac{1}{2}}$. For reviews of the weak instrument local-to-zero asymptotics literature, see Stock et al. (2002) and Andrews and Stock (2007). Staiger and Stock (1997) showed that the TSLS estimator is inconsistent and has a nonstandard asymptotic distribution when the instruments are weak, and the behavior seen in Figure 3 indicates that similar properties will prevail for estimable functions in the nonstandard weakly identified case where the instrument strength is mixed, as indicated in Theorem 3.5.

When applied to DGPs containing both partial and weak identification, the TSLS estimator exhibits behavior very similar to that seen in Figure 3. Unlike

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

the situation in Figure 2, however, consideration of estimable functions does not adjust for the effect of identification failure, neither for the empirical values nor the theoretical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$. The estimable function and non-estimable function values are essentially identical for both. This invariance occurs because not only does $\widehat{\mathbf{H}} = \mathbf{I}$ for the empirical values, for the reasons already given, but the added randomness in the Jacobian from the limiting properties of $\bar{\mathbf{J}} + n^{-\frac{1}{2}}\sum_{t=1}^{n}\boldsymbol{\xi}_t\mathbf{v}_t'\mathcal{S}_w$ introduced by the presence of identification weakness (recall Assumption 3.2 and Theorem 3.5) implies that, bar sets of measure zero, $n^{-1}(\mathbf{N}_n\mathbf{X}'\mathbb{P}_{\xi}\mathbf{X}\mathbf{N}_n)$ will be nonsingular for all n sufficiently large and hence that $\mathbf{H}_{0n} = \mathbf{I}$ for the theoretical values. So $\mathbf{z}'(\mathbf{H}_{0n} - \mathbf{I})(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\mathbf{z}'(\widehat{\mathbf{H}} - \mathbf{I})\boldsymbol{\theta}_0 = 0$.

An interesting feature of Figure 3 is that even though the DGP is identified the behavior of both the theoretical and empirical values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ parallel that of the non-estimable values of $\sqrt{n}\mathbf{q}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})/se_{\mathbf{z}}(\widehat{\boldsymbol{\theta}}_n)$ with $\mathbf{q} = \mathbf{z}$ seen in Figure 2b. In Figure 2b, the DGP instrument strength was strong, $p_s = 5$, but the identification rank $q = p - 2 = 3$ and the rank deficiency $p - q = 2$. The DGP underlying Figure 3 corresponds to an identified model where $r\{\boldsymbol{\Pi}\} = p = 5$, $0 < \|\bar{\mathbf{J}}\| < \infty$ and the parameter identification strength is mixed, with $\lambda_c = 0$, $c = 1, \ldots, p_s = 3$, and $\lambda_c = 0.5$, $c = p - p_w + 1, \ldots, p$, $p_w = 2$, so $\mathbf{N}_n = \mathrm{diag}(1, 1, 1, n^{\frac{1}{2}}, n^{\frac{1}{2}})$. The last two columns of $\bar{\mathbf{J}}\mathbf{N}_n^{-1}$ will be arbitrarily close to zero for all n sufficiently large, suggesting that whereas θ_c, $c = 1, \ldots, p_s$ will be strongly identified as n increases, the weakly identified parameters θ_c, $c = p - p_w + 1, \ldots, p$, will appear to be unidentified. In light of the previous observations, it might therefore seem not unreasonable to conjecture that the behavior of statistics computed from data derived from an identified DGP with p_s strongly identified and p_w weakly identified parameters will ultimately be difficult to distinguish from the behavior that would be seen had the statistics been computed from data produced by a DGP with strongly but partially identified parameters with identification rank $q = p_s < p$ and rank deficiency $p - q = p_w$. In what follows, it will be shown that such a conjecture is false.

4.2. Common Conditional Heteroskedastic Features

In their analysis of common features, Engle and Kozicki (1993) suggested that Hansen's $\mathfrak{J}$-test (Hansen, 1982) could provide a unified testing framework. As shown in Dovonon and Renault (2013), however, in the nonlinear example of testing for common CH features identification failure is an integral part of the model and standard GMM asymptotic theory, and the distribution of the $\mathfrak{J}$-test statistic in particular, breaks down. A p-dimensional stochastic process $\mathbf{r}_t = (r_{1t}, \ldots, r_{pt})'$ (a vector of asset returns) is said to have $p - K$ time invariant CH common features, $K < p$, if it has a conditional covariance matrix given by

$$Var(\mathbf{r}_{t+1}|\mathfrak{F}_t) = \boldsymbol{\Lambda}\mathbf{D}_t\boldsymbol{\Lambda}' + \boldsymbol{\Omega},$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

where $Var[\mathbf{r}_{t+1}|\mathfrak{F}_t]$ denotes the conditional variance of $\mathbf{r}_{t+1}$ given all available information at period t, $\mathbf{\Lambda}$ is a $p \times K$ matrix, $\mathbf{D}_t = \text{diag}\{\sigma_{1t}^2, \ldots, \sigma_{Kt}^2\}$, $\mathbf{\Omega}$ is a $p \times p$ positive definite matrix, and $\{\mathfrak{F}_t\}_{t\geq 0}$ is the filtration to which $\{\mathbf{r}_t\}_{t\geq 0}$ and $\{\sigma_{it}^2\}_{t\geq 0}$, $1 \leq i \leq K$, are adapted. In this framework, a CH common feature is by definition any vector $\boldsymbol{\theta}$ in $\mathbb{R}^p$ such that $Var[\boldsymbol{\theta}'\mathbf{r}_{t+1}|\mathfrak{F}_t] = \boldsymbol{\theta}'\mathbf{\Omega}\boldsymbol{\theta}'$, a constant.

Following Dovonon and Renault (2013), suppose that: (i) $r\{\mathbf{\Lambda}\} = K$ and $Var[\mathbf{d}_t]$, where $\mathbf{d}_t$ is the K component vector $(\sigma_{1t}^2, \ldots, \sigma_{Kt}^2)'$, is nonsingular, (ii) $E[\mathbf{r}_{t+1}|\mathfrak{F}_t] = \mathbf{0}$ and (iii) there exists a k-dimensional vector of $\mathfrak{F}_t$-measurable instruments $\boldsymbol{\xi}_t$ such that $Var[\boldsymbol{\xi}_t]$ is nonsingular and $r\{Cov[\boldsymbol{\xi}_t, \mathbf{d}_t]\} = K$. Then the common features are the solutions $\boldsymbol{\theta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \neq \mathbf{0}$, of the moment conditions

$$\boldsymbol{\mu}_t((\boldsymbol{\xi}', \mathbf{r}'), \boldsymbol{\theta}) = E[\boldsymbol{\xi}_t\{(\boldsymbol{\theta}'\mathbf{r}_{t+1})^2 - c(\boldsymbol{\theta})\}] = \mathbf{0},$$

where $c(\boldsymbol{\theta}) = E[(\boldsymbol{\theta}'\mathbf{r}_{t+1})^2]$ (Dovonon and Renault, 2013, Lem. 2.2). If the process $(\boldsymbol{\xi}_t', \mathbf{r}_{t+1}')'$ is stationary and ergodic with $E[\|\boldsymbol{\xi}_t\|^2] < \infty$ and $E[\|\mathbf{r}_{t+1}\|^4] < \infty$, and $(\boldsymbol{\xi}_t', \mathbf{r}_{t+1}')'$ fulfills the conditions needed for $(\boldsymbol{\xi}_t', \text{vec}(\mathbf{r}_{t+1}\mathbf{r}_{t+1}')')'$ to satisfy a central limit theorem, then

$$\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial E\left[\boldsymbol{\mu}_t((\boldsymbol{\xi}', \mathbf{r}'), \boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}'} = \frac{\partial \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \mathbf{0} \tag{4.5}$$

for *any* common feature $\boldsymbol{\theta}$ (Dovonon and Renault, 2013, Prop. 2.1). Thus, we have that the Jacobian will be null and $r\{\bar{\mathbf{\Delta}}_n(\boldsymbol{\theta})\} = q_n = 0$ for any true value $\boldsymbol{\theta}_0$ and any $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0; \delta)$. Consequently, first-order identification failure becomes an intrinsic property of the model.

Under current assumptions, $Var[\boldsymbol{\theta}'\mathbf{r}_{t+1}|\mathfrak{F}_t]$ equals a constant if and only if $\boldsymbol{\theta}$ equals a nontrivial solution to the homogeneous equation system $\mathbf{\Lambda}'\boldsymbol{\theta} = \mathbf{0}$. The null space of $\mathbf{\Lambda}'$ has dimension $p - K$ and in order to achieve identification, exclusion restrictions must be imposed that characterize a subset of the parameter space Θ that contains at most one unknown common feature $\boldsymbol{\theta}_0$, with Θ taken as an arbitrarily large compact subset of $\mathbb{R}^p$. Various parameter restrictions could be imposed, but a common and interpretable normalization condition is the unit cost constraint that $\sum_{i=1}^p \theta_i = 1$, which leads to the interpretation of $\boldsymbol{\theta}'\mathbf{r}_t$ as the return per dollar invested. Setting $\alpha = 1 - \sum_{i=1}^{p-1} \beta_i$ where $\beta_i = \theta_i$, $i = 1, \ldots, p-1$, we can see that the unit cost constraint parallels Lemma 2.2 and the moment conditions and Jacobian can be viewed as functions of $\boldsymbol{\beta} = (\beta_i, \ldots, \beta_{p-1})' \in \mathfrak{B} \subset \mathbb{R}^{p-1}$.

Suppose that the unit cost condition is imposed. Using a sequence of elementary row transformations, the equation system $\mathbf{\Lambda}'\boldsymbol{\theta} = \mathbf{0}$ can be rewritten in equivalent row echelon form

$$\mathbf{H}_1\boldsymbol{\beta}_1 + \mathbf{H}_2\boldsymbol{\beta}_2 + \mathbf{h}\alpha = \mathbf{0}, \tag{4.6}$$

where $\mathbf{H}_1$ is a $K \times K$ upper triangular matrix with unit diagonal elements, $\mathbf{H}_2$ is a $K \times (p - K - 1)$ matrix, $\mathbf{h}$ is a $K \times 1$ column vector, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ is a

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

conformable partition of $\boldsymbol{\beta}$. If $K = p-1$, then $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = (\beta_i, \ldots, \beta_K)'$, the second term $\mathbf{H}_2\boldsymbol{\beta}_2$ is null, and (4.6) can be rewritten as

$$(\mathbf{H}_1 - \mathbf{h}\mathbf{i}_K')\boldsymbol{\beta}_1 + \mathbf{h} = \mathbf{0},$$

where $\mathbf{i}_K = (1, \ldots, 1)'$, the $K \times 1$ sum vector. Since $r\{[(\mathbf{H}_1 - \mathbf{h}\mathbf{i}_K') : -\mathbf{h}]\} = r\{(\mathbf{H}_1 - \mathbf{h}\mathbf{i}_K')\} = K$, it follows that $\boldsymbol{\theta}$ is uniquely identified from $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = (\beta_i, \ldots, \beta_K)'$ and $\alpha = 1 - \sum_{i=1}^{K} \beta_i$. In line with Engle and Kozicki (1993), both Dovonon and Renault (2013) and Lee and Liao (2018) impose the condition that $K = p-1$ in order to achieve identifiability. If $K < p-1$, however, (4.6) implies that any $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha)'$ in which

$$(\mathbf{H}_1 - \mathbf{h}\mathbf{i}_K')\boldsymbol{\beta}_1 = \mathbf{h}(\mathbf{i}_{p-K-1}'\boldsymbol{\beta}_2 - 1) - \mathbf{H}_2\boldsymbol{\beta}_2,$$

where $\boldsymbol{\beta}_2 = (\beta_{K+1}, \ldots, \beta_{p-1})'$ is arbitrary will satisfy both the unit cost constraint and $\boldsymbol{\Lambda}'\boldsymbol{\theta} = \mathbf{0}$. Thus, if the assumption that $K = p-1$ has been imposed erroneously, the unit cost condition will fail to identify $\boldsymbol{\theta}_0$. Moreover, recognizing that $\boldsymbol{\beta}_1$ is a linear and therefore convex function of $\boldsymbol{\beta}_2$, it follows that, when $K < p-1$,

$$\Theta_{0n} = \{\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \alpha)' : (\mathbf{H}_1 - \mathbf{h}\mathbf{i}_K')\boldsymbol{\beta}_1 + \mathbf{h} = (\mathbf{h}\mathbf{i}_{p-K-1}' - \mathbf{H}_2)\boldsymbol{\beta}_2, \alpha = 1 - \sum_{i=1}^{p-1} \beta_i\},$$

a convex subset of Θ. In what follows, we will examine the consequences of supposing that $K = p-1$ when in truth $K \leq p-1$.

In order to render the previous theoretical structure operative, we will suppose that the DGP is given by the CH factor model

$$\mathbf{r}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_{t+1} + \mathbf{u}_{t+1} \qquad \textbf{(4.7)}$$

with constant factor loadings $\boldsymbol{\Lambda}$ and uncorrelated factors $f_{1t}, \ldots, f_{Kt}$ where: $Var[\mathbf{f}_{t+1}|\mathfrak{F}_t] = \mathbf{D}_t$, $Var[\mathbf{u}_{t+1}|\mathfrak{F}_t] = \boldsymbol{\Omega}$, $E[(\mathbf{f}_t', \mathbf{u}_t')'|\mathfrak{F}_t] = \mathbf{0}$, and $E[\mathbf{f}_t\mathbf{u}_t'|\mathfrak{F}_t] = \mathbf{0}$. We will set our empirical moments conditions to

$$\mathbf{m}_n(\boldsymbol{\theta}) = n^{-1}\mathbb{X}'\mathbf{M}_n\mathbf{r}(\boldsymbol{\theta}), \qquad \textbf{(4.8)}$$

where $\mathbb{X}' = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)$, $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\mathbf{i}_n\mathbf{i}_n'$, and $\mathbf{r}(\boldsymbol{\theta}) = ((\boldsymbol{\theta}'\mathbf{r}_2)^2, \ldots, (\boldsymbol{\theta}'\mathbf{r}_{n+1})^2)'$. Under the conditions as previously stated, Assumptions 2.1 and 2.2 hold, and if in addition Assumption 2.3 also holds with $\bar{\boldsymbol{\Omega}}_n$ positive definite, then by Lemma 3.1 the GMM estimate $\widehat{\boldsymbol{\theta}}_n = \arg\min_{\theta \in \Theta} Q_n(\boldsymbol{\theta})$ where

$$Q_n(\boldsymbol{\theta}) = \mathbf{r}(\boldsymbol{\theta})'\mathbf{M}_n\mathbb{X}\mathbf{W}_n\mathbb{X}'\mathbf{M}_n\mathbf{r}(\boldsymbol{\theta})/n^2, \qquad \textbf{(4.9)}$$

will be consistent for $\boldsymbol{\theta}_0$ provided appropriate identification constraints have been imposed that ensure $\bar{Q}_n(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_0$. Otherwise, Theorem 3.2 will obtain. It is shown in Dovonon and Renault (2013,

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

Props. 3.1 and 3.2) that the nullity of the moment Jacobian at any common feature prevents a $\sqrt{n}$ consistent estimate of $\boldsymbol{\theta}_0$ from being built and that when $\boldsymbol{\theta}_0$ is identified $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| = O_p(n^{-1/4})$. Our focus of interest turns therefore to the asymptotic properties of the GMM criterion function and the nonstandard behavior of $Q_n(\boldsymbol{\theta})$ that might be induced by the nonstandard asymptotic behavior of the GMM estimator.

5. CRITERION-BASED INFERENCE

Our purpose here is to examine the effects of lack of identification and identification strength on the properties of $nQ_n(\boldsymbol{\theta})$ and to investigate the consequences for Hansen's $\mathfrak{J}$-test of overidentification and the implementation of other inferential procedures based on $nQ_n(\boldsymbol{\theta})$. Theorem 5.1 shows that in the presence of singularity in the moment long-run variance, or rank deficiency in the Jacobian, or insufficient identification strength, $nQ_n(\widehat{\boldsymbol{\theta}}_n)$, $nQ_n(\boldsymbol{\theta}_{0n})$, and $nQ_n(\boldsymbol{\theta}_{0n}) - nQ_n(\widehat{\boldsymbol{\theta}}_n)$ have nonstandard asymptotic distributions. In Theorem 5.1, the symbol $\prec$ is used to denote stochastic dominance and the symbolic logic notation $\{a \vee b \vee c\}$ is used to denote the disjunction a or b or c.[9]

THEOREM 5.1. *Assume that the conditions of Theorem 3.5 hold and let $\widehat{\boldsymbol{\theta}}_n$ denote a GMM estimator obtained by minimizing $Q_n(\boldsymbol{\theta})$ where $\mathbf{W}_n$ is chosen such that $\|\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n\| \xrightarrow{p} 0$. Set $\bar{\boldsymbol{\Psi}}_{0n} = \bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\mathbf{K}_{0n}(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})^{+}\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}$ and let $\mathfrak{L}^k(\boldsymbol{\zeta},\mathbf{A})$ denote the probability law of the quadratic form $\boldsymbol{\zeta}'\mathbf{A}\boldsymbol{\zeta}$ where $\boldsymbol{\zeta} = (\zeta_1,\ldots,\zeta_k)' \sim \mathcal{N}(\mathbf{0},\mathbf{I}_k)$ and $\mathbf{A} = \mathbf{A}' \geq 0$. Then*

$$nQ_n(\boldsymbol{\theta}_{0n}) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})\,\{\prec \vee \equiv \vee \succ\}\,\chi^2(k_n),$$

where $k_n = r\{\bar{\boldsymbol{\Sigma}}_{0n}\}$, and $nQ_n(\boldsymbol{\theta}_{0n})$ can be decomposed into the sum of two asymptotically independent components,

$$nQ_n(\widehat{\boldsymbol{\theta}}_n) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}(\mathbf{I} - \bar{\boldsymbol{\Psi}}_{0n})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})\{\prec \vee \equiv \vee \succ\}\chi^2(k_n - q_n)$$

and

$$n\{Q_n(\boldsymbol{\theta}_{0n}) - Q_n(\widehat{\boldsymbol{\theta}}_n)\} \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Psi}}_{0n}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}).\{\succ \vee \equiv \vee \prec\}\chi^2(q_n),$$

where $q_n = r\{\mathbf{K}_{0n}\}$.

[9]For any two probability measures P_1 and P_2 belonging to a set of measures $\mathcal{P}$ on a measurable space $(\Omega,\mathbb{S})$, P_2 is said to dominate P_1, denoted $P_1 \prec P_2$, if and only if $P_1(A) > P_2(A)$ for all $A \in \mathbb{S}$. This endows the space $\mathcal{P}$ with a partial ordering $P_1\{\prec \vee \equiv \vee \succ\}P_2$, meaning that P_2 dominates P_1, or P_1 and P_2 are equivalent, or P_1 dominates P_2. On the real line, this ordering assumes a very simple form: $P_1 \prec P_2$ if and only if $1 - F_1(x) < 1 - F_2(x)$ for all $x \in \mathbb{R}$ where for $i = 1,2$, $F_i(x)$ is the cumulative distribution function (CDF) associated with P_i, namely $F_i(x) = P_i((-\infty,x])$.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

There are aspects of Theorem 5.1 that are worthy of comment:

1. Whenever standard GMM assumptions are violated, the large-sample distributions of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$, $nQ_n(\boldsymbol{\theta}_{0n})$, and $nQ_n(\boldsymbol{\theta}_{0n}) - nQ_n(\widehat{\boldsymbol{\theta}}_n)$ will be given by generalized Laguerre series probability laws, and conventional chi-squared approximations will only provide probability bounds, with the direction and magnitude of the stochastic dominance between the two distributions dependent on the nature and severity of the violation.
2. If $\mathbf{W}_n$ is chosen optimally and identification strength is strong, Theorem 5.1 recovers the standard $\chi^2(k-p)$, $\chi^2(k)$ and $\chi^2(p)$ large-sample distributions that hold in the identified full-rank case (see Hall, 2005, Chap. 5, for example).
3. Even if $\mathbf{W}_n$ is chosen optimally and identification strength is strong, singularity in the moment long-run variance and rank deficiency in the Jacobian result in chi-squared distributions with degrees of freedom determined by $k_n = r\{\bar{\boldsymbol{\Sigma}}_{0n}\} < k$ and $q_n = r\{\mathbf{K}_{0n}\} < p$ that provide probability bounds to the underlying generalized Laguerre series probability laws.
4. Theorem 5.1 indicates that identification strength affects the distributional properties of $nQ_n(\boldsymbol{\theta})$ via its impact upon $\bar{\boldsymbol{\Psi}}_{0n}$. When the identification strength is weak, for example, the parameter values may not be readily discernable from the data because $Q_n(\boldsymbol{\theta})$ will not be able to discriminate among different values of $\boldsymbol{\theta}$ if $\partial\mathbf{D}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \approx \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}) \approx \mathbf{0}$, suggesting that tests and confidence sets based on $nQ_n(\boldsymbol{\theta})$ will not be immune from identification deficiency. Nevertheless, even though the precision of any inference based upon $Q_n(\boldsymbol{\theta})$ may be impaired in the presence of identification deficiency, the generalized Laguerre series probability laws for the limiting distributions of $nQ_n(\boldsymbol{\theta}_{0n})$, $nQ_n(\widehat{\boldsymbol{\theta}}_n)$, and $n\{Q_n(\boldsymbol{\theta}_{0n}) - Q_n(\widehat{\boldsymbol{\theta}}_n)\}$ presented in Theorem 5.1 will adapt to the circumstances and yield correct probability calculations.

In the following section, Section 6.1, these aspects of Theorem 5.1 are demonstrated in the context of the classical linear equations model considered in Section 4.1. Before proceeding, however, we present in Theorem 5.2 a modification to the results presented in Theorem 5.1 that allows for the added complexity introduced by the null Jacobian and first-order identification failure of the common CH features model that was outlined in Section 4.2.

THEOREM 5.2. *Assume that the common CH features model satisfies the conditions stated in Section 4.2 and let* $\widehat{\boldsymbol{\theta}}_n$ *denote a GMM estimator obtained by minimizing* $Q_n(\boldsymbol{\theta})$ *where* $\mathbf{W}_n$ *is chosen such that* $\|\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n\| \xrightarrow{p} 0$. *Let* $\mathfrak{L}^k(\boldsymbol{\zeta}, \mathbf{A}, \mathbf{b})$ *denote the probability law of the quadratic form* $(\boldsymbol{\zeta} - \mathbf{b})'\mathbf{A}(\boldsymbol{\zeta} - \mathbf{b})$ *where* $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_k)' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$, $\mathbf{A} = \mathbf{A}' \geq 0$ *and* $\|\mathbf{b}\| \neq 0$. *Then*

$$nQ_n(\boldsymbol{\theta}_{0n}) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}} \bar{\boldsymbol{\Omega}}_n \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})$$

and

$$nQ_n(\widehat{\boldsymbol{\theta}}_n) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}} \bar{\boldsymbol{\Omega}}_n \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}, -\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}} \bar{\boldsymbol{\mu}}_{0n}),$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

where $\bar{\boldsymbol{\mu}}_{0n} = \bar{\boldsymbol{\mu}}_n(\bar{\boldsymbol{\nu}}_{0n})$,

$$\bar{\boldsymbol{\nu}}_{0n} = \arg\min_{\boldsymbol{\nu}_{0n}\in\mathbb{R}^p} \bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n})'\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}(\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n}).$$

The following features of Theorems 5.1 and 5.2 should be added to the previous list.

5. The generalized Laguerre series probability laws in Theorems 5.1 and 5.2 have closed-form expressions for their cumulative distribution functions that can be readily implemented numerically using standard software without recourse to bootstrap or simulation methods. The probability distribution function of $\mathfrak{L}^k(\boldsymbol{\zeta},\mathbf{A},\mathbf{b})$ ($\mathfrak{L}^k(\boldsymbol{\zeta},\mathbf{A}) \equiv \mathfrak{L}^k(\boldsymbol{\zeta},\mathbf{A},\mathbf{0})$) takes the form

$$L^k(\mathbf{A},\mathbf{b};x) = \frac{1}{2^{k/2}\Gamma(k/2)}\int_0^{x/\beta} t^{(k-2)/2}e^{t/2}dt + \left(\frac{x}{2\beta}\right)e^{-x/2\beta}\sum_{j=1}^{\infty}\omega_j\frac{\Gamma(j)}{\Gamma((k+2j)/2)}L_{j-1}^{(k/2)}(x/2\beta), \tag{5.1}$$

where the weights $\omega_j = (2j)^{-1}\sum_{r=0}^{j-1}\tau_{j-r}\omega_r$, $j = 1,2,\ldots$, and

$$\tau_j = \mathrm{tr}\{(\mathbf{I}-\mathbf{A}/\beta)^j\} - (j/\beta)\mathbf{b}'\mathbf{A}(\mathbf{I}-\mathbf{A}/\beta)^{j-1}\mathbf{b}, \quad j = 1,2,\ldots,$$

$\omega_0 = 1$, and $L_j^{(\gamma)}(\cdot)$, $j = 1,2,\ldots$, are the generalized Laguerre polynomials that form an orthogonal family on $0 \le x < \infty$ with respect to the Gamma distribution with shape parameter γ. The series is uniformly convergent for all $x \ge 0$ for any value of β such that 2β exceeds the largest eigenvalue of $\mathbf{A}$, and reduces to a chi-squared form if $\mathbf{A}^2 = \mathbf{A}$ (for details, see Johnson, Kotz, and Balakrishnan, 1995, Chap. 29.5.3).
6. The use of $nQ_n(\boldsymbol{\theta})$ in conjunction with generalized Laguerre series probability laws therefore presents the applied researcher with a straightforward method of constructing asymptotically valid hypothesis tests and confidence sets that will have size and coverage equal to their respective nominal levels and that will be robust to identification failure and deficiency.

These features are noteworthy because although random-functional characterizations of the limiting probability laws of test statistics of interest have been given elsewhere, and it has been pointed out that the distributions will be nonstandard in the presence of identification failure and deficiency (see Staiger and Stock, 1997; Stock and Wright, 2000; Guggenberger and Smith, 2005; Dovonon and Renault, 2013), explicit closed-form expressions for the probability distributions have not previously been presented in the literature as they have here.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

6. ILLUSTRATIONS II

6.1. The Linear Equations Model

In order for the orthogonal decomposition $nQ_n(\boldsymbol{\theta}) = n\{Q_n(\boldsymbol{\theta}) - Q_n(\widehat{\boldsymbol{\theta}}_n)\} + nQ_n(\widehat{\boldsymbol{\theta}}_n)$ in Theorem 5.1 to be valid, the same weighting matrix must be used when evaluating each component, which here amounts to using the TSLS error variance estimate. The consequence of using $\widehat{\sigma}_u^2$ is that $nQ_n(\boldsymbol{\theta})$ equates to a modified Anderson–Rubin statistic, and Hansen's $\mathfrak{J}$-statistic $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ equals the test statistic of Sargan (1958). Figure 4 graphs the simulated empirical distribution function of the modified Anderson–Rubin (mAR) statistic, the Hansen–Sargan (HS) statistic, and the pseudo-likelihood-ratio (pLR) statistic $(n\{Q_n(\boldsymbol{\theta}) - Q_n(\widehat{\boldsymbol{\theta}}_n)\})$, together with two approximating Laguerre series expansions and the (bounding) chi-squared distributions as given in Theorem 5.1. The approximating Laguerre series expansions were evaluated by calculating the probability distribution function of $\mathfrak{L}^k(\boldsymbol{\zeta}, \mathbf{A})$ with the theoretical value of $\mathbf{A}$ replaced by either; (i) the sample analog

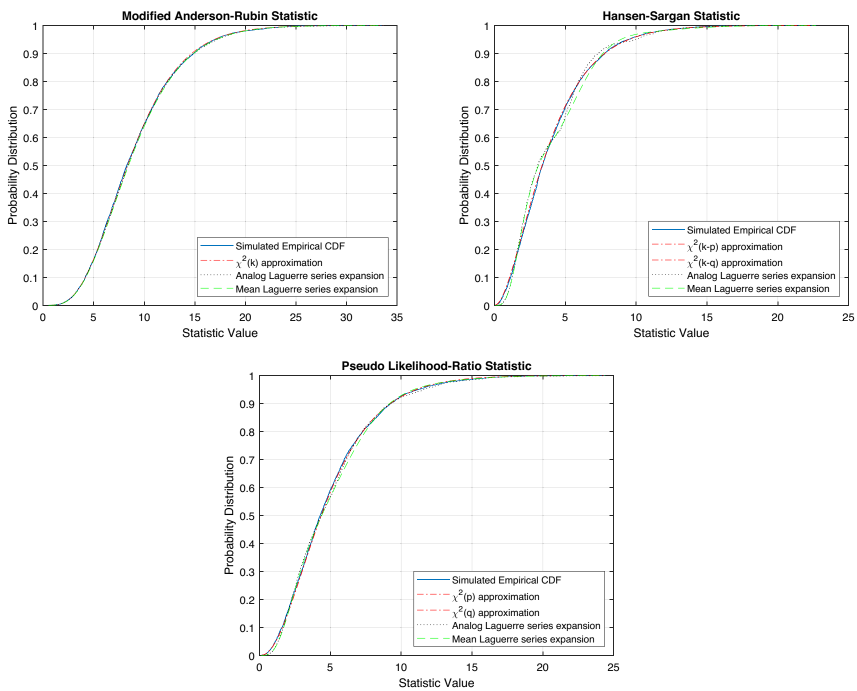


FIGURE 4. Distribution of modified Anderson–Rubin statistic, Hansen–Sargan statistic, and pseudo-likelihood ratio statistic: number of endogenous regressors $p = 5$, number of instruments $k = 9$, sample size $n = 250$. Identified model with identification rank $q = p = 5$ and instrument strength strong $p_s = p = 5$.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

constructed by substituting $n^{-1}(\mathbb{X}'\mathbb{X})\widehat{\sigma}_u^2$ for $\bar{\boldsymbol{\Sigma}}_{0n}$, $\mathbf{W}_n$ for $\bar{\boldsymbol{\Omega}}_n$, $-n^{-1}(\mathbb{X}'\mathbf{X})$ for $\bar{\boldsymbol{\Delta}}_{0n}$, and $n^{-1}(\mathbf{X}'\mathbb{P}_\xi\mathbf{X})\widehat{\sigma}_u^2$ for $\bar{\boldsymbol{\Delta}}_{0n}^{'}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_{0n}$, or (ii) the mean of the sample analogs over the experimental replications. The first of these will be referred to as an analog Laguerre series expansion and denoted $L^k(\mathbf{A}_n)$, and the second as a mean Laguerre series expansion and denoted $L^k(\overline{\mathbf{A}})$.

The results depicted in Figure 4 were obtained using the same experimental design and parameter values as employed in Figure 1, where the level of endogeneity is moderate, the degree of overidentification is quite large, and the instruments are strong, with $p = 5$ and $k = 9$, and sample size $n = 250$. We can see from Figure 4 that in the standard case where the model is identified and the identification strength is strong, $L^k(\mathbf{A}_n)$, $L^k(\overline{\mathbf{A}})$, and the chi-squared distributions are virtually identical for the mAR, HS, and the pLR statistic, and they are all roughly coincident with their simulated empirical distribution functions. Such outcomes are not unexpected and serve as a basis for comparison (as with Figure 1), but they also demonstrate how Theorem 5.1 encompasses the standard results and indicate how generalized Laguerre series probability laws will coalesce with chi-squared distributions when appropriate.

Figure 5 is based on the same experimental design and parameter values as employed in Section 4.1 to explore how data generated by a partially identified model will influence the large-sample properties of statistics of interest. In Figure 5, the DGP is partially identified with strong instruments, the identification rank $q = p - 2 = 3$ and $p_s = p = 5$. It is apparent from Figure 5 that lack of identification has a significant impact on the behavior of $nQ_n(\boldsymbol{\theta})$. We find that whereas both the analog and mean Laguerre series expansions continue to trace out the simulated empirical distribution function of each statistic quite closely, the chi-squared approximations no longer yield accurate guides to the sampling distributions of the statistics. The $\chi^2(q)$ and $\chi^2(p)$ approximations provide upper and lower bounds (respectively) to $L^k(\mathbf{A}_n)$, $L^k(\overline{\mathbf{A}})$ and the simulated empirical distribution function for the pLR statistic, but $L^k(\mathbf{A}_n)$, $L^k(\overline{\mathbf{A}})$ and the simulated empirical distribution function are dominated by the chi-squared approximations for both the mAR statistic and the HS statistic.

In Figure 6, the DGP is identified, but the strength of the instruments is mixed, the identification rank $q = p = 5$, and the instruments contain both strong $p_s = 3$ and weak $p_w = 2$ components. Thus, in the DGP giving rise to Figure 6, the strength of the instruments is mixed, $p_s = 3$ and $p_w = 2$, whereas in the DGP that gave rise to Figure 4, the instruments were uniformly strong, $p_s = p = 5$. The impact of weak instruments can be seen here in the behavior of the mAR statistic and the pLR statistic. In Figure 4, we saw that when the model was identified and the instruments were uniformly strong, $L^k(\mathbf{A}_n)$, $L^k(\overline{\mathbf{A}})$, and the chi-squared approximations were virtually identical, and they were all roughly coincident with their simulated empirical distribution functions. In Figure 6, we now see that when the model is identified but the strength of the instruments is mixed, both the analog and mean Laguerre series expansions continue to match the simulated empirical distribution function of each statistic quite closely, but the chi-squared

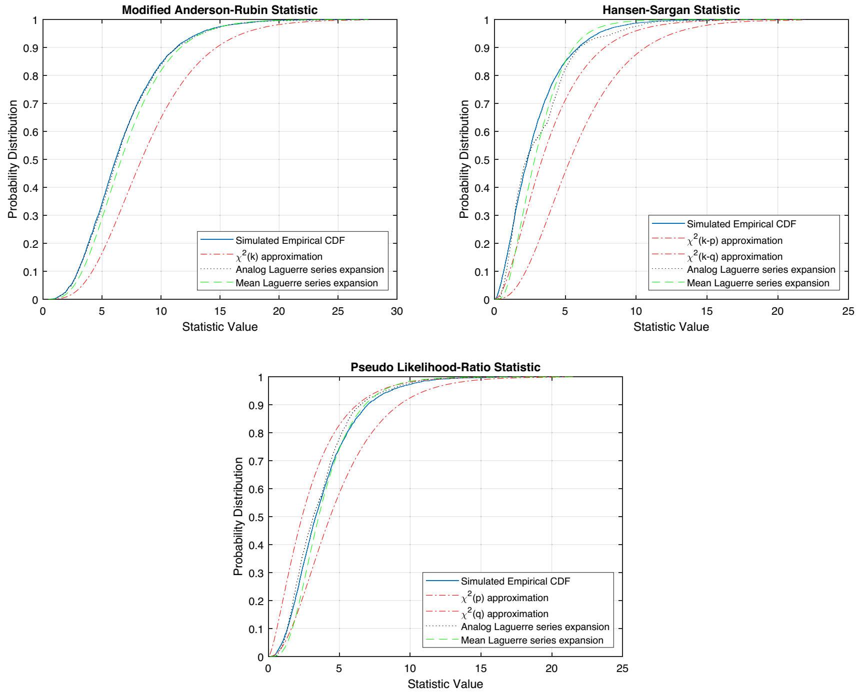


FIGURE 5. Distribution of modified Anderson–Rubin statistic, Hansen–Sargan statistic, and pseudo-likelihood-ratio statistic: number of endogenous regressors $p=5$, number of instruments $k=9$, sample size $n=250$. Partially identified model with identification rank $q=3$ and instrument strength strong $p_s=5$.

approximations no longer yield accurate guides to the sampling distributions of the mAR or the pLR statistic. Interestingly enough, for the HS statistic $L^k(\mathbf{A}_n)$, $L^k(\overline{\mathbf{A}})$, the chi-squared approximation and the simulated empirical distribution are all in close proximity, indicating that the HS statistic appears to exhibit an invariance to mixed levels of instrument strength in identified settings. For the mAR statistic and the pLR statistic, however, the chi-squared approximations are dominated by the Laguerre series expansions and the simulated empirical distribution. The chi-squared approximations are particularly poor in the right-hand tail of the distributions, the third quartile of mAR and pLR corresponding roughly to the 0.875 quantile of their chi-squared approximations, for example. This result is explicative of the over-rejection of test statistics based upon the TSLS estimator found previously by Staiger and Stock (1997) when the instruments are weak.[10]

[10] When the sample size was increased from $n=250$ to $n=5,000$, the distributional properties of the mAR, HS, and pLR statistics changed little despite there being a 20-fold increase in the sample size. This presumably reflects that,

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

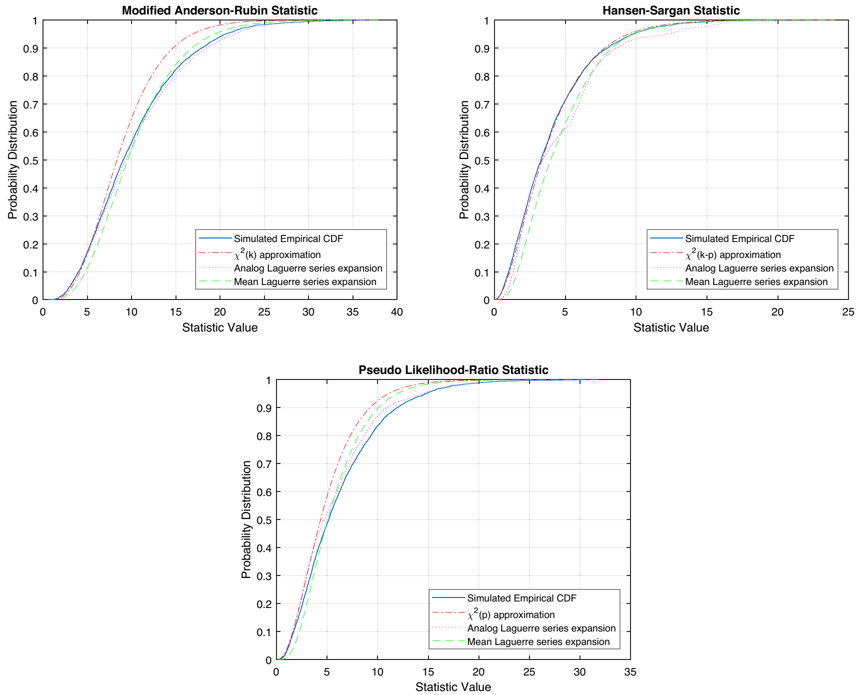


FIGURE 6. Distribution of modified Anderson–Rubin statistic, Hansen–Sargan statistic, and pseudo-likelihood-ratio statistic: number of endogenous regressors $p = 5$, number of instruments $k = 9$, sample size $n = 250$. Identified model with identification rank $q = p = 5$ and instrument strength mixed $p_s = 3$, $p_w = 2$.

A further observation to be made of Figures 5 and 6 is that the only difference in their DGPs is that the level of identification deficiency in Figure 5, $p - q = 2$, matches the extent of the weakness in Figure 6, $p_w = 2$. In both cases, $L^k(\mathbf{A}_n)$ and $L^k(\overline{\mathbf{A}})$ are able to accurately characterize the sampling distribution of all three statistics mAR, HS, and pLR. The chi-squared approximations, on the other hand, dominate the sampling distributions of the mAR and HS statistics in the partially identified case depicted in Figure 5, whereas the stochastic ordering is reversed in the weakly identified case depicted in Figure 6. In Section 4, it was observed that the behavior of estimable functions led to the conjecture that statistics computed from data derived from different DGPs would behave similarly if the degree of identification failure in one DGP matched the level of identification deficiency given by the extent of the weakness in the other. The contrast between the behavior seen in Figures 5 and 6 obviously runs counter to this conjecture.

as indicated in Theorem 3.5, for the DGP used in Figures 3 and 6, the TSLS estimates will be inconsistent and have a nonstandard non-degenerate asymptotic distribution.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

Finally, for the structural equation plus reduced-form model in equations (4.1) and (4.2), the maximal invariants under the group of non-singular linear transformations are the canonical correlations between $\mathbf{x}_t$ and $\boldsymbol{\xi}_t$. For the DGPs examined here, the canonical correlations between $\mathbf{x}_t$ and $\boldsymbol{\xi}_t$ are 0.9761, 0.9153, 0.9079, 0.0, and 0.0 in the partially identified case, and 0.9708, 0.9368, 0.8861, 0.0599, and 0.0227 in the weakly identified case. The maximal invariants map into eigenvalues $\ell_1 \geq \cdots \geq \ell_p$ of the concentration parameter $n\mathbf{T}^{\frac{\pm}{2}}(\mathbf{\Pi}\,\mathbf{\Xi}\,\mathbf{\Pi}')\mathbf{T}^{\frac{\pm}{2}}$ of

$$\{\ell_1,\ldots,\ell_p\} = \{40.7934, 10.8018, 9.8616, 0.0000, 0.0000\}$$

for the partially identified DGP and

$$\{\ell_1,\ldots,\ell_p\} = \{33.2556, 14.8268, 7.7831, 0.0637, 0.0233\}$$

for the weakly identified DGP. Suppose that such values were to be observed in practice and then employed in the widely used test procedure of Cragg and Donald (1993), implemented by assigning a value to either asymptotic bias or size distortion and using the arguments of Stock and Yogo (2005, Sect. 3) to calculate a so-called weak instrument set. Then, following Stock and Yogo (2005), the hypothesis that the instruments lie in the weak instrument set will be rejected if the p-value $Pr\{\chi^2(k,\delta_w) > (n-k)\times \ell_p\}$ is less than α where α is the size of the test and the non-centrality parameter δ_w is determined by the weak instrument set. Whatever value is assigned to δ_w, this p-value would exceed 0.1469 for both the partially identified DGP and the weakly identified DGP. If the likelihood ratio (coefficient of vector alienation) statistic of Poskitt and Skeels (2008) were to be similarly employed, using $Pr\{\chi^2(pk,\delta_s) < -(2n-(p+k+1))\times \log(\prod_{j=1}^{p}(1+\ell_j)^{-\frac{1}{2}})\}$ to approximate its p-value (Poskitt and Skeels, 2008, Thm. 2), then the hypothesis that the instruments are strong will be rejected if the p-value is less than α. For any value of $\delta_s > 0$, both the partially identified DGP and the weakly identified DGP would result in this p-value being approximately zero. In both settings, the models would therefore be deemed to be unidentified or weakly identified at any conventional level of significance. Thus, the practitioner would be made aware of identification issues, although the origin of any identification problems would remain obscure. Nevertheless, the applied researcher can take comfort from the fact that calculations conducted using Laguerre series expansions to approximate the sampling distributions of statistics constructed from the standard GMM criterion function $nQ_n(\boldsymbol{\theta})$ will give accurate probability values in the presence of identification issues, irrespective of the latter's source.

6.2. Common Conditional Heteroskedastic Features

In this subsection, we will examine the finite-sample behavior of the GMM criterion function using the CH common factor model in (4.7). We consider a quadruple vector $\mathbf{r}_t$ with one, two, and three stationary and ergodic generalized

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

Table 1. Parameter values for CH experimental DGPs, $p = 4$

DGP	K_0	Θ_{0n}	$\boldsymbol{\Lambda}_0$	GARCH$(\sigma_0, \phi_0, \rho_0)$
CH_1	1	$\begin{bmatrix} -(1+\beta_2) \\ \beta_2 \\ \beta_3 \\ (2-\beta_3) \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1/2 \\ 1/2 \end{bmatrix}$	(0.2,0.2,0.6)
CH_2	2	$\begin{bmatrix} (\beta_3-6)/8 \\ -(\beta_3+2)/8 \\ \beta_3 \\ (2-\beta_3) \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1/2 & 1/4 \\ 1/2 & 1/8 \end{bmatrix}$	$\begin{pmatrix} 0.2, 0.2, 0.6 \\ 0.2, 0.4, 0.4 \end{pmatrix}$
CH_3	3	$\begin{bmatrix} -1/2 \\ -1/2 \\ -2/7 \\ 16/7 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1/2 & 1/4 & 1 \\ 1/2 & 1/4 & 1/8 \end{bmatrix}$	$\begin{pmatrix} 0.2, 0.2, 0.6 \\ 0.2, 0.4, 0.4 \\ 0.1, 0.1, 0.8 \end{pmatrix}$

autoregressive conditionally heteroskedastic (GARCH) factors. The vector of factors $\mathbf{f}_t = (f_{1t}, \ldots, f_{Kt})')$ are generated from normal GARCH(σ, ϕ, ρ) processes:

$$f_{j(t+1)} = \sigma_{jt}\varepsilon_{j(t+1)} \quad \text{and} \quad \sigma_{jt} = \sigma_j + \phi_j f_{jt}^2 + \rho_j \sigma_{t-1}^2 \quad j = 1, \ldots, K,$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$ is an i.i.d. white noise $\mathcal{N}(\mathbf{0}, \mathbf{I}_k)$ process. The idiosyncratic shock $\mathbf{u}_t$ is an i.i.d. $\mathcal{N}(\mathbf{0}, 0.5\mathbf{I}_4)$ white noise process that is independent of the stochastic disturbance $\boldsymbol{\varepsilon}_t$. The parameter values chosen for the GARCH$(\sigma_j, \phi_j, \rho_j)$ factors, $j = 1, \ldots, K$, the factor loadings considered in the simulation study, and the true and pseudo-true unit cost common features are presented in Table 1. The parameters β_2 and β_3 that appear in the Θ_{0n} column are arbitrary and characterize the unit cost common feature pseudo-true parameter space for $K = p-1 = 3$ when $K_0 = 1$ and $K_0 = 2$. The parameter $\boldsymbol{\mu}$ was set equal to zero throughout. That the conditions stated in Section 4.2 are satisfied for these DGPs is outlined in the Supplementary Material. For each DGP, experiments were conducted for sample sizes $n = 250, 500, 750, 1,250, 2,000, 3,250, 5,250,$ and $8,500$, each with $R = 5,000$ replications. The experimental design and parameter values employed here parallel those employed in Dovonon and Renault (2013) and Lee and Liao (2018) but explore in greater detail a broader range of unidentified models.

For each DGP, $\widehat{\boldsymbol{\theta}}_n = \arg\min_{\theta \in \Theta} Q_n(\boldsymbol{\theta})$ was calculated while imposing the unit cost constraint and assuming that $K = p-1 = 3$. From the norm inequality

$$\|\mathbf{D}_n(\boldsymbol{\theta})'\mathbf{W}_n\mathbf{m}_n(\boldsymbol{\theta})\| \leq \|\mathbf{D}_n(\boldsymbol{\theta})'\mathbf{W}_n^{\frac{1}{2}}\| \cdot \|\mathbf{W}_n^{\frac{1}{2}}\mathbf{m}_n(\boldsymbol{\theta})\|,$$

it follows that the first-order condition for a minimum of $Q_n(\boldsymbol{\theta})$ will be satisfied by any $\boldsymbol{\theta}$ such that $\mathbf{D}_n(\boldsymbol{\theta})'\mathbf{W}_n^{\frac{1}{2}} = \mathbf{0}$. That such a solution is possible in the current setting follows from (4.5) since by Assumption 2.2 (which holds by virtue of the

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

conditions stated in Section 4.2) $\mathbf{D}_n(\boldsymbol{\theta})$ converges uniformly to $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})$ and $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}) = \mathbf{0}$ for any common feature. Substituting $\boldsymbol{\theta} = \mathbf{E}\boldsymbol{\beta} + \mathbf{e}$ where

$$\mathbf{E} = \begin{bmatrix} \mathbf{I}_{p-1} \\ -\mathbf{i}'_{p-1} \end{bmatrix} \quad \text{and} \quad \mathbf{e} = (0, \ldots, 0, 1)'$$

into the moment conditions in (4.8) and differentiating with respect to $\boldsymbol{\beta}$ leads us, after some straightforward if somewhat tedious manipulations, to the following representation for the vectored homogeneous equation $\mathrm{vec}\{\mathbf{D}_n(\boldsymbol{\theta})'\mathbf{W}_n^{\frac{1}{2}}\} = \mathbf{0}$,

$$\frac{1}{n}\sum_{t=1}^{n}(\mathbf{W}_n^{\frac{1}{2}}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}) \otimes \mathbf{E}'\mathbf{r}_t\mathbf{r}_t'(\mathbf{E}\boldsymbol{\beta} + \mathbf{e})) = \mathbf{0}, \tag{6.1}$$

where $\bar{\boldsymbol{\xi}} = n^{-1}\sum_{t=1}^{n}\boldsymbol{\xi}_t$. Let

$$\nabla\mathbf{r}_t = \mathbf{E}'\mathbf{r}_t = ((r_{1t} - r_{pt}), \ldots, (r_{(p-1)t} - r_{pt}))'$$

and set

$$\mathbf{A}_n = \frac{1}{n}\sum_{t=1}^{n}(\mathbf{W}_n^{\frac{1}{2}}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}) \otimes \nabla\mathbf{r}_t\nabla\mathbf{r}_t')$$

and

$$\mathbf{b}_n = \frac{1}{n}\sum_{t=1}^{n}(\mathbf{W}_n^{\frac{1}{2}}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}) \otimes \nabla\mathbf{r}_t r_{pt}).$$

Then, from (6.1), we can conclude that for any given weight matrix $\mathbf{W}_n$, the corresponding criterion minimizing solution is given by

$$\widetilde{\boldsymbol{\theta}}_n = \mathbf{e} - \mathbf{E}\mathbf{A}_n^{+}\mathbf{b}_n.$$

For the purposes of the current simulations, the GMM estimator $\widehat{\boldsymbol{\theta}}_n$ was taken as the iterated two-step estimate $\widetilde{\boldsymbol{\theta}}_n^{(s')}$ where the sequence $\widetilde{\boldsymbol{\theta}}_n^{(s)}$, $s = 1, 2, \ldots$, was initiated with $\widetilde{\boldsymbol{\theta}}_n^{(0)}$ calculated using $\mathbf{W}_n = \mathbf{I}$ and $\widetilde{\boldsymbol{\theta}}_n^{(s)}$ was obtained using $\mathbf{W}_n = \widehat{\widehat{\boldsymbol{\Sigma}}}{}_{0n}^{+}(\widetilde{\boldsymbol{\theta}}_n^{(s-1)})$ where

$$\widehat{\widehat{\boldsymbol{\Sigma}}}_{0n}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}})'\{(\boldsymbol{\theta}'\mathbf{r}_{t+1})^2 - \overline{\mathbf{r}(\boldsymbol{\theta})}\}^2, \quad \overline{\mathbf{r}(\boldsymbol{\theta})} = \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{\theta}'\mathbf{r}_{t+1})^2.$$

The iterations were terminated at $\widetilde{\boldsymbol{\theta}}_n^{(s')}$ where s' is the first index such that $nQ_n(\widetilde{\boldsymbol{\theta}}_n^{(s)}) \leq nQ_n(\widetilde{\boldsymbol{\theta}}_n^{(s-1)})$, $s \leq s'$, and $nQ_n(\widetilde{\boldsymbol{\theta}}_n^{(s')}) < nQ_n(\widetilde{\boldsymbol{\theta}}_n^{(s'+1)})$, $s' \leq 50$.[11]

[11] The average value of s' across the simulations reported below was 3.5193.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

Following Engle and Susmel (1993), the instrument sets employed were $\boldsymbol{\xi}_t = (r_{1t}^2, \ldots, r_{pt}^2)'$ and $\boldsymbol{\xi}_t = (r_{1t}^2, \ldots, r_{pt}^2, r_{1t} \cdot r_{2t}, \ldots, r_{(p-1)t} \cdot r_{pt})'$, namely r_{it}^2, and r_{it}^2 plus $r_{it} \cdot r_{jt}$, for $i = 1, \ldots, p$, $j = i+1, \ldots, p$, $p = 4$, giving $k = 4$ and $k = 10$, respectively. For the AR statistic $nQ_n(\boldsymbol{\theta}_{0n})$, the probability law $\mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}} \bar{\boldsymbol{\Omega}}_n \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})$ and the $\chi^2(k)$ distribution coincide since $\bar{\boldsymbol{\Omega}}_n = \bar{\boldsymbol{\Sigma}}_{0n}^{+}$ and in the current setting $\bar{\boldsymbol{\Sigma}}_{0n}$ is nonsingular. The AR analog and mean Laguerre series expansions correspond to calculating $L^k(\mathbf{A}; x)$ with $\mathbf{A} = \widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{1}{2}}(\boldsymbol{\theta}_{0n}) \widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{+}(\boldsymbol{\theta}_{0n}) \widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{1}{2}}(\boldsymbol{\theta}_{0n})$, which equates to estimating $\bar{\boldsymbol{\Sigma}}_{0n}$ by $\widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{+}(\boldsymbol{\theta}_{0n})$. For $nQ_n(\widehat{\boldsymbol{\theta}}_n)$, the analog Laguerre series expansion $L^k(\mathbf{A}_n, \mathbf{b}_n)$ was evaluated by calculating $L^k(\mathbf{A}, \mathbf{b}; x)$ with $\mathbf{A} = \widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_n) \mathbf{W}_n \widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_n)$ and $\mathbf{b} = -\widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{+}{2}}(\widehat{\boldsymbol{\theta}}_n) \widehat{\mathbf{u}}_{0n} \|\mathbf{m}_n(\widehat{\boldsymbol{\nu}})\|$ where $\widehat{\mathbf{u}}_{0n}$ is the eigenvector corresponding to the smallest eigenvalue of $\widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{+}{2}}(\widehat{\boldsymbol{\theta}}_n)(\widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_n) \mathbf{W}_n \widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_n)) \widehat{\bar{\boldsymbol{\Sigma}}}_{0n}^{\frac{+}{2}}(\widehat{\boldsymbol{\theta}}_n)$ and $\widehat{\boldsymbol{\nu}}$ minimizes $n^{-1} \sum_{t=1}^n (\boldsymbol{\nu}' \mathbf{r}_t)^2$, $\|\boldsymbol{\nu}\| = 1$. These assignments amount to estimating $\bar{\boldsymbol{\Sigma}}_{0n}$ by $\widehat{\bar{\boldsymbol{\Sigma}}}_{0n}(\widehat{\boldsymbol{\theta}}_n)$, $\bar{\boldsymbol{\Omega}}_n$ by $\mathbf{W}_n$, and $\bar{\boldsymbol{\mu}}_{0n}$ by $\widehat{\boldsymbol{\mu}}_{0n} = \widehat{\mathbf{u}}_{0n} \|\mathbf{m}_n(\widehat{\boldsymbol{\nu}})\|$. The mean Laguerre series expansion $L^k(\overline{\mathbf{A}}, \overline{\mathbf{b}})$ was calculated with $\mathbf{A} = \overline{\mathbf{A}}$ and $\mathbf{b} = \overline{\mathbf{b}}$ where $\overline{\mathbf{A}}$ and $\overline{\mathbf{b}}$ denote the average value of the sample analog estimates obtained across the experimental replications.

Figure 7 graphs the distributions of $nQ_n(\boldsymbol{\theta}_{0n})$ and $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ observed with DGP CH_3 for $k = 10$ when $n = 1,250$ and $n = 2,000$. At these sample sizes, the variation in $nQ_n(\boldsymbol{\theta}_{0n})$ from realization to realization generates an observed distribution that is slightly more leptokurtic and skewed than $\chi^2(k)$, but the distributions have virtually coalesced by the time $n = 8,500$, the largest sample size considered. The observed distribution of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ dominates the $\chi^2(k-K) = \chi^2(7)$ distribution and aligns with $L^k(\mathbf{A}_n, \mathbf{b}_n)$ and $L^k(\overline{\mathbf{A}}, \overline{\mathbf{b}})$ alongside $\chi^2(k) = \chi^2(10)$. This accords with the finding of Dovonon and Renault (2013) that Hansen's $\mathfrak{J}$-test (the H$\mathfrak{J}$-test) is over-sized and confirms that the $\chi^2(k)$ distribution will provide asymptotically conservative critical values. The Laguerre series expansions trace out the tail behavior of the distribution of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ more precisely, the non-centrality of the expansions reflecting the over-sized nature of H$\mathfrak{J}$-test.[12] Interestingly enough, as n increases, the Laguerre series expansions maintain the asymptotic first-order stochastic dominance condition

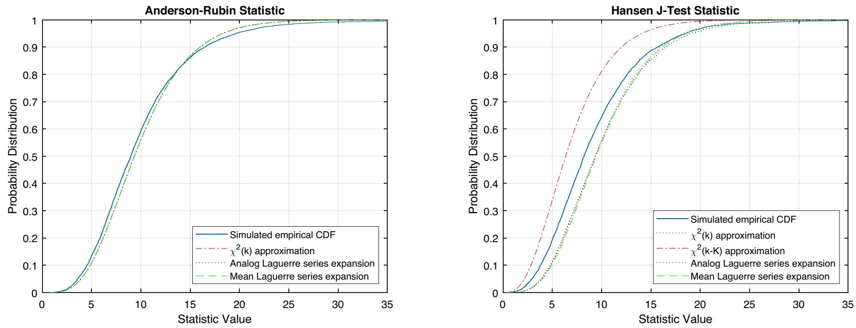
$$\chi^2(k-K) \prec \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}} \bar{\boldsymbol{\Omega}}_n \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}, -\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}} \bar{\boldsymbol{\mu}}_{0n}) \prec \chi^2(k)$$

(Dovonon and Renault, 2013, Thm. 3.2) despite being non-central.

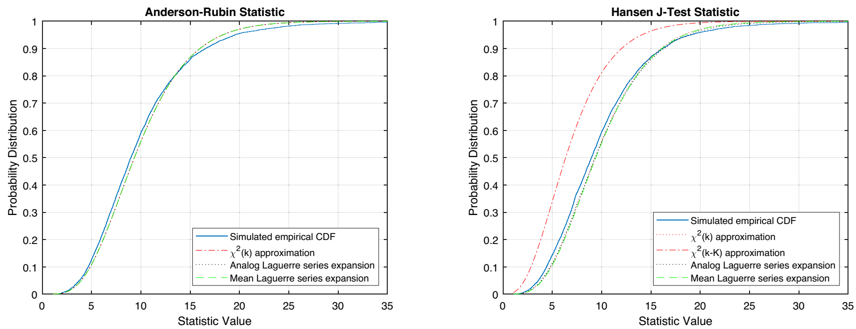
To illustrate the properties of the AR and H$\mathfrak{J}$ statistic in the unidentified case, Figure 8 depicts the outcomes obtained with DGPs CH_2 and CH_1 for $k = 4$ when $n = 1,250$. In Figure 8, the AR statistic was calculated as $nQ_n(\boldsymbol{\theta}_{0n})$ where

[12] In the linear equations IV framework of Section 6.1, the Jacobian is rank deficient, but $\bar{\boldsymbol{\Delta}}_{0n} \neq \mathbf{0}$ and the constancy of $\bar{\boldsymbol{\Delta}}_{0n} = -\boldsymbol{\Xi}\boldsymbol{\Pi}'$ implies that the condition for second-order identification (Dovonon and Renault, 2013, Assum. 5) fails. The parameter $\boldsymbol{\theta}_0$ therefore remains unidentified, resulting in H$\mathfrak{J}$-test being under-sized in the presence of the first- and higher-order identification failure that arises in the linear IV setting.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press



(a) DGP CH_3, $k = 10$ and $n = 1{,}250$.



(b) DGP CH_3, $k = 10$ and $n = 2{,}000$.

FIGURE 7. Distribution of $nQ_n(\boldsymbol{\theta}_{0n})$ and $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ for DGP CH_3 for $k = 10$ and $n = 1,250,\ 2000$.

$\boldsymbol{\theta}_{0n}$ was set equal to the pseudo-true parameter in Θ_{0n} with the smallest norm. Figure 8 also illustrates the behavior of a modified version of the H$\mathfrak{J}$-statistic. An alternative approach to implementing the non-centrality in Theorem 5.2 is to adjust the H$\mathfrak{J}$-statistic by applying a mean correction to the moment conditions. The mean corrected version, or modified H$\mathfrak{J}$-statistic, is given by

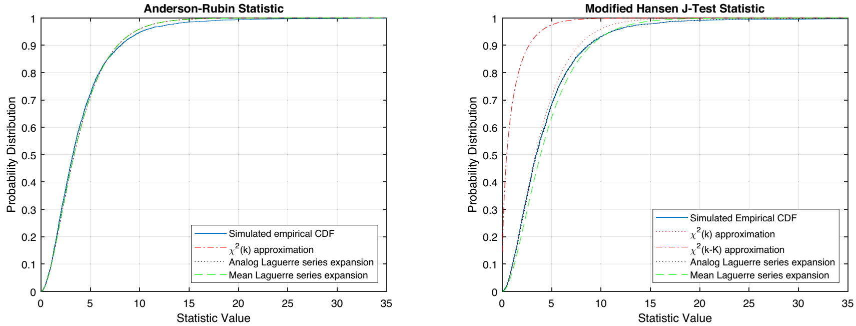
$$nMQ_n(\widehat{\boldsymbol{\theta}}_n) = (\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) - \widehat{\boldsymbol{\mu}}_{0n})'\mathbf{W}_n(\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) - \widehat{\boldsymbol{\mu}}_{0n})\,.$$

The re-centering induced by the mean correction implies that $nMQ_n(\widehat{\boldsymbol{\theta}}_n) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})$ and the cumulative distribution function of $nMQ_n(\widehat{\boldsymbol{\theta}}_n)$ can be approximated by a central Laguerre series expansion $L^k(\mathbf{A};x)$ calculated with $\mathbf{A} = \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}$.

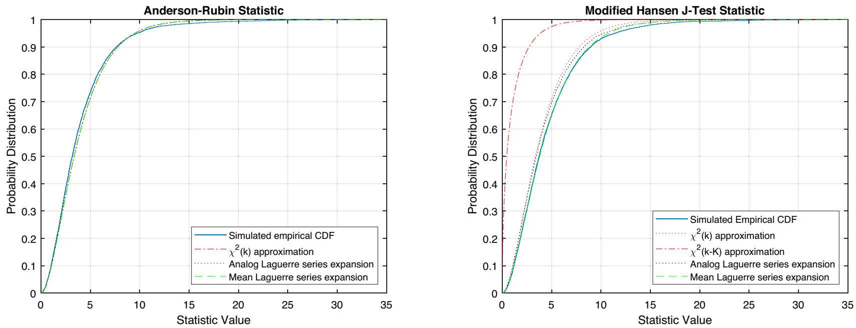
There are some obvious commonalities between the qualitative features seen in Figure 7 and those observed in Figure 8.

Once again, the $\chi^2(k)$ asymptotic approximation provides a relatively accurate characterization of the observed distribution of the AR statistic. The $\chi^2(k)$ percentile values also provide not unreasonable approximations to the distribution

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press



(a) DGPs CH_1 for $k = 4$ and $n = 1{,}250$.



(b) DGPs CH_2 for $k = 4$ and $n = 1{,}250$.

FIGURE 8. Distribution of $nQ_n(\boldsymbol{\theta}_{0n})$ and $nMQ_n(\widehat{\boldsymbol{\theta}}_n)$ for DGPs CH_1, panel (a), and CH_2, panel (b), for $k = 4$ and $n = 1,250$.

of the modified H$\mathfrak{J}$-statistic, but the Laguerre series expansions are again able to more accurately trace out the tail behavior of $nMQ_n(\widehat{\boldsymbol{\theta}}_n)$. That the difference between the Laguerre series expansions and the $\chi^2(k)$ approximation is not inconsequential can be gleaned from the fact that for $nMQ_n(\widehat{\boldsymbol{\theta}}_n)$ the tail probability $1 - L^k(\overline{\mathbf{A}}, \chi^2_{0.95}(k))$ equaled 0.078 when $K_0 = 2$ and 0.0846 when $K_0 = 1$.

Overall, similar features to those seen in Figures 7 and 8 were obtained using DGPs CH_3, CH_2, and CH_1, and various parameter settings. The significant properties that were observed are as follows: (i) The AR statistic can be used as a tool for conducting inference in the common CH features model. Thus, a so-called S-set, namely $\{\boldsymbol{\theta} : nQ_n(\boldsymbol{\theta}) < \chi^2_{(1-\alpha)}(k)\}$ (Stock and Wright, 2000, pp. 1064–1065), will provide an asymptotically valid $(1-\alpha)100\%$ confidence region since $nQ_n(\boldsymbol{\theta})$ will converge in distribution to a $\chi^2(k)$ random variable when evaluated at any parameter point in Θ_{0n}. (ii) The H$\mathfrak{J}$-test will be over-sized if it is employed using a conventional $\chi^2_{(1-\alpha)}(k-K)$ critical value. With moderate to large-sample sizes, a Laguerre series expansion of the sampling distribution of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ or $nMQ_n(\widehat{\boldsymbol{\theta}}_n)$ will provide accurate finite sample p-values.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

6.2.1. *Addendum.* Andrews (1997) has proposed using the H$\mathfrak{J}$-statistic to test for numerical convergence. First, one obtains an initial estimate, $\widehat{\boldsymbol{\theta}}_{n1}$ say. Presupposing that the model is identified, that $\mathbf{W}_n$ has been chosen optimally, and that $r\{\bar{\boldsymbol{\Sigma}}_{0n}\} = k$, one then checks to see if $\widehat{\boldsymbol{\theta}}_{n1}$ satisfies $nQ_n(\widehat{\boldsymbol{\theta}}_{n1}) < \chi^2_{(1-\alpha)}(k-p)$. If $\widehat{\boldsymbol{\theta}}_{n1}$ satisfies the test, one computes $\widehat{\boldsymbol{\theta}}_n$, typically using a local optimization algorithm such as Newton–Raphson starting from the initial estimate $\widehat{\boldsymbol{\theta}}_{n1}$. If $\widehat{\boldsymbol{\theta}}_{n1}$ fails the test, one looks for a new initial estimate, perhaps by considering new starting values for the local optimization algorithm. The previous results indicate that identification failure will compromise such a strategy. Theorem 3.1 and Lemma 3.1 indicate that when the model is unidentified, $Q_n(\boldsymbol{\theta})$ will be "flat" with respect to $\boldsymbol{\theta}$ in neighborhoods of Θ_{0n}, and flatness of $Q_n(\boldsymbol{\theta})$ can of course cause numerical difficulties. The consequences for Andrews' strategy will, however, be contextual.

If a practitioner mistakenly presupposes that a linear equations IV model is identified and compares $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ to the critical value $\chi^2_{(1-\alpha)}(k-p)$, then the probability of rejection may be less than the nominal size if the DGP is in fact unidentified. This will be so if the Laguerre series expansion sampling distribution of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ is dominated by $\chi^2(k-p)$, as seen in Figure 5. This will result in $\Pr(nQ_n(\widehat{\boldsymbol{\theta}}_n) \geq \chi^2_{(1-\alpha)}(k-p)) \ll \alpha$ and indicates that the occurrence of the event $nQ_n(\widehat{\boldsymbol{\theta}}_{n1}) < \chi^2_{(1-\alpha)}(k-p)$ in conjunction with a failure to achieve numerical convergence may not reflect intractability of the global optimization but may be indicative of an unrecognized lack of identification.[13] On the other hand, the H$\mathfrak{J}$-test is over-sized when applied to a common CH features model and $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ can exceed a $\chi^2_{(1-\alpha)}(k-K)$ critical value by a considerable margin even though numerical convergence has occurred, as seen in Figure 7. In both cases, the problem can be rectified by not using the chi-squared critical value. This can be done by employing the Laguerre series expansion sampling distribution of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ to calculate a p-value that will be robust to identification failure.

7. SUMMARY

In this paper, it has been shown that a consequence of working with partially identified GMM models is that although the identified parameter set can be consistently estimated, only certain, so-called, estimable functions of the parameters will possess meaningful limit distributions. When evaluated using theoretical population ensemble averages, estimable functions can ameliorate the adverse effects of identification failure. But due to the fragile relationship between the rank of convergent matrices and their limiting g-inverse, and associated numerical accuracy issues, when estimable functions are calculated using consistent estimates based on sample counterparts to population ensemble averages, they fail to counteract the deleterious consequences of identification failure. Moreover, in

[13] Wright (2003) has proposed an hypothesis test for detecting lack of identification that might be of use here.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

the presence of weak identification, estimable functions will have a nonstandard non-degenerate limiting distribution. These results suggest that although estimable functions are of interest from a theoretical perspective, they do not present the applied researcher with an attractive option and their use in practice may be ill advised.

Fortunately, a new limiting distribution theory for Hansen's $\mathfrak{J}$-test statistic and associated statistics based on the GMM criterion function $Q_n(\boldsymbol{\theta})$ was developed that encompasses standard theory and is also applicable when the standard theory does not apply due to (i) rank deficiency in the moment Jacobian $\bar{\boldsymbol{\Delta}}_{0n}$, (ii) singularity in the moment long-run variance matrix $\bar{\boldsymbol{\Sigma}}_{0n}$, (iii) the weighting matrix $\mathbf{W}_n$ not appropriately converging to $\bar{\boldsymbol{\Sigma}}_{0n}$, or (iv) weak identification strength. The new limiting distributions can be expressed as uniformly convergent series expansions in generalized Laguerre polynomials and can be readily evaluated using standard software without the need to make recourse to bootstrap or simulation methods. Monte Carlo experiments indicated that generalized Laguerre series probability laws will adapt to the prevailing circumstances and provide accurate guides to the sampling distributions of the different GMM statistics based on $Q_n(\boldsymbol{\theta})$. The use of the standard GMM criterion function $Q_n(\boldsymbol{\theta})$ in conjunction with generalized Laguerre series probability laws therefore offers the practitioner a reliable and robust inferential tool.

APPENDICES

A. Proofs

Proof. (Theorem 3.1). It follows from Assumption 2.1, and by the definition of the differential, that given $\epsilon > 0$, there exists a $\delta(\epsilon))$ such that $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0; \delta(\epsilon))$ implies

$$|Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}_0) - 2\mathbf{m}_n(\boldsymbol{\theta}_0)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)| \leq \epsilon\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|, \quad \textbf{(A.1)}$$

wherein we have used the abbreviated notation $\mathbf{m}_n(\boldsymbol{\theta})$ for $\mathbf{m}_n(\mathbf{w}, \boldsymbol{\theta})$, and $\mathbf{D}_n(\boldsymbol{\theta})$ for $\mathbf{D}_n(\mathbf{w}, \boldsymbol{\theta})$. Now, set $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \lambda\mathbf{u}$ where $|\lambda| < \delta(\epsilon)$ and $\mathbf{u}$ denotes a vector of unit length that belongs to the null space of $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)$. Then $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0; \delta(\epsilon))$ and by construction $\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \mathbf{0}$, so $\boldsymbol{\theta}$ belongs to the tangent plane of the level set $\{\boldsymbol{\theta} : \bar{Q}_n(\boldsymbol{\theta}) = \bar{Q}_n(\boldsymbol{\theta}_0)\}$ at $\boldsymbol{\theta}_0$ (see Marsden, 1974, Sects. 6.4 and 6.6, for a definition and discussion of tangent planes).

Substituting $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \lambda\mathbf{u}$ in (A.1), we obtain the inequality

$$\begin{aligned}|Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}_0)| &\leq \epsilon\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + 2|\mathbf{m}_n(\boldsymbol{\theta}_0)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)| \\ &= |\lambda|\left\{\epsilon + 2|\mathbf{m}_n(\boldsymbol{\theta}_0)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_0)\mathbf{u}|\right\} \\ &< \delta(\epsilon)\left\{\epsilon + 2|\mathbf{m}_n(\boldsymbol{\theta}_0)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_0)\mathbf{u}|\right\}. \quad \textbf{(A.2)}\end{aligned}$$

Using Assumptions 2.2 and 2.3 in conjunction with Slutsky's theorem, it is a straightforward exercise to verify that $\|\mathbf{m}_n(\boldsymbol{\theta}_0)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_0) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\| \xrightarrow{p} 0$, and since by construction $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}_0)'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_0)\mathbf{u} = 0$, it follows that

$$\lim_{n\to\infty} \Pr(|\mathbf{m}_n(\boldsymbol{\theta}_0)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_0)\mathbf{u}| > \epsilon) = 0. \quad \textbf{(A.3)}$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

From (A.2) and (A.3), we can therefore deduce that

$$|Q_n(\boldsymbol{\theta}_0) - Q_n(\boldsymbol{\theta})| \leq \delta(\epsilon)\left\{\epsilon + 2o_p(1)\right\},$$

and choosing ϵ such that $3\epsilon\delta(\epsilon) < \eta$ and n_η such that

$$\Pr(|\mathbf{m}_n(\boldsymbol{\theta}_0)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}_0)\mathbf{u}| \leq \epsilon) > 1-\eta$$

for $n > n_\eta$ gives the stated result. □

Proof. (Lemma 3.1). Applying the Cauchy–Schwartz inequality in the form $|\mathbf{x}'\mathbf{A}\mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \|\mathbf{A}\|$ to the expansion

$$\begin{aligned}|Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta})| &= |\mathbf{m}_n(\boldsymbol{\theta})'\mathbf{W}_n\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})'\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})| \\ &\leq |(\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}))'\mathbf{W}_n(\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}))| \\ &\quad + 2|\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})'\mathbf{W}_n(\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta}))| \\ &\quad + |\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})'(\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n)\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})|\end{aligned}$$

gives

$$\begin{aligned}|Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta})| &\leq \|\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\|^2\|\mathbf{W}_n\| \\ &\quad + 2\|\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\|\|\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\|\|'\mathbf{W}_n\| \\ &\quad + \|\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\|^2\|\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n\|.\end{aligned}$$

By assumption, $\|\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\| \xrightarrow{p} 0$ uniformly in $\boldsymbol{\theta}$. Since Θ is compact and $\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, there exists a constant $M < \infty$ such that $\sup_{\theta \in \Theta}\|\bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\| < M$. Furthermore, since $\|W_n\| \leq \|\Omega_n\| + \|W_n - \Omega_n\|$ where $\|\Omega_n\| < \infty$ and $\|\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n\| \xrightarrow{p} 0$, it follows that $\|W_n\|$ is $O_p(1)$.

Hence, we can conclude that

$$\lim_{n\to\infty}\Pr(\sup_{\theta \in \Theta}|Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta})|) = 1,$$

as required. □

Proof. (Theorem 3.3). Multiplying the left-hand side of equation (3.6) by $\mathbf{z}'\bar{\mathbf{H}}_{0n}$ gives

$$\mathbf{z}'\bar{\mathbf{H}}_{0n}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = \sqrt{n}\mathbf{q}_{0n}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}),$$

since $\mathbf{q}_{0n}' = \mathbf{z}'\bar{\mathbf{H}}_{0n}$, and multiplying the first term on the right-hand side of (3.6) by $\mathbf{z}'\bar{\mathbf{H}}_{0n}$ gives

$$-\mathbf{z}'\bar{\mathbf{H}}_{0n}\bar{\boldsymbol{\Gamma}}_{0n}\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) = -\mathbf{z}'\bar{\boldsymbol{\Gamma}}_{0n}\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n})$$

since $\bar{\mathbf{H}}_{0n}\bar{\boldsymbol{\Gamma}}_{0n} = \bar{\boldsymbol{\Gamma}}_{0n}$.

Multiplying equation (3.6) through by $\mathbf{z}'\bar{\mathbf{H}}_{0n}$ and reexpressing the result using the previous equalities now yields the expansion

$$\sqrt{n}\mathbf{q}_{0n}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -\mathbf{z}'\bar{\boldsymbol{\Gamma}}_{0n}\bar{\boldsymbol{\Delta}}_{0n}'\bar{\boldsymbol{\Omega}}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + o_p(1), \quad \textbf{(A.4)}$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

and using the representation from Assumption 3.1

$$\mathbb{G}_n(\boldsymbol{\theta}) = n^{\frac{1}{2}}\{\mathbf{m}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\mu}}_n(\boldsymbol{\theta})\} = \bar{\boldsymbol{\Sigma}}_n^{\frac{1}{2}}(\boldsymbol{\theta})\mathbb{B}_n(\boldsymbol{\theta})$$

evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_{0n}$ gives the desired result. □

Proof. (Theorem 3.4). Expanding $\partial Q_n(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}$ about $\partial Q_n(\boldsymbol{\theta}_{0n})/\partial\boldsymbol{\theta}$ using a first-order Taylor series with Peano's form for the remainder gives us

$$\frac{\partial Q_n(\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}} = \frac{\partial Q_n(\boldsymbol{\theta}_{0n})}{\partial\boldsymbol{\theta}} + \frac{\partial^2 Q_n(\boldsymbol{\theta}_{0n})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{0n}) + o(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{0n}\|),$$

which we rewrite as

$$\mathbf{Q}_{0n}^{(2)}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -2\mathbf{D}_n(\boldsymbol{\theta}_{0n})'\mathbf{W}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + o_p(1)$$

since by definition $\partial Q_n(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta} = \mathbf{0}$ and $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{0n}\| \xrightarrow{p} 0$. Employing Theorem 2.3.1(b) of Rao and Mitra (1971) gives us

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -2\mathbf{Q}_{0n}^{(2+)}\mathbf{D}_n(\boldsymbol{\theta}_{0n})'\mathbf{W}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + (\mathbf{Q}_{0n}^{(2+)}\mathbf{Q}_{0n}^{(2)} - \mathbf{I})\mathbf{z} + o_p(1),$$

where $\mathbf{z}$ is arbitrary.

Multiplying the previous equation through by $\mathbf{z}'\mathbf{Q}_{0n}^{(2+)}\mathbf{Q}_{0n}^{(2)}$ gives us the unique value $\sqrt{n}\mathbf{q}_{0n}'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ on the left-hand side (Rao and Mitra, 1971, Thm. 2.3.1(c)). On the right-hand side, we get $-2\mathbf{z}'\mathbf{Q}_{0n}^{(2+)}\mathbf{D}_n(\boldsymbol{\theta}_{0n})'\mathbf{W}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + o_p(1)$. The stated result now follows since by Assumption 3.1 $n^{\frac{1}{2}}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) = \bar{\boldsymbol{\Sigma}}_n^{\frac{1}{2}}(\boldsymbol{\theta}_{0n})\mathbb{B}_n(\boldsymbol{\theta}_{0n})$. □

Proof. (Lemma 3.2). The proof of this lemma is given in the Supplementary Material. □

Proof. (Theorem 3.5). For ease of reference, recall that the general solution to the first-order condition in (3.3) can be expressed as

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = -\mathbf{G}_n\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + (\mathbf{H}_n - \mathbf{I})\mathbf{z},$$

where $\mathbf{G}_n = (\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}^*))^+$, $\mathbf{H}_n = (\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}^*))^+(\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\mathbf{D}_n(\boldsymbol{\theta}^*))$ and $\mathbf{z}$ is arbitrary. Multiplying through by $\mathbf{M}_n^{-1}\mathbf{P}$, recognizing that $\mathbf{P}\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{P} = \mathbf{I}$ since $\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{P}\mathbf{P} = \mathbf{I}$, gives us

$$\begin{aligned}\sqrt{n}\mathbf{M}_n^{-1}\mathbf{P}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n}) = &- (\mathbf{M}_n^{-1}\mathbf{P}\mathbf{G}_n^+\mathbf{P}\mathbf{M}_n^{-1})\mathbf{M}_n\mathbf{P}'\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)'\mathbf{W}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) \\ &+ ((\mathbf{M}_n^{-1}\mathbf{P}\mathbf{H}_n\mathbf{P}\mathbf{M}_n) - \mathbf{I})\mathbf{M}_n^{-1}\mathbf{P}\mathbf{z}.\end{aligned} \tag{A.5}$$

From Assumption 3.2, we have that $n^{\frac{1}{2}}\{\mathbf{D}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\} = \mathbb{F}_n(\boldsymbol{\theta})$, and from the law of the iterated logarithm for Brownian motion (Wiener) processes, we have

$$\sqrt{\frac{n}{2\log\log n}}\sup_{\theta\in\Theta}\left\|\mathbf{D}_n(\boldsymbol{\theta}) - \bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\right\| \le 1, \text{ almost surely},$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

from which it follows that $n^{\kappa}\sup_{\theta\in\Theta}\left\|\mathbf{D}_n(\boldsymbol{\theta})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\right\| = o_p(1)$ for $0\le\kappa<\frac{1}{2}$. We can therefore deduce that $\sup_{\theta\in\Theta}\|\mathbf{M}_n\mathbf{P}'\{\mathbf{D}_n(\boldsymbol{\theta})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\}'(\mathbf{W}_n-\bar{\boldsymbol{\Omega}}_n)\| \xrightarrow{p} 0$ and that $\sup_{\theta\in\Theta}\|\mathbf{M}_n\mathbf{P}'\{\mathbf{D}_n(\boldsymbol{\theta})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\}'(\mathbf{W}_n-\bar{\boldsymbol{\Omega}}_n)\{\mathbf{D}_n(\boldsymbol{\theta})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta})\}\mathbf{P}\mathbf{M}_n\| \xrightarrow{p} 0$, and the equality $\{\mathbf{D}_n(\boldsymbol{\theta})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\}\mathbf{P}\mathbf{M}_n=\{\mathbf{D}_n(\boldsymbol{\theta})-\mathbf{D}_n(\boldsymbol{\theta}_{0n})\}\mathbf{P}\mathbf{M}_n+\{\mathbf{D}_n(\boldsymbol{\theta}_{0n})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\}\mathbf{P}\mathbf{M}_n$ implies that

$$\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)\mathbf{P}\mathbf{M}_n-\bar{\mathbf{J}}_n(\boldsymbol{\theta}_{0n})\mathbf{P}=\{\mathbf{D}_n(\boldsymbol{\theta}_{0n})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\}\mathbf{P}\mathbf{M}_n+o_p(1)$$

and

$$\mathbf{D}_n(\boldsymbol{\theta}^*)\mathbf{P}\mathbf{M}_n-\bar{\mathbf{J}}_n(\boldsymbol{\theta}_{0n})\mathbf{P}=\{\mathbf{D}_n(\boldsymbol{\theta}_{0n})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\}\mathbf{P}\mathbf{M}_n+o_p(1),$$

where $\|\boldsymbol{\theta}^*-\boldsymbol{\theta}_{0n}\|\le\|\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n}\|=o_p(n^{-\frac{1}{2}})$ by Theorem 3.2. The upshot of this is that $\mathbf{D}_n(\widehat{\boldsymbol{\theta}}_n)\mathbf{P}\mathbf{M}_n$ and $\mathbf{D}_n(\boldsymbol{\theta}^*)\mathbf{P}\mathbf{M}_n$ both equal $\mathbf{K}_{0n}+o_p(1)$ where

$$\mathbf{K}_{0n}=\bar{\mathbf{J}}_n(\boldsymbol{\theta}_{0n})\mathbf{P}+\{\mathbf{D}_n(\boldsymbol{\theta}_{0n})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\}\mathbf{P}\mathbf{M}_n\,,$$

and substituting into (A.5) yields the limiting expression

$$\sqrt{n}\mathbf{M}_n^{-1}\mathbf{P}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})=-\boldsymbol{\Gamma}_{0n}\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n})+(\mathbf{H}_{0n}-\mathbf{I})\mathbf{M}_n^{-1}\mathbf{P}\mathbf{z}+o_p(1), \qquad \textbf{(A.6)}$$

where $\boldsymbol{\Gamma}_{0n}=(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})^+$ and $\mathbf{H}_{0n}=(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})^+(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})$ for the general solution.

From Rao and Mitra (1971, Thm. 2.3.1), it follows that $\mathbf{q}_{0n}'\mathbf{P}\boldsymbol{\theta}$ is estimable whenever $\mathbf{q}_{0n}'=\mathbf{z}'\mathbf{H}_{0n}$, $\mathbf{z}\neq\mathbf{0}$, giving the first part of the theorem. Multiplying (A.6) through by $\mathbf{z}'\mathbf{H}_{0n}=\mathbf{q}_{0n}'$ then yields

$$\begin{aligned}\mathbf{z}'\mathbf{H}_{0n}\sqrt{n}\mathbf{M}_n^{-1}\mathbf{P}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n}) =&\mathbf{q}_{0n}'\sqrt{n}\mathbf{M}_n^{-1}\mathbf{P}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})\\ =&-\mathbf{z}'(\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\mathbf{K}_{0n})^+\mathbf{K}_{0n}'\bar{\boldsymbol{\Omega}}_n\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n})+o_p(1)\,,\end{aligned}$$

from which the second statement in the theorem follows since $n^{\kappa}\|\mathbf{D}_n(\boldsymbol{\theta}_{0n})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\|=o_p(1)$ for $0\le\kappa<\frac{1}{2}$ and $n^{\frac{1}{2}}\mathrm{vec}\{\mathbf{D}_n(\boldsymbol{\theta}_{0n})-\bar{\boldsymbol{\Delta}}_n(\boldsymbol{\theta}_{0n})\}=\mathrm{vec}\{\mathbb{F}_n(\boldsymbol{\theta}_{0n})\}=\bar{\boldsymbol{\Lambda}}_{0n}^{\frac{1}{2}}\mathbb{W}_n(\boldsymbol{\theta}_{0n})$. □

In order to establish Theorem 5.1 appeal will be made to the following result, the proof of which is given in the Supplementary Material.

Lemma A.1. *Assume that* $\mathbf{x}_n\in\mathbb{R}^k$ *converges to* $\mathbf{x}$ *as* $n\to\infty$, *and that* $\|\mathbf{A}_n-\mathbf{A}\|\to 0$ *where* $\mathbf{A}_n=\mathbf{A}_n'$ *and* $\mathbf{A}=\mathbf{A}'$ *are positive semi-definite* $k\times k$ *matrices and* $0\le\|\mathbf{A}\|<\infty$. *Then* $|\mathbf{x}_n\mathbf{A}_n\mathbf{x}_n-\mathbf{x}\mathbf{A}\mathbf{x}|\to 0$ *as* $n\to\infty$. *Furthermore,* $\liminf_{n\to\infty}\mathbf{x}_n\mathbf{A}_n^+\mathbf{x}_n\ge\mathbf{x}\mathbf{A}^+\mathbf{x}$, *and* $|\mathbf{x}_n\mathbf{A}_n^+\mathbf{x}_n-\mathbf{x}\mathbf{A}^+\mathbf{x}|\to 0$ *as* $n\to\infty$ *if and only if* $r\{\mathbf{A}_n\}=r\{\mathbf{A}\}$ *for all n sufficiently large.*

Proof. (Theorem 5.1). From Assumption 5, it follows that $\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n})=\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n}$ where $\boldsymbol{\zeta}_{0n}\Rightarrow\boldsymbol{\zeta}\sim N(\mathbf{0},\mathbf{I})$. Using Lemma A.1 together with Skorokhod's representation theorem, we are therefore led to the conclusion that

$$\begin{aligned} nQ_n(\boldsymbol{\theta}_{0n}) &= n\mathbf{m}_n(\boldsymbol{\theta}_{0n})'\mathbf{W}_n\mathbf{m}_n(\boldsymbol{\theta}_{0n}) \\ &= \boldsymbol{\zeta}'_{0n}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\mathbf{W}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} \\ &= \boldsymbol{\zeta}'_{0n}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} + o_p(1), \end{aligned}$$

where $\|\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n\| \to 0$. A quadratic function of $\boldsymbol{\zeta}$ is obviously Fréchet differentiable, and therefore Lipschitzian and hence uniformly continuous, and it follows upon application of the Continuous Mapping Theorem (CMT) in conjunction with Lemma A.1 that $nQ_n(\boldsymbol{\theta}_{0n}) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})$. If $\bar{\boldsymbol{\Omega}}_n = \bar{\boldsymbol{\Sigma}}_{0n}^+$, then $nQ_n(\boldsymbol{\theta}_{0n}) \Rightarrow \chi^2(k_n)$ where $k_n = tr\{\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^+\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\} = tr\{\bar{\boldsymbol{\Sigma}}_{0n}^+\bar{\boldsymbol{\Sigma}}_{0n}\} = r\{\bar{\boldsymbol{\Sigma}}_{0n}\} \leq k$. This follows from Theorem 9.2.1 of Rao and Mitra (1971) since $\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^+\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}$ is a symmetric and idempotent matrix. The stated stochastic dominance follows since $\bar{\boldsymbol{\Omega}}_n\{< \vee = \vee >\}\bar{\boldsymbol{\Sigma}}_{0n}^+$ implies that $\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\{< \vee = \vee >\}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^+\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}$.

Solving (A.6) for $(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0n})$ and substituting back into the expansion of $\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n)$ in (3.2) gives us

$$\begin{aligned} \mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) &= \mathbf{m}_n(\boldsymbol{\theta}_{0n}) - \mathbf{D}_n(\boldsymbol{\theta}^*)\mathbf{P}\mathbf{M}_n\{\boldsymbol{\Gamma}_{0n}\mathbf{K}'_{0n}\bar{\boldsymbol{\Omega}}_n\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + (\mathbf{H}_{0n} - \mathbf{I})\mathbf{M}_n^{-1}\mathbf{P}\mathbf{z}/\sqrt{n} + o_p(1)\} \\ &= \{\mathbf{I} - \mathbf{K}_{0n}\boldsymbol{\Gamma}_{0n}\mathbf{K}'_{0n}\bar{\boldsymbol{\Omega}}_n\}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \mathbf{K}_{0n}(\mathbf{I} - \mathbf{H}_{0n})\mathbf{M}_n^{-1}\mathbf{P}\mathbf{z}/\sqrt{n} + o_p(1), \qquad \textbf{(A.7)} \end{aligned}$$

where $\mathbf{z}$ is arbitrary. We can therefore infer that

$$\begin{aligned} \sqrt{n}\mathbf{W}_n^{\frac{1}{2}}\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) &= (\mathbf{I} - \bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\mathbf{K}_{0n}\boldsymbol{\Gamma}_{0n}\mathbf{K}'_{0n}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} + \bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\mathbf{K}_{0n}(\mathbf{I} - \mathbf{H}_{0n})\mathbf{M}_n^{-1}\mathbf{P}\mathbf{z} + o_p(1) \\ &= (\mathbf{I} - \mathbf{A}_{0n}(\mathbf{A}'_{0n}\mathbf{A}_{0n})^+\mathbf{A}'_{0n})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} + \mathbf{A}_{0n}(\mathbf{I} - \mathbf{H}_{0n})\mathbf{M}_n^{-1}\mathbf{P}\mathbf{z} + o_p(1), \end{aligned}$$

where $\mathbf{H}_{0n} = (\mathbf{A}'_{0n}\mathbf{A}_{0n})^+(\mathbf{A}'_{0n}\mathbf{A}_{0n})$ and $\mathbf{A}'_{0n} = \mathbf{K}'_{0n}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}$. Since by Lemma 2.2.6(b) of Rao and Mitra (1971) $\mathbf{A}_{0n}(\mathbf{I} - \bar{\mathbf{H}}_{0n}) = \mathbf{0}$, it follows that

$$nQ_n(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{\zeta}'_{0n}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}(\mathbf{I} - \mathbf{A}_{0n}(\mathbf{A}'_{0n}\mathbf{A}_{0n})^+\mathbf{A}'_{0n})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} + o_p(1),$$

and Lemma 2.2.6(d) of Rao and Mitra (1971) implies that $\mathbf{A}_{0n}(\mathbf{A}'_{0n}\mathbf{A}_{0n})^+\mathbf{A}'_{0n}$ is invariant to the choice of g-inverse. Application of the CMT in conjunction with Lemma A.1 now indicates that

$$\begin{aligned} nQ_n(\widehat{\boldsymbol{\theta}}_n) &= \boldsymbol{\zeta}'_{0n}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}(\mathbf{I} - \mathbf{A}_{0n}(\mathbf{A}'_{0n}\mathbf{A}_{0n})^+\mathbf{A}'_{0n})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} + o_p(1) \\ &\Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}(\mathbf{I} - \bar{\boldsymbol{\Psi}}_{0n})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}). \end{aligned}$$

Furthermore, straightforward algebraic manipulation shows that

$$\nabla_{0n} = \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}(\mathbf{I} - \mathbf{A}_{0n}(\mathbf{A}'_{0n}\mathbf{A}_{0n})^+\mathbf{A}'_{0n})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}$$

is a symmetric nonnegative definite matrix and that ∇_{0n} is idempotent if and only if $\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}\bar{\boldsymbol{\Omega}}_n = \bar{\boldsymbol{\Omega}}_n$. Theorem 9.2.1 of Rao and Mitra (1971) therefore implies that $nQ_n(\widehat{\boldsymbol{\theta}}_n) \Rightarrow \chi^2(k_n - q_n)$ when $\bar{\boldsymbol{\Omega}}_n = \bar{\boldsymbol{\Sigma}}_{0n}^+$ where $k_n = tr\{\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^+\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\} = tr\{\bar{\boldsymbol{\Sigma}}_{0n}^+\bar{\boldsymbol{\Sigma}}_{0n}\} = r\{\bar{\boldsymbol{\Sigma}}_{0n}\}$ and

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

$$q_n = tr\{\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}(\bar{\boldsymbol{\Sigma}}_{0n}^{+}\mathbf{K}_{0n}(\mathbf{K}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\mathbf{K}_{0n})^{+}\mathbf{K}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+})\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\}$$
$$= tr\{(\mathbf{K}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\mathbf{K}_{0n})^{+}(\mathbf{K}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\mathbf{K}_{0n})\}$$
$$= r\{\mathbf{K}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{+}\mathbf{K}_{0n}\}.$$

That $\mathfrak{L}^k(\boldsymbol{\zeta},\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}(\mathbf{I}-\bar{\boldsymbol{\Psi}}_{0n})\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})\{\succ\vee\equiv\vee\prec\}\chi^2(k_n-q_n)$ follows.

Via an application of a similar logic, we are also led to the conclusion that

$$n\{Q_n(\boldsymbol{\theta}_{0n})-Q_n(\widehat{\boldsymbol{\theta}}_n)\} = \boldsymbol{\zeta}_{0n}'\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}(\mathbf{A}_{0n}(\mathbf{A}_{0n}'\mathbf{A}_{0n})^{+}\mathbf{A}_{0n}')\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n}+o_p(1)$$
$$\Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta},\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Psi}}_{0n}\bar{\boldsymbol{\Omega}}_n^{\frac{1}{2}}\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})\{\succ\vee\equiv\vee\prec\}\chi^2(q_n).$$

Since $(\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}-\nabla_{0n})\nabla_{0n}=\mathbf{0}$, it follows from Theorem 9.4.1 of Rao and Mitra (1971) that $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ and $n\{Q_n(\boldsymbol{\theta}_{0n})-Q_n(\widehat{\boldsymbol{\theta}}_n)\}$ are asymptotically independent. □

Proof. (Theorem 5.2). The first part of Theorem 5.2 concerning the distribution of $nQ_n(\boldsymbol{\theta}_{0n})$ parallels that of Theorem 5.1 and can be verified in an identical manner. To establish the second part of Theorem 5.2, we will adapt the argument used in Dovonon and Renault (2013, p. 2576) and consider the second-order Taylor expansion

$$\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \sum_{i=1}^{p}\frac{\partial\mathbf{m}_n(\boldsymbol{\theta}_{0n})}{\partial\theta_i}(\widehat{\theta}_n-\theta_{0n})_i + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}(\widehat{\theta}_n-\theta_{0n})_i'\frac{\partial^2\mathbf{m}_n(\boldsymbol{\theta}_{0n})}{\partial\theta_i\partial\theta_j}(\widehat{\theta}_n-\theta_{0n})_j$$
$$= \mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \mathbf{D}_n(\boldsymbol{\theta}_{0n})(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n}) + \frac{1}{2}(\mathbf{I}_k\otimes(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})')\frac{\partial\mathrm{vec}(\mathbf{D}_n(\boldsymbol{\theta})')}{\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n}). \quad \textbf{(A.8)}$$

Since by Corollary 3.1 and Proposition 3.1 of Dovonon and Renault (2013) the second term in (A.8) $\mathbf{D}_n(\boldsymbol{\theta}_{0n})(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})=O_p(n^{-3/4})$, and

$$\frac{\partial\mathrm{vec}(\mathbf{D}_n(\boldsymbol{\theta})')}{\partial\boldsymbol{\theta}'} = \frac{2}{n}\sum_{t=1}^{n}((\boldsymbol{\xi}_t-\bar{\boldsymbol{\xi}})\otimes\mathbf{r}_t\mathbf{r}_t'),$$

it follows that

$$\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \frac{1}{n}\sum_{t=1}^{n}((\boldsymbol{\xi}_t-\bar{\boldsymbol{\xi}})\otimes(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})'\mathbf{r}_t\mathbf{r}_t'(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})) + o_p(n^{-\frac{1}{2}})$$
$$= \mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \frac{1}{n}\sum_{t=1}^{n}((\boldsymbol{\xi}_t-\bar{\boldsymbol{\xi}})\cdot((\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})'\mathbf{r}_t)^2 + o_p(n^{-\frac{1}{2}}). \quad \textbf{(A.9)}$$

Let $\widehat{\boldsymbol{\nu}}_{0n}=n^{\frac{1}{4}}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{0n})$ and note from Assumption 2.2 and (A.9) that

$$\sqrt{n}\mathbf{m}_n(\widehat{\boldsymbol{\theta}}_n) = \sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \bar{\boldsymbol{\mu}}_n(\widehat{\boldsymbol{\nu}}_{0n}) + o_p(1)$$

and hence that

$$nQ_n(\widehat{\boldsymbol{\theta}}_n) = (\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n})+\bar{\boldsymbol{\mu}}_n(\widehat{\boldsymbol{\nu}}_{0n}))'\mathbf{W}_n(\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n})+\bar{\boldsymbol{\mu}}_n(\widehat{\boldsymbol{\nu}}_{0n})) + o_p(1).$$

Now, set

$$n\mathcal{Q}_n(\boldsymbol{\nu}_{0n}) = n\mathbf{m}_n(\boldsymbol{\theta}_{0n}+n^{-\frac{1}{4}}\boldsymbol{\nu}_{0n})'\mathbf{W}_n\mathbf{m}_n(\boldsymbol{\theta}_{0n}+n^{-\frac{1}{4}}\boldsymbol{\nu}_{0n}),$$

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

where $\boldsymbol{\nu}_{0n} = n^{\frac{1}{4}}(\boldsymbol{\theta} - \boldsymbol{\theta}_{0n})$, $\boldsymbol{\theta} \in \Theta$. Using a second-order Taylor expansion as in (A.8) and setting $\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) = \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n}$ where $\boldsymbol{\zeta}_{0n} \Rightarrow \boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$, we have

$$\begin{aligned}
n\mathcal{Q}_n(\boldsymbol{\nu}_{0n}) &= (\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n}))'\mathbf{W}_n(\sqrt{n}\mathbf{m}_n(\boldsymbol{\theta}_{0n}) + \bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n})) + o_p(1) \\
&= (\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} + \bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n}))'\mathbf{W}_n(\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\zeta}_{0n} + \bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n})) + o_p(1) \\
&= (\boldsymbol{\zeta}_{0n} + \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n}))'\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\boldsymbol{\Omega}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}(\boldsymbol{\zeta}_{0n} + \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n})) + o_p(1),
\end{aligned}$$

since by assumption $\|\mathbf{W}_n - \bar{\boldsymbol{\Omega}}_n\| \xrightarrow{p} 0$ (Assumption 2.2). This implies that for any $\boldsymbol{\nu}_{0n} \in \mathbb{R}^p$

$$n\mathcal{Q}_n(\boldsymbol{\nu}_{0n}) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}, -\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n})).$$

By definition,

$$nQ_n(\widehat{\boldsymbol{\theta}}_n) = n\mathcal{Q}_n(\widehat{\boldsymbol{\nu}}_{0n}) = \min_{\boldsymbol{\nu}_{0n} \in \mathbb{R}^p} n\mathcal{Q}_n(\boldsymbol{\nu}_{0n}).$$

It follows that the distribution of $nQ_n(\widehat{\boldsymbol{\theta}}_n)$ is either equivalent to or is dominated by the distribution of $n\mathcal{Q}_n(\boldsymbol{\nu}_{0n})$ for all $\boldsymbol{\nu}_{0n} \neq \widehat{\boldsymbol{\nu}}_{0n}$. From (5.1), the distribution function $L^k(\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}, -\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n}); x)$ is a monotonically decreasing function of the non-centrality parameter

$$\kappa(\boldsymbol{\nu}_{0n}) = \bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n})'\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}(\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}})\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\mu}}_n(\boldsymbol{\nu}_{0n}),$$

for any $x \geq 0$, and it follows that $nQ_n(\widehat{\boldsymbol{\theta}}_n) \Rightarrow \mathfrak{L}^k(\boldsymbol{\zeta}, \bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}\bar{\boldsymbol{\Omega}}_n\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{1}{2}}, -\bar{\boldsymbol{\Sigma}}_{0n}^{\frac{+}{2}}\bar{\boldsymbol{\mu}}_{0n})$ where $\bar{\boldsymbol{\mu}}_{0n} = \bar{\boldsymbol{\mu}}_n(\bar{\boldsymbol{\nu}}_{0n})$, $\bar{\boldsymbol{\nu}}_{0n} = \arg\min \kappa(\boldsymbol{\nu}_{0n})$. □

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit: https://doi.org.10.1017/S0266466623000221

REFERENCES

Ambrisio, L. & P. Tilli (2004) *Topics on Analysis in Metric Spaces*. Oxford University Press.

Andrews, D.K.W. (1997) A stopping rule for the computation of generalized method of moments estimators. *Econometrica* 65(4), 913–931.

Andrews, D.W.K. & P. Guggenberger (2017) Asymptotic size of Kleibergen's LM and conditional LR tests for moment condition models. *Econometric Theory* 33, 1046–1080.

Andrews, D.W.K. & P. Guggenberger (2019) Identification- and singularity-robust inference for moment condition models. *Quantitative Economics* 10, 1703–1746.

Andrews, D.W.K. & J.H. Stock (2007) Inference with weak instruments. In R. Blundell, W.K. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. III*, pp. 122–173. Cambridge University Press.

Antoine, B. & E. Renault (2009) Efficient GMM with nearly-weak instruments. *Journal of Econometrics* 12, S135–S171.

Bodnar, T., S. Mazur, & K. Podgorski (2016) Singular inverse Wishart distribution and its application to portfolio theory. *Journal of Multivariate Analysis* 143, 314–326.

Caner, M. (2010) Testing, estimation in GMM and CUE with nearly weak identification. *Econometric Reviews* 29, 330–363.

Chao, J.C. & N.R. Swanson (2005) Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.

Chernozhukov, V., H. Hong, & E. Tamer (2007) Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica* 75(5), 1243–1284.

Choi, I. & P.C.B. Phillips (1992) Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified models. *Journal of Econometrics* 51(1–2), 113–150.

Cragg, J.G. & S.G. Donald (1993) Testing Identifiability and specification in instrument variable models. *Econometric Theory* 9, 222–240.

Dovonon, P. & E. Renault (2013) Testing for common conditionally heteroskedastic factors. *Econometrica* 81, 2561–2586.

Engle, R.F. & S. Kozicki (1993) Testing for common features. *Journal of Buisiness and Economic Statistics* 11, 369–395.

Engle, R.F. & R. Susmel (1993) Common volatility in international equity markets. *Journal of Business and Economic Statistics* 11, 369–395.

Guggenberger, P. & R.J. Smith (2005) Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory* 21, 667–709.

Hahn, J. & J. Hausman (2003) Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review* 93(2), 118–125.

Hall, A.R. (2005) *Generalized Method of Moments*. Oxford University Press.

Han, C. & P.C.B. Phillips (2006) GMM with many moment conditions. *Econometrica* 74, 147–192.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P., J. Heaton, & A. Yaron (1996) Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14, 262–280.

Johnson, N.L., S. Kotz, & N. Balakrishnan (1995) *Continuous Univariate Distributions*, vol. 2, 2nd ed. Wiley.

Kleibergen, F. (2005) Testing parameters in GMM without assuming that they are identified. *Econometrica* 73(4), 1103–1123.

Lee, J. & Z. Liao (2018) On standard inference for GMM with local identification failure of known forms. *Econometric Theory* 34, 790–814.

Marsden, J.E. (1974) *Elementary Classical Analysis*. W.H. Freeman.

Menzel, K. (2014) Consistent estimation with many moment inequalities. *Journal of Econometrics* 182, 329–350.

Newey, W.K. (1985) Generalized method of moments specification testing. *Journal of Econometrics* 29, 229–256.

Newey, W.K. & F. Windmeijer (2009) Generalized method of moments with many weak moment conditions. *Econometrica* 77, 687–719.

Pagan, A. & J. Robertson (1997) GMM and Its Problems. Discussion paper, Australia National University.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5(2), 181–240.

Phillips, P.C.B. (2006) A remark on bimodality and weak instrumentation in structural equation models. *Econometric Theory* 22(5), 947–960.

Phillips, P.C.B. (2016) Inference in near-singular regression. In G. González-Rivera, R.C. Hill, & T.-H. Lee (eds.), *Essays in Honor of Aman Ullah*. Advances in Econometrics, vol. 36, pp. 461–486. Emerald Group Publishing.

Podivinsky, J.M. (1999) Finite sample properties of GMM estimators and tests. In L. Matyas (ed.), *Generlized Method of Moments Estimation*, Chap. 5, pp. 128–148. Cambridge University Press.

Poskitt, D.S. & C.L. Skeels (2008) Assessing the magnitude of the concentration parameter in a simultaneous equations model. *Econometrics Journal* 12(1), 26–44.

Poskitt, D.S. & C.L. Skeels (2013) *Inference in the Presence of Weak Instruments: A Selected Survey*. Foundations and Trends in Econometrics, vol. 6.1. Now Publishers Inc.

https://doi.org/10.1017/S0266466623000221 Published online by Cambridge University Press

Puri, M.L., C.T. Russell, & T. Mathew (1984) Convergence of generalized inverses with applications to asymptotic hypothesis testing. *Sankhya: The Indian Journal of Statistics, Series A* 46(2), 277–286.

Rao, C.R. & S.K. Mitra (1971) *Generalized Inverse of Matrices and its Applications*. Wiley.

Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica* 39(3), 577–591.

Sargan, J.D. (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26(3), 393–415.

Srivastava, M. (2003) Singular Wishart and multivariate beta distributions. *Annals of Statistics* 31, 1537–1560.

Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3), 557–586.

Stock, J.H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.

Stock, J.H., J.H. Wright, & M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4), 518–529.

Stock, J.H. & M. Yogo (2005) Testing for weak instruments in linear IV regression. In D.W.K. Andrews and J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Chap. 5, pp. 80–108. Cambridge University Press.

Tauchen, G. (1986) Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data. *Journal of Business and Economic Statistics* 4, 397–416.

Windmeijer, F. (2005) A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126, 25–51.

Wright, J.H. (2003) Detecting lack of identification in GMM. *Econometric Theory* 19, 322–330.