cemmap

centre for microdata methods and practice

Generalised Anderson-Rubin statistic based inference in the presence of a singular moment variance matrix

Nicky L. Grant
Richard J Smith

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP05/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Generalised Anderson-Rubin Statistic Based Inference In The Presence Of A Singular Moment Variance Matrix*

Nicky L. Grant†
University of Manchester
nicky.grant@manchester.ac.uk

Richard J Smith‡
*c*emmap, U.C.L and I.F.S.
University of Cambridge
University of Melbourne
ONS Economic Statistics Centre of Excellence
rjs27@econ.cam.ac.uk

This Draft: January 2019

Abstract

The particular concern of this paper is the construction of a confidence region with pointwise asymptotically correct size for the true value of a parameter of interest based on the generalized Anderson-Rubin (GAR) statistic when the moment variance matrix is singular. The large sample behaviour of the GAR statistic is analysed using a Laurent series expansion around the points of moment variance singularity. Under a condition termed first order moment singularity the GAR statistic is shown to possess a limiting chi-square distribution on parameter sequences converging to the true parameter value. Violation, however, of this condition renders the GAR statistic unbounded asymptotically. The paper details an appropriate discretisation of the parameter space to implement a feasible GAR-based confidence region that contains the true parameter value with pointwise asymptotically correct size. Simulation evidence is provided that demonstrates the efficacy of the GAR-based approach to moment-based inference described in this paper.

Keywords: Laurent series expansion; moment indicator; parameter sequences; singular moment matrix.

JEL Classifications: C13, C31, C51, C52.

*Address for correspondence: N.L. Grant, Arthur Lewis Building, Department of Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom.

†Arthur Lewis Building, Department of Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom.

‡Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK.

1 Introduction

The generalized Anderson-Rubin (GAR) statistic is often used as the basis for the construction of an asymptotically valid confidence region for the true value θ_0 of a d_θ-vector $\theta \in \Theta$ of unknown parameters with $\Theta \subseteq \mathbb{R}^{d_\theta}$ the corresponding parameter space. A GAR-based confidence region estimator for θ_0 with asymptotic level α, $0 < \alpha < 1$, is formed by the inversion of the non-rejection region of a GAR-based test with asymptotic size $1 - \alpha$ of the hypothesis $\mathrm{H}_0 : \ \theta = \theta_0$.

To be more precise the moment indicator vector $g(z, \theta)$, a d_g-vector of known functions of the d_z-dimensional data observation vector z and θ, forms the basis for inference on θ_0 in the following discussion and analysis. It is assumed that θ_0 satisfies the population unconditional moment equality condition

$$\mathbb{E}_{\mathcal{P}_0}[g(z, \theta)] = 0. \tag{1.1}$$

where $\mathbb{E}_{\mathcal{P}_0}[\cdot]$ denotes expectation taken with respect to the true population probability law ($\mathcal{P}_0$) of z. Throughout the paper z_i, $(i = 1, \ldots, n)$, will denote a random sample of size n of observations on z. Let $g_i(\theta) = g(z_i, \theta)$ and $G_i(\theta) = \partial g_i(\theta)/\partial\theta'$, $(i = 1, \ldots, n)$, $\hat{g}_n(\theta) = \sum_{i=1}^n g_i(\theta)/n$, $\hat{G}_n(\theta) = \sum_{i=1}^n G_i(\theta)/n$ and $\hat{\Omega}_n(\theta) = \sum_{i=1}^n g_i(\theta) g_i(\theta)'/n$. A GAR-based confidence region estimator for θ_0 is defined in terms of the GAR statistic

$$\hat{T}_n(\theta) = n\hat{g}_n(\theta)'\hat{\Omega}_n(\theta)^{-1}\hat{g}_n(\theta). \tag{1.2}$$

The particular context for this study concerns circumstances in which the variance matrix $\Omega = \mathbb{E}_{\mathcal{P}_0}[g(z, \theta_0)g(z, \theta_0)']$ of the moment indicator vector $g(z, \theta)$ at the true parameter value θ_0 is singular. Since the sample second moment matrix $\hat{\Omega}_n(\theta)$ at $\theta = \theta_0$ is consequentially rendered singular, a number of theoretical issues then arise with the GAR-based approach to confidence region estimation which are highlighted in this paper. First, the GAR statistic (1.2) does not exist for certain parameter sequences θ_n converging to θ_0. Next, even for those parameter sequences for which $\hat{T}_n(\theta_n)$ does exist, the probability limit of $\hat{\Omega}_n(\theta_n)$ is singular. This paper addresses both of these concerns and provides conditions for the construction of a feasible GAR-based confidence region that contains θ_0 with pointwise asymptotically correct size. To derive the requisite asymptotic properties of the GAR statistic $\hat{T}_n(\theta)$ (1.2), the paper adopts an approach new to the literature using a Laurent series expansion of the inverse of the sample moment variance matrix $\hat{\Omega}_n(\theta)$ around points of singularity. The paper places minimal restrictions on the rank and form of the moment variance matrix Ω and the expected Jacobian $G = \mathbb{E}_{\mathcal{P}_0}[\partial g(z, \theta_0)/\partial\theta']$ and provides a direct extension of those results for the GAR statistic with nonsingular Ω in Stock and Wright (2000).

We devote attention primarily to a relatively mild assumption on the columns of the sample Jacobian matrix $\hat{G}_n(\theta_0)$ which we term *first order moment singularity.* This condition enables a Laurent series expansion of $\hat{\Omega}_n(\theta_n)^{-1}$ to be established and, consequently, the existence of the GAR statistic on particular parameter sequences θ_n converging to true parameter value θ_0. We also show that, in the absence of this condition, the GAR statistic is asymptotically unbounded on a subset of such sequences. When Ω is singular the asymptotic size of a GAR-based confidence region depends crucially on the properties of the discretized parameter space Θ_n on which the GAR statistic is inverted in practice. Therefore, the paper details how to discretize appropriately the parameter space Θ to guarantee that it contains parameter sequences for which the GAR statistic is asymptotically chi-square distributed. The feasible GAR-based confidence region contains θ_0 with correct size under relatively mild assumptions and requires no knowledge of points of singularity, so that all such points need not be included in the discretised parameter space Θ_n. Furthermore, feasible GAR-based inference does not require any regularization or pre-testing, and so is less computationally burdensome than a regularization approach. A number of examples of moment functions with singular variance matrix are provided. A simulation study illustrates the results of this paper.

Much of the literature on identification-robust inference originating with Anderson and Rubin (1949) has been concerned with linear instrumental variable (IV) models. An array of alternative approaches providing asymptotically valid inference on θ_0 under a minimal set of assumptions has been developed since this seminal work, including, but not limited to, Andrews and Marmer (2008), Andrews et al. (2007), Chernozhukov et al. (2009), Guggenberger et al. (2012), Kleibergen (2002), Kleibergen and Mavroeidis (2009), Magnusson (2010) and Moreira (2003). Important extensions to nonlinear moment indicator functions have been developed in which Ω is maintained non-singular. For example, Stock and Wright (2000) study confidence regions formed by inverting the GAR statistic (1.2) non-rejection region under weak identification. These results are generalized to the many weak moment setup in Newey and Windmeijer (2009). Guggenberger and Smith (2005, 2008), Kleibergen (2005) and Guggenberger et al. (2012) consider GMM and generalized empirical likelihood based inference with the GAR statistic as a special case.

In nonlinear models identification failure may directly result in the singularity of Ω. Consequently, research has focused on methods of inference that do not require Ω to be full rank. Andrews and Guggenberger (2018) study the asymptotic properties of, among others, a singularity-robust-GAR (SR-GAR) statistic that deletes redundant directions of the moment indicator vector across the parameter space Θ. This method allows for general forms of Ω. However, to obtain correct asymptotic size for an SR-GAR-based confidence region, all points of singularity are required to be included in the discretized parameter space used in practice. Dufour and Valéry (2016) develop a regularized Wald statistic providing valid pointwise inference on strongly identified functions of θ_0 in the presence of a singular Ω. Peñaranda and Sentana (2012) study GMM inference when the rank of Ω is known using a generalized inverse of the sample moment variance matrix. Another strand in the literature has studied the asymptotic properties of confidence regions based on various statistics with specific forms of singular moment indicator variance matrix. Andrews and Cheng (2012, 2013, 2014) and Cheng (2014) derive conditions for valid subvector inference (in a uniform sense) using t, Wald, quasi-likelihood ratio and maximum likelihood (ML) statistics. They consider moment indicators with a singular variance matrix arising from identification failure in a class of nonlinear models. Rotnitzky, Cox, Bottai and Robins (2000) study the asymptotic properties of inference based on the likelihood ratio statistic when Ω is rank $d_\theta - 1$. In related work, Bottai (2003) considers inference from various ML-based statistics where $d_\theta = 1$ and $\Omega = 0$.

The remainder of the paper is set out as follows. Section 2 details the notation used in the paper. Section 3 studies asymptotic properties of the GAR statistic (1.2) when Ω is singular. Section 4 details feasible GAR-based confidence regions formed by inverting the GAR statistic over a discretization Θ_n of the parameter space Θ. Section 5 establishes sufficient conditions appropriate for nonlinear conditional moment restriction models. A simulation study is provided in Section 6, corroborating the main results of this paper. Section 7 presents conclusions and directions for further research. The Appendices collect proofs of the main theorems and subsidiary lemmas used in this paper. Supplements E and S contain examples and additional simulation evidence.

2 Notation

$\xrightarrow{p}$, $\xrightarrow{d}$ denote convergence in probability and distribution, respectively, 'w.p.(a.)1' is 'with probability (approaching) 1' and 'i.i.d.' is 'independent and identically distributed'. $o_p(a)$ and $O_p(a)$ respectively indicate a variate that, after division by a, converges to zero w.p.a.1 and to a variate that is bounded w.p.a.1 by a bounded non-stochastic sequence; similar definitions apply for their deterministic counterparts $o(a)$, $O(a)$. For an arbitrary random variable x, a.s.(x) denotes 'almost surely' x.

$\mathbb{E}_{\mathcal{P}_0}[\cdot]$ and $\mathrm{Var}_{\mathcal{P}_0}[\cdot]$ denote expectation and variance taken with respect to the true population probability law ($\mathcal{P}_0$) of z. For arbitrary random variables y and x, $\mathbb{E}_{\mathcal{P}_0}[y|x]$ is the conditional expectation of y given x. χ^2_k denotes a central chi-square distributed random variable with k degrees of freedom.

$\mathrm{rk}(A)$ and $\mathcal{N}(A)$ denote the rank and (right) null space, respectively, of a matrix A whereas $\mathrm{tr}(A)$ and $\det(A)$ are the trace and determinant, respectively, of a square matrix A. For integer $k > 0$, I_k denotes a $k \times k$ identity matrix. For a full column rank $k \times p$ matrix A and $k \times k$ nonsingular matrix K, $P_A(K)$ denotes the oblique projection matrix $A(A'K^{-1}A)^{-1}A'K^{-1}$ and $M_A(K) = I_k - P_A(K)$ its orthogonal counterpart. We abbreviate this notation to P_A and M_A if $K = I_k$ and, if $p = 0$, set $M_A = I_k$. For square matrices A and B, $\mathrm{diag}(A, B)$ denotes a block-diagonal matrix with diagonal blocks A and B.

$\|A\| = \mathrm{tr}(A'A)^{1/2}$ denotes the Euclidean norm of the matrix A and $d(x, y) = \|x - y\|$ the Euclidean distance between vectors $x, y \in \mathbb{R}^d$ for integer $d \geq 1$. The Hausdorff distance between sets A and B is defined as $d_H(A, B) = \max\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(A, b)\}$, where $d(a, B) = \inf_{b\in B} \|b - a\|$, and $d_H(A, B) = \infty$ if either A or B is the null set $\emptyset$.

The expectations of the moment vector, Jacobian matrix and moment function second moment matrix are defined as $g(\theta) = \mathbb{E}_{\mathcal{P}_0}[g(z, \theta)]$, $G(\theta) = \mathbb{E}_{\mathcal{P}_0}[\partial g(z, \theta)/\partial\theta']$ and $\Omega(\theta) = \mathbb{E}_{\mathcal{P}_0}[g(z, \theta)g(z, \theta)']$ respectively, $\theta \in \Theta$.

At the true value θ_0 the dependence on θ_0 is suppressed where there can be no confusion. Thus, $g_i = g_i(\theta_0)$, $G_i = G_i(\theta_0)$, $(i = 1, \ldots, n)$, $\hat{g}_n = \hat{g}_n(\theta_0)$ and $\hat{\Omega}_n = \hat{\Omega}_n(\theta_0)$. Recall $G = G(\theta_0)$ and $\Omega = \Omega(\theta_0)$.

3 Asymptotic Properties of the GAR Statistic with Singular Moment Variance

When Ω is singular, $\hat{\Omega}_n$ is also singular w.p.1 so that the GAR statistic $\hat{T}_n(\theta)$ (1.2) does not exist at $\theta = \theta_0$. Consequently, its large sample properties can no longer be studied using the standard analysis to be found in, e.g., Stock and Wright (2000). To deal with this difficulty, this section provides conditions under which $\hat{T}_n(\theta_n)$ exists w.p.a.1 and $\hat{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$ for suitable parameter sequences $\theta_n = \theta_0 + o(n^{-1/2})$.

Remark 3.1. Another form of GAR statistic is $\tilde{T}_n(\theta) = n\hat{g}_n(\theta)'\tilde{\Omega}_n(\theta)^{-1}\hat{g}_n(\theta)$ based on the alternative moment variance matrix estimator $\tilde{\Omega}_n(\theta) = \sum_{i=1}^n (g_i(\theta) - \bar{g}_n(\theta))(g_i(\theta) - \bar{g}_n(\theta))'/n$. Although the discussion below primarily concerns the GAR statistic $\hat{T}_n(\theta)$, the results for the alternative GAR statistic $\tilde{T}_n(\theta_n)$ are almost identical to those for $\hat{T}_n(\theta_n)$.

By symmetry, the moment indicator second moment matrix $\Omega(\theta)$ satisfies the spectral decomposition

$$\begin{aligned} \Omega(\theta) &= P(\theta)\Lambda(\theta)P(\theta)' \\ &= P_+(\theta)\Lambda_+(\theta)P_+(\theta)' + P_0(\theta)\Lambda_0(\theta)P_0(\theta)' \\ &= P_+(\theta)\Lambda_+(\theta)P_+(\theta)' \end{aligned} \tag{3.1}$$

where the eigen-vectors and eigen-values of $\Omega(\theta)$ respectively constitute the columns of the $d_g \times d_g$ orthonormal matrix $P(\theta)$ and the diagonal elements, arranged in non-increasing order of magnitude, of the $d_g \times d_g$ diagonal matrix $\Lambda(\theta)$. Given $\mathrm{rk}(\Omega(\theta)) = r_\Omega(\theta)$, defining $\bar{r}_\Omega(\theta) = d_g - r_\Omega(\theta)$, $P(\theta) = (P_+(\theta), P_0(\theta))$ and $\Lambda(\theta) = \mathrm{diag}(\Lambda_+(\theta), \Lambda_0(\theta))$ are partitioned such that $P_+(\theta)$ and $P_0(\theta)$ are of dimensions $d_g \times r_\Omega(\theta)$ and $d_g \times \bar{r}_\Omega(\theta)$ respectively with $\Lambda_+(\theta)$ and $\Lambda_0(\theta)$ $r_\Omega(\theta) \times r_\Omega(\theta)$ and $\bar{r}_\Omega(\theta) \times \bar{r}_\Omega(\theta)$ diagonal matrices respectively with the $r_\Omega(\theta)$ positive and $\bar{r}_\Omega(\theta)$ zero eigen-values as diagonal elements. At the true value θ_0, we write $r_\Omega = r_\Omega(\theta_0)$, $\bar{r}_\Omega = d_g - r_\Omega$, $P_+ = P_+(\theta_0)$, $P_0 = P_0(\theta_0)$, $\Lambda_+ = \Lambda_+(\theta_0)$ and $\Lambda_0 = \Lambda_0(\theta_0)$.

To study the limiting properties of the GAR statistic $\hat{T}_n(\theta)$ (1.2), define the set of sequences

$$\Theta_n^0(\delta) = \{\theta_n \in \Theta : \ \theta_n = \theta_0 + n^{-\varepsilon}\delta_n, \ \varepsilon > 1/2, \ n^{1/2}(\delta_n - \delta) \to 0, \ \delta \neq 0, \ \delta \in \mathbb{R}^{d_\theta}\}. \tag{3.2}$$

Section 3.1 provides conditions under which $\hat{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$ for $\theta_n \in \Theta_n^0(\delta)$ (3.2) when Ω has deficient rank. Section 3.2 shows that the GAR statistic is asymptotically unbounded, and, thus, GAR-based confidence regions empty, if a particular hypothesis of Section 3.1 fails to hold. Section 3.3 discusses the potential importance of the singularity of Ω for the construction of the GAR statistic (1.2) and the consequent GAR-based confidence region for θ_0.

3.1 Asymptotic Properties when $P_0'G\delta = 0$

When Ω is singular, then $P_0'\Omega P_0 = 0$ and, thus, $P_0'g_i = 0$ w.p.1., $(i = 1, \ldots, n)$, i.e., $\bar{r}_\Omega$ linearly independent combinations of the moment indicator function $g(z, \theta)$ evaluated at θ_0 are degenerate. This section provides results on the large sample behaviour of the GAR statistic (1.2) under a condition that we term *first order moment singularity*, namely, there exists $\delta \in \mathbb{R}^{d_\theta}$ such that $\text{rk}(\text{Var}_{\mathcal{P}_0}[P_0'(G_i\delta - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- g_i)]) = \bar{r}_\Omega$, where $\Omega^- = P_+\Lambda_+^{-1}P_+'$ is the Moore-Penrose inverse of Ω.

The following example highlights the importance of first order moment singularity for establishing the asymptotic properties of $\hat{T}_n(\theta_n)$ for sequences $\theta_n \in \Theta_n^0(\delta)$ when Ω is singular. Suppose $g_i(\theta)$ is linear in $\theta \in \Theta$, i.e., $G_i(\theta) = G_i$, $(i = 1, \ldots, n)$. Setting $\delta_n = \delta$ for simplicity, then, for $\theta_n \in \Theta_n^0(\delta)$, substituting for θ_n,

$$n^\varepsilon P_0' g_i(\theta_n) = P_0' G_i \delta \text{ w.p.1}$$

since $P_0' g_i = 0$ w.p.1, i.e., there exists $\bar{r}_\Omega$ linearly independent combinations of $n^\varepsilon g_i(\theta_n)$ that do not involve g_i, $(i = 1, \ldots, n)$. Unlike the full rank case, the first order asymptotic properties of $\hat{T}_n(\theta_n)$ depend not only on the mean and variance matrix of $P_+' g_i$ but those of $P_0' G_i \delta$ or, more precisely, $P_0'(G_i\delta - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- g_i)$. For $\theta_n \in \Theta_n^0(\delta)$,

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}_0}[n^\varepsilon P_0' g_i(\theta_n) - P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- g_i(\theta_n)] &= P_0' G\delta - n^{-\varepsilon} P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- G\delta \\
&= P_0' G\delta + o(1)
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}_{\mathcal{P}_0}[n^\varepsilon P_0' g_i(\theta_n) - P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- g_i(\theta_n)] &= P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i'] - G\delta\delta' G' \\
&\quad -\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']')P_0 + o(1).
\end{aligned}$$

Moreover, and importantly,

$$\text{Cov}_{\mathcal{P}_0}[P_+' g_i(\theta_n), n^\varepsilon P_0' g_i(\theta_n) - P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- g_i(\theta_n)] = o(1).$$

For the GAR statistic (1.2) to exist and be asymptotically bounded for sequences $\theta_n \in \Theta_n^0(\delta)$, we require that $n^\varepsilon P_0' g_i(\theta_n)$ has zero mean, i.e., $P_0' G\delta = 0$. Moreover, not only should $n^\varepsilon P_0' g_i(\theta_n)$ have full rank variance matrix, i.e., $\text{rk}(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0) = \bar{r}_\Omega$, but, crucially, the variance matrix of $n^\varepsilon P_0' g_i(\theta_n) - P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- g_i(\theta_n)$ should also be full rank, i.e., $\text{rk}(P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i'] - G\delta\delta' G' - P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']')P_0) = \bar{r}_\Omega$. If both of these conditions are met, then $\hat{T}_n(\theta_n)$ can be expressed w.p.a.1 as a quadratic form in d_g moment functions with zero mean and full rank variance matrix, namely, $P_+'\hat{g}$ and $P_0'(\hat{G}\delta - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- \hat{g})$ that, asymptotically, are both mean zero with full rank r_Ω and $\bar{r}_\Omega$ variance matrices respectively and, critically, are asymptotically uncorrelated. Hence, for any sequence $\theta_n \in \Theta_n^0(\delta)$ satisfying these conditions, $\hat{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$.

More generally, under the regularity condition Assumption 3.1, Theorem 3.1 below states that $\hat{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$ for any sequence $\theta_n \in \Theta_n^0(\delta)$ that satisfies Assumption 3.2.

Assumption 3.1. ***(a)*** *z_i, $(i = 1, \ldots, n)$, is a random sample of size n on the d_z-dimensional observation vector z;* ***(b)*** *$\mathbb{E}_{\mathcal{P}_0}[\sup_{\theta\in\Theta}\|g(z,\theta)\|^2] < \infty$, $\mathbb{E}_{\mathcal{P}_0}[\sup_{\theta\in\Theta}\|\partial g(z,\theta)/\partial\theta'\|^2] < \infty$;* ***(c)*** *$\|\hat{\Omega}_n(\theta_a) - \hat{\Omega}_n(\theta_b)\| \leq \hat{M}_{\Omega,n}\|\theta_a - \theta_b\|$ uniformly $\theta_a, \theta_b \in \Theta$ for some $\hat{M}_{\Omega,n} = O_p(1)$;* ***(d)*** *$\|\hat{G}_n(\theta_a) - \hat{G}_n(\theta_b)\| \leq \hat{M}_{G,n}\|\theta_a - \theta_b\|$ uniformly $\theta_a, \theta_b \in \Theta$ for some $\hat{M}_{G,n} = O_p(1)$.*

Remark 3.2. Assumption 3.1 provides a set of relatively mild regularity conditions similar to those commonly made in the literature on GAR-based inference. The random sampling Assumption 3.1(a) is primarily made for simplicity but could be weakened straightforwardly to allow for non-i.i.d. data. Assumptions A.1(b) and (c) ensure $\hat{g}_n(\theta) \xrightarrow{p} g(\theta)$, $\hat{G}_n(\theta) \xrightarrow{p} G(\theta)$ and $\hat{\Omega}_n(\theta) \xrightarrow{p} \Omega(\theta)$ uniformly $\theta \in \Theta$ by an i.i.d. uniform weak law of large numbers.

Define the sets

$$
\begin{aligned}
\Delta_b &= \{\delta \in \mathbb{R}^{d_\theta} : \ \mathrm{rk}(\mathrm{Var}_{\mathcal{P}_0}[P_0'(G_i\delta - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^- g_i)] = \bar{r}_\Omega\}, \\
\Delta_c &= \{\delta \in \mathbb{R}^{d_\theta} : \ P_0'G\delta = 0\}.
\end{aligned}
$$

Assumption 3.2. ***(a)*** $\mathrm{rk}(\Lambda_+) = r_\Omega$; ***(b)*** $\Delta_b \neq \emptyset$; ***(c)*** $\Delta_b \cap \Delta_c \neq \emptyset$.

Remark 3.3. A necessary and sufficient condition for Assumptions 3.2(a) and (b) is that the variance matrix of $P_+'g_i$ and $P_0'G_i\delta$ is full rank d_g.

Define the set of sequences

$$
\Theta_n^0(\Delta_b \cap \Delta_c) = \{\theta_n \in \Theta_n^0(\delta) : \ \delta \in \Delta_b \cap \Delta_c\}.
$$

Theorem 3.1. *Let Assumptions 3.1 and 3.2 hold. Then, for any sequence $\theta_n \in \Theta_n^0(\Delta_b \cap \Delta_c)$,* ***(a)*** $\hat{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$; ***(b)*** $\tilde{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$.

Remark 3.4. If Assumptions 3.1 and 3.2(a) hold, it is straightforward to establish Theorem 3.1 whenΩ is full rank, i.e., $\bar{r}_\Omega = 0$; cf. Remark 3.3. A key step in the Proof of Theorem 3.1 is an expansion for $\hat{T}_n(\theta_n)$ based on a Laurent series expansion of $\hat{\Omega}_n(\theta_n)^{-1}$ around points of singularity; see Lemma A.3 in Appendix A. Assumption 3.2(b), first order moment singularity, is a relatively mild requirement and is sufficient if Assumption 3.1 holds for the Laurent series expansion of $\hat{\Omega}_n(\theta_n)^{-1}$ needed to establish the large sample properties of the statistic $\hat{T}_n(\theta_n)$ and those sequences $\theta_n \in \Theta$ for which $\hat{T}_n(\theta_n)$ exists w.p.a.1.[1]

Remark 3.5. Given Assumption 3.2(b), Assumption 3.2(c) requires that there exists $\delta \in \Delta_b$ such that $P_0'G\delta = 0$. Thus $\hat{T}_n(\theta_n)$ is asymptotically bounded; see the Proof of Theorem 3.1. A sufficient condition for Assumption 3.2(c) if Assumption 3.2(b) is satisfied, is $\mathcal{N}(\Omega) \subseteq \mathcal{N}(G')$, i.e., $G'P_0 = 0$, and, thus, $\Delta_c = \mathbb{R}^{d_\theta}$. Section 5 provides general conditions for $\mathcal{N}(\Omega) \subseteq \mathcal{N}(G')$ for nonlinear conditional moment functions; these conditions always hold for single equation nonlinear least squares moment functions but may not do so in multiple equation models.

3.2 Asymptotic Properties when $P_0'G\delta \neq 0$

The next theorem establishes that $\hat{T}_n(\theta_n) = O_p(n)$ for sequences $\theta_n \in \Theta_n^0(\delta)$ that satisfy Assumption 3.2(b) but when Assumption 3.2(c) fails, i.e., $P_0'G\delta \neq 0$.

[1] To illustrate, consider the nonlinear regression $\epsilon(\theta) = y - h(x,\theta)$. Non-linear LS estimation of θ_0 defines $g(z,\theta) = \epsilon(\theta)\partial h(x,\theta)/\partial\theta$ and $\Omega(\theta) = \mathbb{E}_{\mathcal{P}_0}[\epsilon(\theta)^2 \partial h(x,\theta)/\partial\theta\partial h(x,\theta)/\partial\theta']$. If $P_0'g(z,\theta_0) = 0$, i.e., $P_0'\partial h(x,\theta_0)/\partial\theta = 0$, Assumption 3.2(b) fails if and only if $\mathrm{Var}[P_0'\partial^2 h(x,\theta_0)/\partial\theta\partial\theta'\delta] < \bar{r}_\Omega$ for all $\delta \in \mathbb{R}^{d_\theta}$ noting $\partial g(x,\theta)/\partial\theta' = \epsilon(\theta)\partial^2 h(x,\theta)/\partial\theta\partial\theta' + \partial h(x,\theta)/\partial\theta\partial\epsilon(\theta)/\partial\theta'$.

Let $\bar{\Delta}_c$ denote the complement of Δ_c, i.e.,

$$\bar{\Delta}_c = \{\delta \in \mathbb{R}^{d_\theta} : \; P_0' G\delta \neq 0\}.$$

Also let $\Upsilon(\delta) = P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0$ and $\Sigma(\delta) = \Upsilon(\delta) - \Psi(\delta)\Omega^- \Psi(\delta)'$, where $\Psi(\delta) = P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']$ and $\Omega^- = P_+\Lambda_+^{-1}P_+'$. Define $\tilde{\Upsilon}(\delta) = \Upsilon(\delta) - P_0'G\delta\delta G' P_0$ and $\tilde{\Sigma}(\delta) = \tilde{\Upsilon}(\delta) - \Psi(\delta)\Omega^-\Psi(\delta)'$.

Theorem 3.2. *Suppose Assumptions 3.1, 3.2(a) and (b) are satisfied. Then, for all $\theta_n \in \Theta_n^0(\Delta_b \cap \bar{\Delta}_c)$, **(a)** $\hat{T}_n(\theta_n)/n \xrightarrow{p} \delta' G' P_0 \Sigma(\delta) P_0' G\delta$; **(b)** $\tilde{T}_n(\theta_n)/n \xrightarrow{p} \delta' G' P_0 \tilde{\Sigma}(\delta) P_0' G\delta$.*

Remark 3.6. Unless $P_0' G = 0$, i.e., $\mathcal{N}(\Omega) \subseteq \mathcal{N}(G')$, there exist sequences $\theta_n \in \Theta_n^0(\Delta_b \cap \bar{\Delta}_c)$ for which the GAR statistic $\hat{T}_n(\theta_n)$ is asymptotically unbounded. Section 4 shows how an appropriate discretisation Θ_n of the parameter space Θ to a sufficiently fine degree ensures sequences $\theta_n \in \Theta_n \cap \Theta_n^0(\Delta_b \cap \Delta_c)$ for large enough n and, thereby, that Theorem 3.1 is satisfied for such sequences.

3.3 Degeneracy in the Parameter Space

The singularity of $\Omega(\theta)$, $\theta \in \Theta$, leads to degeneracies in the parameter space, i.e., parameter subsets where particular functions of z and θ are constant w.p.1.[2]

To illustrate, consider the scalar moment indicator $g(z, \theta) = \exp(\theta z) - (1 + \theta)$ in which case $d_g = d_\theta = 1$. Suppose $\bar{r}_\Omega = 1$. Hence, $g(z, \theta_0) = 0$ for all $z \in \mathbb{R}$ and $\theta_0 = 0$ w.p.1, i.e, the true value $\theta_0 = 0$ is identified w.p.1 from a single observation z. Although it is obviously still the case that $\mathbb{E}_{\mathcal{P}_0}[g(z, \theta)] = 0$ at $\theta_0 = 0$, any GAR-based confidence region is rendered irrelevant. No matter that GAR-based inference would not be considered, the hypotheses of Theorem 3.1 may hold. In this example, $P_0 = 1$, so $\mathrm{Var}_{\mathcal{P}_0}[P_0' G_i\delta] > 0$ satisfies Assumption 3.2(b) for $\Delta_b = \{\delta \in \mathbb{R} : \; |\delta| > 0\}$. Since $G = \mathbb{E}_{\mathcal{P}_0}[z] - 1$, Assumption 3.2(c) holds if $G\delta = 0$ which is possible if and only if $G = 0$ given $\delta \neq 0$, i.e., $\mathbb{E}_{\mathcal{P}_0}[z] = 1$ and, thus, $\Delta_c = \mathbb{R}$. Then, from Theorem 3.1, $\hat{T}(\theta_n) \xrightarrow{d} \chi_1^2$ for all sequences $\theta_n \in \Theta_n^0(\Delta_b \cap \Delta_c)$, i.e, $\theta_n \in \Theta_n^0(\Delta_b)$ when $G = 0$. In fact, this result may be shown directly from a Taylor expansion of $\hat{\Omega}(\theta_n)^{-1}$ around θ_0. When $G \neq 0$, then $\Delta_c = \emptyset$ and $\hat{T}(\theta_n) = O_p(n)$ for all $\theta_n \in \Theta_n^0(\Delta_b)$ by Theorem 3.2 as $\Delta_b \cap \bar{\Delta}_c = \Delta_b$ in this case too.

In the one moment, one parameter, setting, the singularity of Ω, i.e., $\Omega = 0$, implies that the moment condition holds w.p.1 and the true value θ_0 of the parameter vector θ can be deduced from a single observation of z. Similarly, and more generally, in the just-identified case, i.e., $d_g = d_\theta$, a single observation of z is sufficient to deduce the true value θ_0. However, in the more general over-identified setting when $d_g > d_\theta$ and in which $\bar{r}_\Omega(\theta) > 0$ linear combinations of $g(z, \theta)$ are degenerate, the value θ is restricted to a $d_{\theta - \bar{r}_\Omega(\theta)}$ dimensional stochastic (sub-)manifold of Θ w.p.1 by the restriction $P_0(\theta)' g(z, \theta) = 0$ w.p.1. Let $g_\theta(z, \theta) = P_0(\theta)' g(z, \theta)$, i.e., $g_\theta(z, \theta) = 0$ w.p.1, and is, of course, invariant to z. Partition $\theta = (\theta'_{d_\theta - \bar{r}_\Omega(\theta)}, \theta'_{\bar{r}_\Omega(\theta)})'$, where $\theta_{d_\theta - \bar{r}_\Omega(\theta)}$ and $\theta_{\bar{r}_\Omega(\theta)}$ are, respectively, $d_\theta - \bar{r}_\Omega(\theta)$ and $\bar{r}_\Omega(\theta)$ sub-vectors of θ with the sub-vector $\theta_{\bar{r}_\Omega(\theta)}$ chosen so that the derivative matrix $\partial g_\theta(z, \theta)/\partial\theta_{\bar{r}_\Omega(\theta)}$ is nonsingular. Thus, by the implicit function theorem, there exists a function $\theta_{\bar{r}_\Omega(\theta)}(\cdot)$ such that $\theta_{\bar{r}_\Omega(\theta)} = \theta_{\bar{r}_\Omega(\theta)}(\theta_{d_\theta - \bar{r}_\Omega(\theta)})$. Consequently, the parametric dimension d_θ is reduced to $d_{\theta - \bar{r}_\Omega(\theta)}$, i.e., $\theta_{d_{\theta - \bar{r}_\Omega(\theta)}}$, and, correspondingly, the moment function

[2] We are indebted to P.C.B. Phillips for raising this issue.

dimension d_g to $r_\Omega(\theta)$, i.e., $P_+(\theta_{d_\theta-\bar{r}_\Omega(\theta)}, \theta_{\bar{r}_\Omega(\theta)}(\theta_{d_\theta-\bar{r}_\Omega(\theta)}))'g(z, \theta_{d_\theta-\bar{r}_\Omega(\theta)}, \theta_{\bar{r}_\Omega(\theta)}(\theta_{d_\theta-\bar{r}_\Omega(\theta)}))$, since the moment restriction $P_0(\theta_{d_\theta-\bar{r}_\Omega(\theta)}, \theta_{\bar{r}_\Omega(\theta)}(\theta_{d_\theta-\bar{r}_\Omega(\theta)}))'g(z, \theta_{d_\theta-\bar{r}_\Omega(\theta)}, \theta_{\bar{r}_\Omega(\theta)}(\theta_{d_\theta-\bar{r}_\Omega(\theta)})) = 0$ w.p.1 is now redundant. To illustrate, consider Example E.1 in Supplement E in which $P_0(\theta)'g(z,\theta) = 0$ at $\theta = \theta_0$ implies $y_1 - y_2 = x_1\theta_{10} - x_2\theta_{20}$ w.p.1. Hence, $\theta_{10} = (y_1 - y_2)/x_1 + \theta_{20}x_2/x_1$ w.p.1 if $\mathcal{P}_0\{x_1 = 0\} = 0$. Thus, $g(z, \theta_0)$ can be expressed as a function of θ_{20} only; written this way the first two moment indicators are identical and the first moment function can be dropped in this example.

In principle, then, the singularity of $\Omega(\theta)$ allows the dimensions d_g of the moment indicator function $g(z,\theta)$ and d_θ of the parameter vector θ to be reduced to, respectively, $r_\Omega(\theta)$ and $d_{\theta-\bar{r}_\Omega(\theta)}$ with the $\bar{r}_\Omega(\theta)$ redundant directions $P_0(\theta)'g(z,\theta)$ of the moment indicator vector $g(z,\theta)$ consequently deleted. A GAR-type statistic may then be based on the reduced moment indicator vector $P_+(\theta)'g(z,\theta)$ which, by definition, has full rank variance matrix. This is, essentially, the approach taken in Andrews and Guggenberger (2018) to the formulation of the SR-GAR statistic with $P_+(\theta)$ and $P_0(\theta)$ estimated from $\hat{\Omega}_n(\theta)$. In the just-identified Example E.3 in Supplement E, where singularity of Ω occurs at $\sigma_0 = a_0$, setting parameter values so that $\sigma = a$ and deleting the last moment indicator reduces the parametric dimension d_θ to $d_\theta - \bar{r}_\Omega(\theta) = 3$ and the moment indicator dimension d_g to $r_\Omega(\theta) = 3$. The difficulty with the Andrews and Guggenberger (2018) approach is that, in general, all points of singularity, i.e., those $\theta \in \Theta$ such that $\bar{r}_\Omega(\theta) > 0$, cannot be known *a priori*. Even if the points of singularity were to be known, it may not be possible to include all such points in the discretised parameter set Θ_n necessitated for practical implementation of a GAR statistic based on the reduced moment indicator vector $P_+(\theta)'g(z,\theta)$ and of a consequent GAR-based confidence region for θ_0. See Section 4 for further discussion of the construction of the discretised parameter set Θ_n.

Remark 3.7. The Andrews and Guggenberger (2018) SR-GAR method is particularly suited to the case in which $P_0(\theta) = P_0$ and, thus, the moment indicator vector $g(z,\theta)$ is redundant for all $\theta \in \Theta$ in the column directions of P_0. Hence, the GAR statistic (1.2) does not exist for any value of $\theta \in \Theta$. Indeed, this case does not satisfy the hypotheses of Theorem 3.1. To see this, by the mean value theorem, $P_0'g_i(\theta_a) = P_0'g_i(\theta_b) + P_0'G_i(\bar{\theta})(\theta_a - \theta_b)$ for any $\theta_a, \theta_b \in \Theta$, where $\bar{\theta}$ lies on the line segment joining θ_a and θ_b, and, thus, $P_0'G_i(\bar{\theta})(\theta_a - \theta_b) = 0$. Hence, since θ_a and θ_b are arbitrary, $P_0'G_i(\theta) = 0$ for all $\theta \in \Theta$, a negation of Assumption 3.2(b). With identities in the moment indicator vector, redundant moments can be deleted straightforwardly and standard GAR-based inference undertaken if the reduced dimension moment indicator vector at θ_0 is not subject to further, possibly unknown, singularities.

4 Feasible GAR-Based Confidence Regions

This section details a discretisation Θ_n of the parameter space Θ required to implement a feasible GAR-based confidence region to ensure, for n large enough, coverage of (a subset of) sequences $\theta_n \in \Theta_n^0(\delta)$ for any arbitrary $\delta \in \mathbb{R}^{d_\theta}$ where $\theta_n \in \Theta_n^0(\delta)$ is defined in (3.2). The approach described below then exploits Theorem 3.1 to construct a GAR-based confidence region which is shown to contain θ_0 with asymptotically correct size.

The concern then is a study of the asymptotic properties of the feasible GAR-based confidence region

$$\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha)) = \{\theta \in \Theta_n : \ \hat{T}_n(\theta) \leq \chi^2_{d_g}(\alpha)\}, \tag{4.1}$$

where $\chi^2_{d_g}(\alpha)$ denotes the $100 \times \alpha$ percentile of the $\chi^2_{d_g}$ distribution. The confidence region $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ (4.1) is formed by the inversion over Θ_n of the non-rejection region of a GAR-based test with asymptotic size $1-\alpha$ of the hypothesis $\mathrm{H}_0 : \ \theta = \theta_0$.[3]

Remark 4.1. If Ω is non-singular, i.e., $r_\Omega = d_g$, under suitable regularity conditions, see, e.g., Stock and Wright (2000), $\hat{T}_n(\theta_0) \to \chi^2_{d_g}$. Hence, pointwise, $\lim_{n\to\infty} \mathcal{P}_0\{\theta_0 \in \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))\} = \alpha$ for θ_0 satisfying (1.1).

In general, however, there is no guarantee that Θ_n contains θ_0 or any particular parameter sequence θ_n converging to θ_0. Unless $P_0'G = 0$, i.e., $\mathcal{N}(\Omega) \subseteq \mathcal{N}(G')$, there exist sequences $\theta_n \in \Theta^0_n(\Delta_b \cap \bar{\Delta}_c)$ for which the GAR statistic $\hat{T}_n(\theta_n)$ is asymptotically unbounded by Theorem 3.2. Consequently, GAR-based confidence regions are empty asymptotically, i.e., $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha)) = \emptyset$ w.p.a.1 for any $\alpha \in (0,1)$. If $\Theta_n \cap \Theta^0_n(\Delta_b \cap \Delta_c) \neq \emptyset$ for n large enough, then, from Theorem 3.1, $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ contains sequences θ_n such that $d(\theta_n, \theta_0) = o(n^{-1/2})$ and has correct asymptotic level α. Our concern then is to show how discretising Θ to a sufficiently fine degree ensures that $\Theta_n \cap \Theta^0_n(\Delta_b \cap \Delta_c) \neq \emptyset$ for large enough n.

Remark 4.2. Theorem 15.1, p.20, in the Supplement to Andrews and Guggenberger (2018) demonstrates that a confidence region formed by inverting the SR-GAR based non-rejection region constructed using a Moore-Penrose pseudoinverse of $\Omega_n(\theta)$ over Θ has correct asymptotic size. However, for this result to hold when the SR-GAR based confidence region is constructed in practice with a discretised parameter space Θ_n, $\theta_0 \in \Theta_n$ (or $\theta_n \in \Theta_n$ with $\mathrm{rk}(\hat{\Omega}_n(\theta_n)) = \mathrm{rk}(\hat{\Omega}_n(\theta_0))$ w.p.a.1) is required for correct asymptotic level when Ω is singular. The asymptotic properties of the SR-GAR statistic in this case may be studied using the methods in this paper, or similar.

The asymptotic properties of the GAR-based confidence region estimator $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ (4.1) are studied for Θ_n either fixed or selected at random. Hence the events $\{\theta_n \in \Theta_n\}$ and $\{\hat{T}_n(\theta_n) \leq \chi^2_{d_g}(\alpha)\}$ are independent, i.e.,

$$\mathcal{P}_0\{\theta_n \in \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))\} = \mathcal{P}_0\{\theta_n \in \Theta_n\}\mathcal{P}_0\{\hat{T}_n(\theta_n) \leq \chi^2_{d_g}(\alpha)\}. \tag{4.2}$$

Remark 4.3. Even if Ω is full rank, when $\hat{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$, from (4.2), $\lim_{n\to\infty} \mathcal{P}_0\{\theta_0 \in \Theta_n\} = 1$ is a necessary condition for the confidence region $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ (4.1) to have correct asymptotic size α. However, this condition is non-trivial and not readily verifiable in practice, irrespective of how finely the parameter space Θ is discretized.[4]

The issue noted in Remark 4.3 is especially important when Ω is singular, as indicated by Theorem 3.1, since if $\mathcal{N}(\Omega) \not\subseteq \mathcal{N}(G')$, then the GAR statistic $\hat{T}_n(\theta_n)$ (1.2) is asymptotically unbounded for $\theta_n \in \Theta^0_n(\Delta_b \cap \bar{\Delta}_c)$ and the confidence region $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ (4.1) correspondingly empty

[3]Mikusheva (2010) proposed a similar solution that inverts $\hat{T}_n(\theta)$ over some Θ_n in a one parameter linear IV model with non-singular moment variance matrix that is equivalent to the solution of a system of quadratic inequalities inverted over Θ_n that can be solved explicitly.

[4]Consider $d_\theta = 1$ with $\Theta = [0,1]$. Suppose $\Theta_n = \{0, \frac{1}{n}, ..., \frac{n-1}{n}, 1\}$ and $\theta_0 = 1/\sqrt{2}$. Then $\theta_0 \notin \Theta_n$ for all n and, thus, $\mathcal{P}_0\{\theta_0 \in \Theta_n\} = 0$. Hence $\mathcal{P}_0\{\theta_n \in \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))\} = 0$ by (4.2).

w.p.a.1 if $\Theta_n \subseteq \Theta_n^0(\Delta_b \cap \bar{\Delta}_c)$. Hence, in general, the discretized set Θ_n should contain a subset of sequences $\theta_n \in \Theta_n^0(\Delta_b \cap \Delta_c)$ such that, by Theorem 3.1, the GAR statistic is asymptotically distributed as a $\chi^2_{d_g}$ random variate. Therefore, the GAR-based confidence region $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ will include sequences $\theta_n \in \Theta_n^0(\Delta_b \cap \Delta_c)$ with correct asymptotic size α. To construct a confidence region that includes θ_0 with probability α asymptotically, first consider the discretisation

$$\Theta_n = \{-k_n, -\frac{[n^\kappa]-1}{n^\kappa}k_n, ..., -\frac{1}{n^\kappa}, 0, \frac{1}{n^\kappa}, ..., \frac{[n^\kappa]-1}{n^\kappa}k_n, k_n\}^{d_\theta} \tag{4.3}$$

when $d_\theta = 1$ where $k_n > 0$, $k_n \to \infty$ and $\kappa > 1$; the extension to the case $d_\theta > 1$ is straightforward by applying the same argument element-wise to $\theta \in \Theta$.

Remark 4.4. The discretisation Θ_n (4.3) is chosen so that θ_0 is at most a $1/n^\kappa$ perturbation from some element of Θ_n for n large enough and, thus, θ_0 is in the convex hull of Θ_n as $k_n \to \infty$. Then, for any $\epsilon > 1/2$ such that $\kappa \geq \epsilon + 1/2$ and some $\underline{b}_n, \bar{b}_n \in B_n = \{1, \ldots, [k_n n^\kappa]\}$, where $[\cdot]$ is the integer part of $\cdot$, there exists $\tilde{\theta}_n \in \Theta_n$ such that $-\underline{b}_n/n^{\kappa-\epsilon} \leq n^\epsilon(\tilde{\theta}_n - \theta_0) \leq \bar{b}_n/n^{\kappa-\epsilon}$. For any $\delta \in \mathbb{R}$, for n large enough, there exists $\underline{b}_n = \underline{c}_n - n^{\kappa-\epsilon}\delta$ and $\bar{b}_n = \bar{c}_n + n^{\kappa-\epsilon}\delta$ for some bounded $\underline{c}_n, \bar{c}_n$ in the convex hull of B_n so that $-\underline{c}_n/n^{\kappa-\epsilon} \leq n^\epsilon(\tilde{\theta}_n - \theta_0) - \delta \leq \bar{c}_n/n^{\kappa-\epsilon}$. Thus, writing $\delta_n = n^\epsilon(\tilde{\theta}_n - \theta_0)$, since $\kappa - \epsilon > 1/2$ and $\underline{c}_n, \bar{c}_n$ are $O(1)$, $n^{1/2}(\tilde{\theta}_n - \theta_0) - \delta \to 0$, i.e., $\tilde{\theta}_n \in \Theta_n^0(\delta)$.

The argument of Remark 4.4 holds for any $\delta \in \mathbb{R}^{d_\theta}$ so that Θ_n includes sequences $\theta_n \in \Theta_n^0(\delta)$ for $1/2 < \epsilon \leq \kappa - 1/2$. Therefore, the feasible confidence region $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha)) = \{\theta \in \Theta_n : \hat{T}_n(\theta) \leq \chi^2_{d_g}(\alpha)\}$ with Θ_n defined above will contain sequences in $\Theta_n^0(\delta)$ with probability α as $n \to \infty$ under Assumptions 3.1 and 3.2. Importantly this discussion does not establish that $\theta_0 \in \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ with asymptotic probability α unless $\hat{T}_n(\theta_0)$ exists and $\theta_0 \in \Theta_n$ for n large enough but rather that $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ covers certain $o(n^{-1/2})$ perturbations to θ_0 with asymptotic probability α. Further modifications to $\hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ are therefore required for a feasible GAR-based confidence region that covers θ_0 with asymptotic level α.

Consider the set

$$\hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha)) = \{\theta \in \Theta : \; d_H(\theta, \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))) \leq Cn^{-v}\} \tag{4.4}$$

for some $C > 0$ and $v > 1/2$. Theorem 4.1 below shows that a (piecewise) continuous confidence region $\hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ where $\hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha)) \supseteq \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))$ contains θ_0 with asymptotic probability α.

Theorem 4.1. *Suppose Assumptions 3.1 and 3.2 are satisfied. Then, if $\kappa > v + 1/2$ and $v > 1/2$, $\lim_{n\to\infty} \mathcal{P}_0\{\theta_0 \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))\} = \alpha$.*

Remark 4.5. The feasible GAR-based confidence region $\hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ (4.4) can be constructed by forming Θ_n (4.3) for some large $k_n > 0$ and $\kappa > 1$ where $v < \kappa - 1/2$. Under the relatively mild assumptions needed for Theorem 3.1, Theorem 4.1 establishes that $\hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ includes θ_0 asymptotically with probability α with little restriction on the form of Ω or *a priori* knowledge of points of singularity.

Remark 4.6. The construction of the confidence set $\hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ essentially ignores the information $P_0' g(z, \theta_0) = 0$ w.p.1 associated with the singular moment variance matrix Ω; see the discussion in Section 3.3. As noted in Remark 4.2, however, the Andrews and Guggenberger (2018) SR-GAR confidence region requires $\theta_0 \in \Theta_n$ for correct asymptotic level if Ω is singular.

5 Conditions For $\mathcal{N}(\Omega) \subseteq \mathcal{N}(G')$

As noted above, $\Delta_c = \mathbb{R}^{d_\theta}$ when $\mathcal{N}(\Omega) \subseteq \mathcal{N}(G')$, i.e., $P_0'G\delta = 0$ for all $\delta \in \Delta_b$. Thus, $\hat{T}_n(\theta_n) \xrightarrow{d} \chi^2_{d_g}$ for all sequences $\theta_n \in \Theta_n^0(\delta)$, $\delta \in \Delta_b$, i.e., sequences θ_n such that $\hat{T}_n(\theta_n)$ exists w.p.a.1.

In this section, the moment function $g(z,\theta)$ is defined via a residual d_ρ-dimensional vector $\rho(z,\theta)$ and the conditional moment restriction

$$\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta)|x] = 0 \text{ a.s.}(x).$$

We distinguish between two types of moment function depending upon whether an initial estimation of $\mathbb{E}_{\mathcal{P}_0}[\partial\rho(z,\theta)/\partial\theta'|x]$ is required.

Let $D(x,\theta) = \mathbb{E}_{\mathcal{P}_0}[\partial\rho(z,\theta)/\partial\theta'|x]$ and $V(x,\theta) = \mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta)\rho(z,\theta)'|x]$. Then the optimal vector of instruments is $D(x,\theta)'V(x,\theta)^{-1}\rho(z,\theta)$; see, e.g., Newey (1993).

5.1 Case A. $\mathbb{E}_{\mathcal{P}_0}[\partial\rho(z,\theta)/\partial\theta'|x] = \partial\rho(\theta)/\partial\theta'$ a.s.(x)

Since θ_0 is unknown, the moment indicator vector is often formed in practice as

$$g(z,\theta) = \frac{\partial\rho(\theta)'}{\partial\theta}\rho(z,\theta),$$

i.e., as if the conditional moment variance matrix $V(x,\theta_0)$ is the identity matrix I_{d_ρ}. Here, $d_g = d_\theta$. Hence, $G = (\partial\rho(\theta_0)/\partial\theta')'(\partial\rho(\theta_0)/\partial\theta')$ and $\Omega = (\partial\rho(\theta_0)/\partial\theta')'\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)'](\partial\rho(\theta_0)/\partial\theta')$. If the unconditional variance matrix $\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)']$ of the residual vector $\rho(z,\theta_0)$ is non-singular then, for any $\beta \in \mathbb{R}^{d_\theta}$, $\Omega\beta = 0$ if and only if $(\partial\rho(\theta_0)/\partial\theta')\beta = 0$, i.e., $D(x,\theta_0)\beta = 0$. Hence, $G'\beta = 0$ if and only if $(\partial\rho(\theta_0)/\partial\theta')\beta = 0$. Therefore $\mathcal{N}(\Omega) = \mathcal{N}(G')$.

Proposition 5.1. *Let $g(z,\theta) = (\partial\rho(\theta)/\partial\theta')\rho(z,\theta_0)$. Then, if $\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)']$ is non-singular, $\mathcal{N}(\Omega) = \mathcal{N}(G')$.*

Remark 5.1. More generally, if $\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)']$ is singular, then $\mathcal{N}(G') \subseteq \mathcal{N}(\Omega)$. Proposition 5.1 is applicable to the nonlinear IV simultaneous equation model.

5.2 Case B. $\mathbb{E}_{\mathcal{P}_0}[\partial\rho(z,\theta)/\partial\theta'|x] \neq \partial\rho(\theta)/\partial\theta'$ a.s.(x)

When the condition $\mathbb{E}_{\mathcal{P}_0}[\partial\rho(z,\theta)/\partial\theta'|x] = \partial\rho(\theta)/\partial\theta'$ a.s.(x) fails to hold, it often the case in practice that the unconditional moment vector $g(z,\theta)$ is constructed using a d_ψ-vector of functions $\psi(x) = (\psi_1(x), ..., \psi_{d_\psi}(x))'$ of the instruments x, i.e.,

$$g(z,\theta) = \rho(z,\theta) \otimes \psi(x);$$

cf. *inter alia* Jorgenson and Laffont (1974). Here, $d_g = d_\rho d_\psi$. Hence, $G = \mathbb{E}_{\mathcal{P}_0}[\mathbb{E}_{\mathcal{P}_0}[\partial\rho(z,\theta_0)/\partial\theta'|x]\otimes \psi(x)]$ and $\Omega = \mathbb{E}_{\mathcal{P}_0}[\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)'|x]\otimes\psi(x)\psi(x)']$. We assume that $\psi(x)$ does not include any linearly redundant components, i.e., $\mathbb{E}_{\mathcal{P}_0}[\psi(x)\psi(x)']$ is non-singular. Hence, Ω is singular only if $\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)'|x]$ has deficient rank a.s.(x).

Define $\delta = (\delta_1',\ldots,\delta_{d_\rho}')'$ where $\delta_j \in \mathbb{R}^{d_\psi}$, $(j = 1,\ldots,d_\psi)$.

Proposition 5.2. *Let $g(z,\theta) = \rho(z,\theta)\otimes\psi(x)$. Then, if $\mathbb{E}_{\mathcal{P}_0}[\psi(x)\psi(x)']$ is non-singular, $\mathcal{N}(\Omega)\subseteq\mathcal{N}(G')$ if and only if $G'\delta = 0$ for all $\delta\in\mathbb{R}^{d_g}$ such that $(I_{d_\rho}\otimes\psi(x)')\delta\in\mathcal{N}(\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)'|x]\otimes I_{d_\psi})$ a.s.(x).*

Remark 5.2. If the residual function $\rho(z,\theta)$ is conditionally homoskedastic a.s.(x), i.e., $\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)'|x] = \mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)']$, then $\mathrm{rk}(\Omega) = r_\rho d_\psi$ where $r_\rho = \mathrm{rk}(\mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta_0)\rho(z,\theta_0)'])$.

6 Simulation Evidence

6.1 Preliminaries

Consider the bivariate linear IV regression model, cf. Example E.1 in Supplement E,

$$y_j = \theta_{0j}x_j + \varepsilon_j,\ (j = 1,2),$$

where $x_j = \pi w_1 + \eta_j$, $(j = 1,2)$, and $w = (w_1,w_2)'$ denotes the vector of instruments.

In all cases, the true value of the parameter vector θ_0 is given by $\theta_{01} = 1.0$, $\theta_{02} = 0.5$ and the instruments w_j, $(j = 1,2)$, are independent standard normal variates.

Since $\pi_1 = \pi_2 = \pi$, it follows from Example E.1 that $P_0'G\delta = -\pi(\delta_1-\delta_2)/\sqrt{2}$. We consider $\theta_n = \theta_0 + n^{-1}\delta_n$ where $\delta_n = (\delta_{1n},\delta_{2n})'$; cf. (3.2). By Theorem 3.2, in the common shock case of Example E.1, i.e., $\varepsilon_1 = \varepsilon_2$, $\hat{T}(\theta_n)/n \to \delta'G'P_0\Phi(\delta)^{-1}P_0'G\delta \propto \frac{\pi^2}{2}(\delta_1-\delta_2)^2$; here $\lim_{n\to\infty}\delta_{1n}-\delta_{2n} = \delta_1-\delta_2$.

6.2 Design

The structural innovations ε_1, ε_2 are described by

$$\begin{aligned}\varepsilon_1 &= \sqrt{\frac{1+\rho}{2}}\upsilon_1 + \sqrt{\frac{1-\rho}{2}}\upsilon_2\\ \varepsilon_2 &= \sqrt{\frac{1+\rho}{2}}\upsilon_1 - \sqrt{\frac{1-\rho}{2}}\upsilon_2.\end{aligned}$$

and $(\upsilon_1,\upsilon_2,\eta_1,\eta_2)'$ is distributed conditional on w as $N(0,\Xi)$ where

$$\Xi = \begin{pmatrix} 1.0 & 0.0 & 0.3 & 0.0\\ 0.0 & 1.0 & 0.5 & 0.0\\ 0.3 & 0.0 & 1.0 & 0.0\\ 0.0 & 0.5 & 0.0 & 1.0\end{pmatrix}$$

For sequences θ_n, we set $\delta_{1n} = \delta_1 = 1$ and $\delta_{2n} = 1 + \epsilon_n$ where $\epsilon_n = 10\exp(0.15j)/\exp(0.15J_n)$, $j \in \{0, \cdots, J_n\}$, $J_n \in \{90, 100, 110, 115\}$ corresponding to $n = 100$, 500, 1000 and 5000 respectively. This specification allows for sequences θ_n such that δ_{2n} deviates from δ_{1n} by ϵ_n ranging from $10/\exp(-0.15J_n)$ to 10 with smaller increments for larger sample sizes. For any such θ_n, the minimum of $\delta_2 - \delta_1$, i.e., the limit of $\delta_{1n} - \delta_{2n}$, is 9. This set of sequences θ_n includes $\epsilon_n = o(n^{-1/2})$. Hence, some sequences $\theta_n \in \Theta_n^0(\Delta_b \cap \Delta_c)$ are also covered so that the limiting distribution of $\hat{T}(\theta_n)$ is χ_3^2 by Theorem 3.1; see Section 3.1.

Sample sizes $n = 100$, 500, 1000 and 5000 are examined. All results are based on $1,000$ random draws.

6.3 Results

Figure 1 plots $\mathcal{P}_0\{\hat{\mathcal{C}}_n^0(\chi_{d_g}^2(0.90)\}$ (4.4) against $\delta_{2n} - \delta_{1n}$ on a logarithmic scale.

It is immediately apparent that, for all sample sizes n, sequences $\delta_{2n} - \delta_{1n}$, ρ and π, the coverage $\mathcal{P}_0\{\theta_0 \in \hat{\mathcal{C}}_n^0(\chi_{d_g}^2(0.90)\}$ is approximately 0.1 for $\delta_{2n} - \delta_{1n}$ close to zero, i.e., those sequences $\{\theta_n\}$ satisfying Theorem 3.1. Overall, test size is increasing in ρ and π with size approaching 0.1 for $\rho < 1.0$ and at a faster rate as n increases the further ρ is from 1.0. When $\rho = 1.0$, test size quickly approaches 1.0 as n increases with $\delta_{2n} - \delta_{1n}$ bounded away from zero. The GAR statistic $\hat{T}(\theta_n)$ is oversized for those sequences $\delta_{2n} - \delta_{1n} \to 9$ as π increases even if $\rho = 0.9$ when Ω is non-singular.

Note that the sequences θ_n are all $O(n^{-1})$ perturbations to θ_0. Figure 1 shows that the coverage $\mathcal{P}_0\{\theta_0 \in \hat{\mathcal{C}}_n^0(\chi_{d_g}^2(0.90)\}$ can be highly sensitive to very small perturbations to θ_0, being oversized for moment indicator functions with some elements with correlation at most 0.9. Figure 1 also highlights the necessity for Θ_n to be discretized sufficiently finely in order that, for asymptotically correct size, the feasible confidence region contains sequences θ_n converging to θ_0; see Section 3.

7 Summary and Conclusions

This paper derives the asymptotic properties of GAR-based confidence regions in the presence of a singular moment variance matrix. To do so, the inverse of the sample moment variance matrix $\hat{\Omega}_n(\theta)$ is expanded around points of singularity of Ω using a Laurent series expansion. This approach is new in the literature and does not presuppose the form of singularity of the moment variance matrix is known. The main results of this paper allow both the expected Jacobian G and moment variance matrix Ω to have arbitrary rank and form and provide a direct extension to those for the GAR statistic with nonsingular Ω in Stock and Wright (2000).

We devote attention to first order moment singularity. This is a sufficient condition that enables the Laurent series expansion to be established, which is a key step in showing that the GAR statistic exists asymptotically on a set of parameter sequences converging to the true parameter value θ_0. We show that a sufficient condition for the GAR statistic to converge in distribution to a $\chi_{d_g}^2$ variate on all such sequences is that the null space of the moment variance matrix Ω is a subset of that of the transposed expected Jacobian matrix G'. In the absence of this condition, the GAR statistic may be asymptotically unbounded on a subset of such sequences.

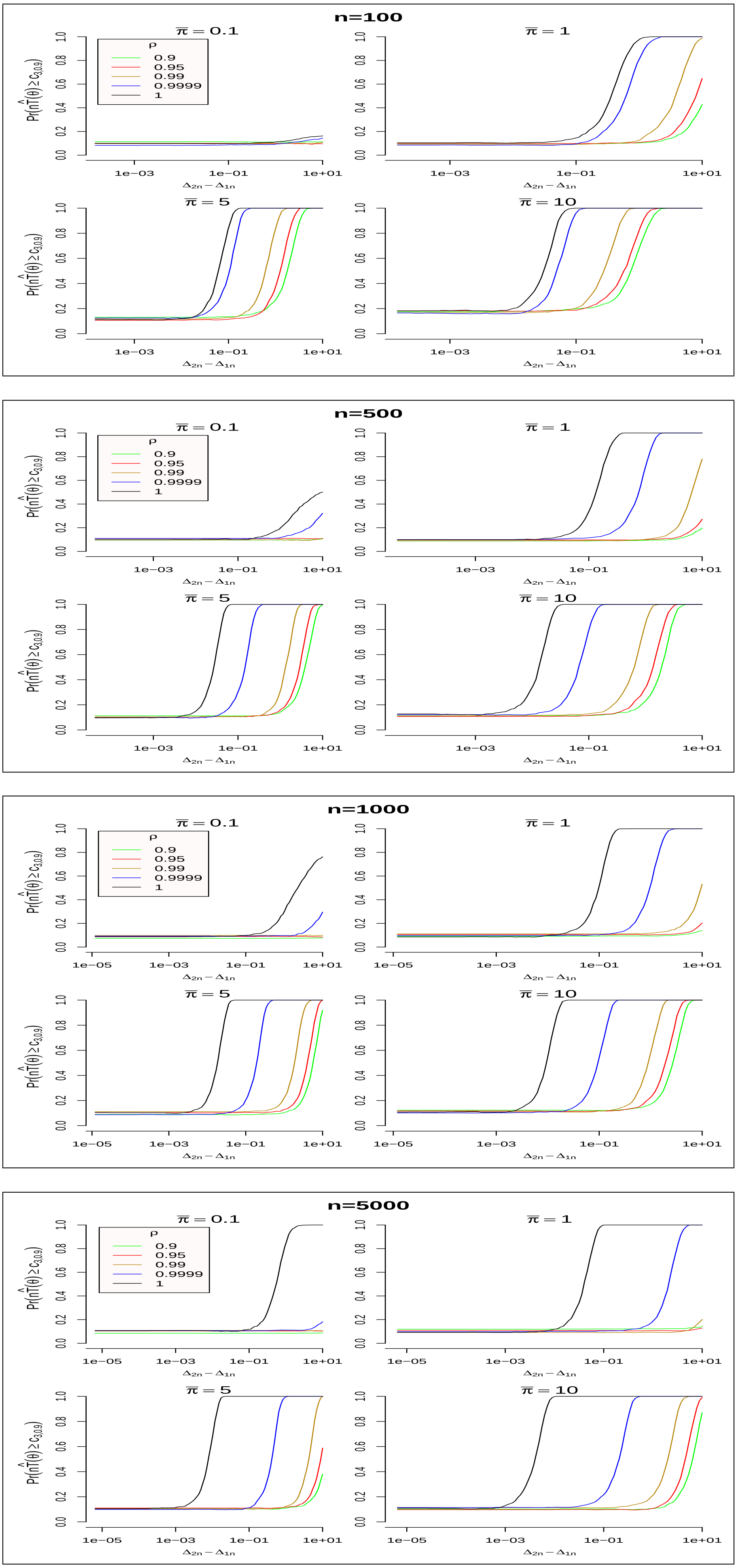

Figure 1: GAR rejection probabilities plotted against $\delta_{2n} - \delta_{1n}$ for $n = 100, 500, 1000, 5000$.

On the basis of these results, the paper details how to discretise appropriately the parameter space over which the non-rejection region of a GAR-based test is inverted to guarantee that parameter sequences for which the GAR statistic is asymptotically chi-square distributed are covered. A feasible GAR-based confidence region contains the value θ_0 with correct asymptotic size under relatively mild assumptions and requires no knowledge of points of singularity. Furthermore, GAR-based inference does not require any regularization or pre-testing, and so is less computationally burdensome than a regularization approach. Supplement E provides a number of examples of moment functions with singular variance matrix. A simulation study illustrates the results of this paper. Additional simulation evidence is provided in Supplement S.

Useful extensions of the results in this paper would be to the many weak moments case considered in Newey and Windmeijer (2009) and to accommodate subvector inference. Another avenue for future research is the extension of the results on generalised empirical likelihood based inference with weak identification of Guggenberger and Smith (2005), Guggenberger and Smith (2008) and Guggenberger et al. (2012) to allow for a singular moment variance matrix.

Appendix

The following auxiliary lemmas are established under Assumptions 3.1 and 3.2 and are used in the proofs of Theorems 3.1, 3.2 and 4.1 in Appendix B. Lemmas A.1 and A.2 are used to show Lemma A.3, which expands $\hat{\Omega}_n(\theta_n)^{-1}$ and $\tilde{\Omega}_n(\theta_n)^{-1}$ around points of singularity by a Laurent series expansions for sequences $\theta_n \in \Theta_n^0(\Delta_b)$, a key step in proving Theorem 3.1.

The argument θ is suppressed for expositional simplicity throughout the Appendices where there is no possibility of confusion.

Throughout the Appendices, C will denote a generic positive constant that may be different in different uses with CS, M and T the Cauchy-Schwartz, Markov and triangle inequalities respectively. In addition UWL and CLT refer to, respectively, a uniform weak law of large numbers and a central limit theorem for i.i.d. random variables.

Appendix A: Auxiliary Lemmas

Lemma A.1. *Under Assumption 3.1, then for any $\theta_n \in \Theta_n^0(\delta)$ for all $\delta \in \mathbb{R}^{d_\theta}$ **(a)** $n^{2\varepsilon} P_0' \hat{\Omega}_n(\theta_n) P_0 \xrightarrow{p} P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta \delta' G_i'] P_0$; **(b)** $n^{2\varepsilon} P_0' \tilde{\Omega}_n(\theta_n) P_0 \xrightarrow{p} P_0' (\mathbb{E}_{\mathcal{P}_0}[G_i \delta \delta' G_i'] - G \delta \delta' G') P_0$.*

Proof. **(a)**. By the mean value theorem

$$g_i(\theta_n) = g_i + n^{-\varepsilon} G_i(\theta_n^*) \delta_n$$

where θ_n^* is on the line segment joining θ_n and θ_0. Hence

$$\hat{\Omega}_n(\theta_n) = \hat{\Omega}_n + \frac{1}{n^{1+2\varepsilon}} \sum_{i=1}^n G_i(\theta_n^*) \delta_n \delta_n' G_i(\theta_n^*)' + \frac{1}{n^{1+\varepsilon}} \sum_{i=1}^n G_i(\theta_n^*) \delta_n g_i' + \frac{1}{n^{1+\varepsilon}} \sum_{i=1}^n g_i \delta_n' G_i(\theta_n^*)'.$$

Since $P_0'g_i = 0$ w.p.1

$$n^{2\varepsilon}P_0'\hat{\Omega}_n(\theta_n)P_0 = \frac{1}{n}\sum_{i=1}^{n}P_0'G_i(\theta_n^*)\delta_n\delta_n'G_i(\theta_n^*)'P_0.$$

Now

$$\frac{1}{n}\sum_{i=1}^{n}G_i(\theta_n^*)\delta_n\delta_n'G_i(\theta_n^*)' = \frac{1}{n}\sum_{i=1}^{n}G_i\delta\delta'G_i' + R_n.$$

where $R_n = \sum_{k=1}^{4}R_{kn}$ and

$$\begin{aligned}
R_{1n} &= \frac{1}{n}\sum_{i=1}^{n}(G_i(\theta_n^*) - G_i)\delta_n\delta_n'(G_i(\theta_n^*) - G_i)', \\
R_{2n} &= \frac{1}{n}\sum_{i=1}^{n}(G_i(\theta_n^*) - G_i)\delta_n\delta_n'G_i' = R_{3n}', \\
R_{4n} &= \frac{1}{n}\sum_{i=1}^{n}G_i(\delta_n\delta_n' - \delta\delta')G_i'.
\end{aligned}$$

By T and Assumption 3.1(d),

$$\begin{aligned}
\|R_{1n}\| &\leq \|\delta_n\|^2\frac{1}{n}\sum_{i=1}^{n}\|G_i(\theta_n^*) - G_i\|^2 \\
&\leq \|\delta_n\|^2 O_p(1)\|\theta_n - \theta_0\|^2 = O_p(n^{-2\varepsilon}) = o_p(1).
\end{aligned}$$

Similarly, by CS and UWL,

$$\begin{aligned}
\|R_{2n}\| &\leq \|\delta_n\|^2\frac{1}{n}\sum_{i=1}^{n}\|G_i(\theta_n^*) - G_i\|\|G_i\| \\
&\leq \|\delta_n\|^2(\frac{1}{n}\sum_{i=1}^{n}\|G_i(\theta_n^*) - G_i\|^2)^{1/2}(\frac{1}{n}\sum_{i=1}^{n}\|G_i\|^2)^{1/2} = O_p(n^{-\varepsilon}) = o_p(1).
\end{aligned}$$

Finally,

$$\begin{aligned}
\|R_{4n}\| &\leq \|\delta_n\delta_n' - \delta\delta'\|\frac{1}{n}\sum_{i=1}^{n}\|G_i\|^2 \\
&\leq O_p(n^{-1/2})O_p(1) = o_p(1).
\end{aligned}$$

Therefore,

$$n^{2\varepsilon}P_0'\hat{\Omega}_n(\theta_n)P_0 = \frac{1}{n}\sum_{i=1}^{n}P_0'G_i\delta\delta'G_i'P_0 + o_p(1)$$

and the conclusion follows, noting $\mathbb{E}_{\mathcal{P}_0}[\|G_i\delta\delta'G_i'\|] \leq \|\delta\|^2\mathbb{E}_{\mathcal{P}_0}[\|G_i\|^2] < \infty$.

(b). Recall $\tilde{\Omega}_n(\theta_n) = \hat{\Omega}_n(\theta_n) - \hat{g}_n(\theta_n)\hat{g}_n(\theta_n)'$. From part (a)

$$n^{\varepsilon}P_0'\hat{g}_n(\theta_n) = \frac{1}{n}\sum_{i=1}^{n}P_0'G_i(\theta_n^*)\delta_n.$$

Now

$$\frac{1}{n}\sum_{i=1}^{n} G_i(\theta_n^*)\delta_n = \frac{1}{n}\sum_{i=1}^{n} G_i\delta + R_n.$$

where $R_n = \sum_{k=1}^{2} R_{kn}$ and

$$\begin{aligned} R_{1n} &= \frac{1}{n}\sum_{i=1}^{n}(G_i(\theta_n^*) - G_i)\delta_n, \\ R_{2n} &= \frac{1}{n}\sum_{i=1}^{n} G_i(\delta_n - \delta). \end{aligned}$$

By similar arguments to those in part (a) $\|R_{1n}\| \leq O_p(n^{-\varepsilon})$ and $\|R_{2n}\| \leq O_p(n^{-1/2})$. Hence, by UWL, $n^{\varepsilon}P_0'\hat{g}_n(\theta_n) \xrightarrow{p} P_0'G\delta$ and the conclusion follows.■

Lemma A.2. *Under Assumption 3.1, then for any $\theta_n \in \Theta_n^0(\delta)$ for all $\delta \in \mathbb{R}^{d_\theta}$* ***(a)*** $n^{\varepsilon}P_0'\hat{\Omega}_n(\theta_n) \xrightarrow{p} P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']$; ***(b)*** $n^{\varepsilon}P_0'\tilde{\Omega}_n(\theta_n) \xrightarrow{p} P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']$.

PROOF. **(a)**. From the Proof of Lemma A.1(a), by the mean value theorem and since $P_0'g_i = 0$ w.p.1

$$n^{\varepsilon}P_0'\hat{\Omega}_n(\theta_n) = \frac{1}{n}\sum_{i=1}^{n} P_0'G_i(\theta_n^*)\delta_n\delta_n'G_i(\theta_n^*)' + \frac{1}{n}\sum_{i=1}^{n} P_0'G_i(\theta_n^*)\delta_n g_i'.$$

where θ_n^* is on the line segment joining θ_n and θ_0. Recall

$$\frac{1}{n}\sum_{i=1}^{n} G_i(\theta_n^*)\delta_n\delta_n'G_i(\theta_n^*)' \xrightarrow{p} \mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'].$$

Similarly to the Proof of Lemma A.1(b)

$$\frac{1}{n}\sum_{i=1}^{n} G_i(\theta_n^*)\delta_n g_i' = \frac{1}{n}\sum_{i=1}^{n} G_i\delta g_i' + R_n.$$

where $R_n = \sum_{k=1}^{2} R_{kn}$ and

$$\begin{aligned} R_{1n} &= \frac{1}{n}\sum_{i=1}^{n}(G_i(\theta_n^*) - G_i)\delta_n g_i', \\ R_{2n} &= \frac{1}{n}\sum_{i=1}^{n} G_i(\delta_n - \delta)g_i'. \end{aligned}$$

By similar arguments to those in the Proof of Lemma A.1

$$\begin{aligned} \|R_{1n}\| &\leq \|\delta_n\|\frac{1}{n}\sum_{i=1}^{n}\|G_i(\theta_n^*) - G_i\|\|g_i\| \\ &\leq \|\delta_n\|(\frac{1}{n}\sum_{i=1}^{n}\|G_i(\theta_n^*) - G_i\|^2)^{1/2}(\frac{1}{n}\sum_{i=1}^{n}\|g_i\|^2)^{1/2} = O_p(n^{-\varepsilon}) = o_p(1) \end{aligned}$$

and

$$\begin{aligned}\|R_{2n}\| &\leq \|\delta_n - \delta\| \frac{1}{n}\sum_{i=1}^{n}\|G_i\|\|g_i\| \\ &\leq O(n^{-1/2})(\frac{1}{n}\sum_{i=1}^{n}\|G_i\|^2)^{1/2}(\frac{1}{n}\sum_{i=1}^{n}\|g_i\|^2)^{1/2} = O_p(n^{-1/2}) = o_p(1).\end{aligned}$$

Hence, by UWL, $n^{\varepsilon}P_0'\hat{\Omega}_n(\theta_n) \xrightarrow{p} P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']$ giving the conclusion.

(b). Follows immediately from part (a) as $n^{\varepsilon}P_0'\hat{g}_n(\theta_n) \xrightarrow{p} P_0'G\delta$ from the Proof of Lemma A.1(b) and $\hat{g}_n(\theta_n) = o_p(1)$ by UWL.■

Central to the Proof of Lemma A.3 is a Laurent series expansion for $\hat{\Omega}_n(\theta_n)$ that relies on Avrachenkov et al. (2013, Theorems 2.10 and 2.11, pp.22-25). Let A, B be $m \times m$ matrices where $\text{rk}(A) = r_A < m$ and $B = O_p(1)$. Let $A(z) = A + zB$, $z \in \mathbb{R}$. Then, for all $z \to 0$, if $A(z)^{-1}$ exists in a neighbourhood of $z = 0$ and B satisfies $\det(V_0'BV_0) \neq 0$ w.p.a.1 where V_0 is a $m \times (m - r_A)$ matrix such that $AV_0 = 0$ w.p.a.1,

$$A(z)^{-1} = \frac{1}{z}V_0(V_0'BV_0)^{-1}V_0' + M_{V_0}(B)A^{-}\sum_{j=0}^{\infty}(-M_{V_0}(B)(zB)M_{V_0}(B)'A^{-})^{j}M_{V_0}(B)',$$

where A^{-} is the Moore-Penrose generalized inverse of A.

Let $\partial\hat{\Omega}_n = \hat{\Omega}_n(\theta_n) - \Omega$. Recall $M_{P_0}(\partial\hat{\Omega}_n) = I_{d_g} - P_0(P_0'\partial\hat{\Omega}_nP_0)^{-1}P_0'\partial\hat{\Omega}_n$.

Lemma A.3. *Let Assumptions 3.1, 3.2(a) and (b) be satisfied. Then, if $\bar{r}_{\Omega} > 0$, for all sequences $\theta_n \in \Theta_n^0(\Delta_b)$, apart from an $o_p(\|M_{P_0}(\partial\hat{\Omega}_n)\|^2)$ term,*
(a)

$$\begin{aligned}\hat{\Omega}_n(\theta_n)^{-1} &= P_0(P_0'\partial\hat{\Omega}_nP_0)^{-1}P_0' \\ &\quad + M_{P_0}(\partial\hat{\Omega}_n)\Omega^{-}(\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i\delta'G_i']P_0\Sigma(\delta)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'])\Omega^{-}M_{P_0}(\partial\hat{\Omega}_n)'\end{aligned}$$

where $\Sigma(\delta) = P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'] - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^{-}\mathbb{E}_{\mathcal{P}_0}[g_i\delta'G_i'])P_0$;
(b)

$$\begin{aligned}\tilde{\Omega}_n(\theta_n)^{-1} &= P_0(P_0'\partial\tilde{\Omega}_nP_0)^{-1}P_0' \\ &\quad + M_{P_0}(\partial\tilde{\Omega}_n)\Omega^{-}(\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i\delta'G_i']P_0\tilde{\Sigma}(\delta)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'])\Omega^{-}M_{P_0}(\partial\tilde{\Omega}_n)'\end{aligned}$$

where $\tilde{\Sigma}(\delta) = P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'] - G\delta\delta'G' - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^{-}\mathbb{E}_{P_0}[g_i\delta'G_i'])P_0$.

Proof. **(a)**. Set $z = n^{-2\varepsilon}$, $A = \Omega$, $B = n^{2\varepsilon}\partial\hat{\Omega}_n$, $V_0 = P_0$ and, thus, $A(z) = \hat{\Omega}_n(\theta_n)$. Now, using Lemma A.1(a),

$$V_0'BV_0 = n^{2\varepsilon}P_0'\partial\hat{\Omega}_nP_0 \xrightarrow{p} P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i']P_0$$

with $P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i']P_0$ full rank under Assumption 3.2(b). Hence, $\text{rk}(n^{2\varepsilon}P_0'\partial\hat{\Omega}_nP_0) = \bar{r}_{\Omega}$ w.p.a.1. Therefore, with these definitions, $\hat{\Omega}_n(\theta_n)$ satisfies the hypotheses of Avrachenkov et al. (2013, Theorems 2.10 and 2.11, pp.22-25) and, thereby, noting $zB = \partial\hat{\Omega}_n$, admits the Laurent series expansion

$$\hat{\Omega}_n(\theta_n)^{-1} = P_0(P_0'\partial\hat{\Omega}_n P_0)^{-1}P_0' + M_{P_0}(\partial\hat{\Omega}_n)\Omega^- \sum_{j=0}^{\infty}(-M_{P_0}(\partial\hat{\Omega}_n)\partial\hat{\Omega}_n M_{P_0}(\partial\hat{\Omega}_n)'\Omega^-)^j M_{P_0}(\partial\hat{\Omega}_n)'$$

for all sequences $\theta_n \in \Theta_n^0(\Delta_b)$.

Consider

$$\begin{aligned} M_{P_0}(\partial\hat{\Omega}_n)\partial\hat{\Omega}_n M_{P_0}(\partial\hat{\Omega}_n)'\Omega^- &= (\partial\hat{\Omega}_n - \partial\hat{\Omega}_n P_0(P_0'\partial\hat{\Omega}_n P_0)^{-1}P_0'\partial\hat{\Omega}_n)\Omega^- \\ &= (\partial\hat{\Omega}_n - \hat{\Omega}_n(\theta_n)P_0(P_0'\hat{\Omega}_n(\theta_n)P_0)^{-1}P_0'\hat{\Omega}_n(\theta_n))\Omega^-. \end{aligned}$$

By Lemmas A.1(a) and A.2(a), since $P_0'\partial\hat{\Omega}_n P_0$ is non-singular w.p.a.1,

$$n^{\varepsilon}\hat{\Omega}_n(\theta_n)P_0(n^{2\varepsilon}P_0'\hat{\Omega}_n(\theta_n)P_0)^{-1}n^{\varepsilon}P_0'\hat{\Omega}_n(\theta_n) \xrightarrow{p} \mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'].$$

By T, $\|\partial\hat{\Omega}_n\| \le \|\hat{\Omega}_n(\theta_n) - \hat{\Omega}_n\| + \|\hat{\Omega}_n - \Omega\| = o_p(1)$, using Assumption 3.1(d) and by UWL. Hence, noting $\Omega^- = O_p(1)$,

$$M_{P_0}(\partial\hat{\Omega}_n)\partial\hat{\Omega}_n M_{P_0}(\partial\hat{\Omega}_n)'\Omega^- \xrightarrow{p} -\mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^-.$$

Let $X = \mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0$ and $V = P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0$. Rewrite

$$\begin{aligned} \Omega^- \sum_{j=0}^{\infty}(-M_{P_0}(\partial\hat{\Omega}_n)\partial\hat{\Omega}_n M_{P_0}(\partial\hat{\Omega}_n)'\Omega^-)^j &= \Omega^- \sum_{j=0}^{\infty}(XV^{-1}X'\Omega^- + o_p(1))^j \\ &= \Omega^- + \Omega^- X(V^{-1}\sum_{j=0}^{\infty}(X'\Omega^- XV^{-1} + o_p(1))^j)X'\Omega^-. \end{aligned}$$

Repeated use of the Sherman-Morrison-Woodbury formula, see Horn and Johnson (2013, p.19), yields

$$V^{-1}\sum_{j=0}^{\infty}(X'\Omega^- XV^{-1})^j = (V - X'\Omega^- X)^{-1} = \Sigma^{-1}$$

since $\Sigma = V - X\Omega^- X'$ is full rank by Assumption 3.2(b).

Therefore, apart from an $o_p(\|M_{P_0}(\partial\hat{\Omega}_n)\|^2)$ term,

$$\hat{\Omega}_n(\theta_n)^{-1} = P_0(P_0'\partial\hat{\Omega}_n P_0)^{-1}P_0' + M_{P_0}(\partial\hat{\Omega}_n)\Omega^-(\Omega + X'\Sigma^{-1}X)\Omega^- M_{P_0}(\partial\hat{\Omega}_n)'$$

for all sequences $\theta_n \in \Theta_n^0(\Delta_b)$.

(b). Let $\partial\tilde{\Omega}_n = \tilde{\Omega}_n(\theta_n) - \Omega$. Recall $M_{P_0}(\partial\tilde{\Omega}_n) = I_{d_g} - P_0(P_0'\partial\tilde{\Omega}_n P_0)^{-1}P_0'\partial\tilde{\Omega}_n$.

The proof is the same as that of Lemma A.3(a) except that Lemmas A.1(b) and A.2(b) substitute for Lemmas A.1(a) and A.2(a).

Apart from setting $B = n^{2\varepsilon}\partial\tilde{\Omega}_n$ and, thus, $A(z) = \tilde{\Omega}_n(\theta_n)$, the other definitions remain the same as in the Proof of Lemma A.3. Using Lemma A.1(b),

$$V_0'BV_0 = n^{2\varepsilon}P_0'\partial\tilde{\Omega}_nP_0 \xrightarrow{p} P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'] - G\delta\delta'G')P_0$$

with $P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'] - G\delta\delta'G')P_0$ full rank under Assumption 3.2(b). Hence, $\mathrm{rk}(n^{2\varepsilon}P_0'\partial\tilde{\Omega}_nP_0) = \bar{r}_\Omega$ w.p.a.1. Therefore, $\tilde{\Omega}_n(\theta_n)$, noting $zB = \partial\tilde{\Omega}_n$, admits the Laurent series expansion

$$\tilde{\Omega}_n(\theta_n)^{-1} = P_0(P_0'\partial\tilde{\Omega}_nP_0)^{-1}P_0' + M_{P_0}(\partial\tilde{\Omega}_n)\Omega^-\sum_{j=0}^{\infty}(-M_{P_0}(\partial\tilde{\Omega}_n)\partial\tilde{\Omega}_nM_{P_0}(\partial\tilde{\Omega}_n)'\Omega^-)^jM_{P_0}(\partial\tilde{\Omega}_n)'$$

for all sequences $\theta_n \in \Theta_n^0(\Delta_b)$.

By the same arguments as in the Proof of Lemma A.3(a), by Lemma A.1(b) and A.2(b),

$$M_{P_0}(\partial\tilde{\Omega}_n)\partial\tilde{\Omega}_nM_{P_0}(\partial\tilde{\Omega}_n)'\Omega^- \xrightarrow{p} -\mathbb{E}_{\mathcal{P}_0}[g_i\delta'G_i']P_0(P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'] - G\delta\delta'G')P_0)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^-.$$

With $X = \mathbb{E}_{\mathcal{P}_0}[g_i\delta'G_i']P_0$ and now $V = P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'] - G\delta\delta'G')P_0$, again defining $\Sigma = V - X\Omega^-X'$,

$$\tilde{\Omega}_n(\theta_n)^{-1} = P_0(P_0'\partial\tilde{\Omega}_nP_0)^{-1}P_0' + M_{P_0}(\partial\tilde{\Omega}_n)\Omega^-(\Omega + X'\Sigma^{-1}X)\Omega^-M_{P_0}(\partial\tilde{\Omega}_n)'$$

apart from an $o_p(\|M_{P_0}(\partial\tilde{\Omega}_n)\|^2)$ term, for all sequences $\theta_n \in \Theta_n^0(\Delta_b)$. ■

Appendix B: Proofs of Theorems

Proof of Theorem 3.1. **(a)**. Using the mean value theorem, $\hat{g}_n(\theta_n) = \hat{g}_n + n^{-\varepsilon}\hat{G}_n(\theta_n^*)\delta_n$, where θ_n^* is on the line segment joining θ_n and θ_0. Hence,

$$\begin{aligned}\hat{T}_n(\theta_n) &= n\hat{g}_n'\hat{\Omega}_n(\theta_n)^{-1}\hat{g}_n + 2n^{1-\varepsilon}\hat{g}_n'\hat{\Omega}_n(\theta_n)^{-1}\hat{G}_n(\theta_n^*)\delta_n + n^{1-2\varepsilon}\delta_n'\hat{G}_n(\theta_n^*)'\hat{\Omega}_n(\theta_n)^{-1}\hat{G}_n(\theta_n^*)\delta_n \\ &= \hat{T}_n^1 + \hat{T}_n^2 + \hat{T}_n^3.\end{aligned}$$

Lemma A.3 is applied to each term $\hat{T}_n^j$, $(j = 1, 2, 3)$, in turn. Recall $\Sigma(\delta) = P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta'G_i'] - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^-\mathbb{E}_{\mathcal{P}_0}[g_i\delta'G_i'])P_0$.

First, noting $P_0'\hat{g}_n = 0$ w.p.1, $M_{P_0}(\partial\hat{\Omega}_n)'\hat{g}_n = \hat{g}_n$ and $o_p(\|M_{P_0}(\partial\hat{\Omega}_n)'\hat{g}_n\|) = o_p(\|\hat{g}_n\|) = o_p(n^{-1/2})$ w.p.1. Hence,

$$\hat{T}_n^1 = n\hat{g}_n'\Omega^-(\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i\delta'G_i']P_0\Sigma(\delta)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'])\Omega^-\hat{g}_n + o_p(1). \tag{B.1}$$

Secondly, by Assumption 3.1(d) and CLT $n^{1/2}(\hat{G}_n - G) = O_p(1)$,

$$\begin{aligned}n^{1/2}(\hat{G}_n(\theta_n^*) - G) &= n^{1/2}(\hat{G}_n(\theta_n^*) - \hat{G}_n) + n^{1/2}(\hat{G}_n - G) \\ &= n^{1/2}(\hat{G}_n - G) + o_p(n^{1/2-\varepsilon}).\end{aligned}$$

Hence, from Assumption 3.2(c), since $\delta_n = \delta + o(1)$,

$$\begin{aligned} n^{1/2} P_0' \hat{G}_n(\theta_n^*) \delta_n &= n^{1/2} P_0' \hat{G}_n(\theta_n^*)(\delta + o(1)) \\ &= n^{1/2} P_0' (\hat{G}_n(\theta_n^*) - G)(\delta + o(1)) \\ &= n^{1/2} P_0' (\hat{G}_n - G)\delta + o_p(n^{1/2-\varepsilon}). \end{aligned}$$

Then, by a similar argument to that in the Proof of Lemma A.3, using Lemmas A.1(a) and A.2(a),

$$\begin{aligned} n^{1/2-\varepsilon} M_{P_0}(\partial\hat{\Omega}_n)' \hat{G}_n(\theta_n^*)\delta_n &= (n^{-\varepsilon} I_{d_g} - n^{\varepsilon}\hat{\Omega}_n(\theta_n) P_0 (n^{2\varepsilon} P_0' \hat{\Omega}_n(\theta_n) P_0)^{-1} n^{1/2} P_0' \hat{G}_n(\theta_n^*)\delta_n \\ &= -\mathbb{E}_{\mathcal{P}_0}[g_i \delta' G_i'] P_0 (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta\delta' G_i'] P_0)^{-1} n^{1/2} P_0' (\hat{G}_n - G)\delta + o_p(1). \end{aligned}$$

Therefore,

$$\begin{aligned} \hat{T}_n^2 &= -2n^{1/2} \hat{g}_n' \Omega^- (\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i \delta' G_i'] P_0 \Sigma(\delta)^{-1} P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'])\Omega^- \\ &\quad \times \mathbb{E}_{\mathcal{P}_0}[g_i \delta' G_i'] P_0 (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta\delta' G_i'] P_0)^{-1} n^{1/2} P_0' (\hat{G}_n - G)\delta + o_p(1) \\ &= -2n^{1/2} \hat{g}_n' \Omega^- \mathbb{E}_{\mathcal{P}_0}[g_i \delta' G_i'] P_0 \Sigma(\delta)^{-1} n^{1/2} P_0' (\hat{G}_n - G)\delta + o_p(1) \end{aligned} \tag{B.2}$$

noting $\Sigma(\delta)^{-1} = (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta\delta' G_i'] P_0)^{-1} \sum_{j=0}^{\infty} (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'])\Omega^- \mathbb{E}_{\mathcal{P}_0}[g_i \delta' G_i'])P_0 (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta\delta' G_i'] P_0)^{-1})^j$.

Finally, similarly,

$$\begin{aligned} \hat{T}_n^3 &= n^{1/2} \delta' (\hat{G}_n - G)' P_0 (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta\delta' G_i'] P_0)^{-1} n^{1/2} P_0' (\hat{G}_n - G)\delta \\ &\quad + n^{1/2} \delta' (\hat{G}_n - G)' P_0 (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta\delta' G_i'] P_0)^{-1} P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \\ &\quad \times \Omega^- (\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i \delta' G_i'] P_0 \Sigma(\delta)^{-1} P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'])\Omega^- \\ &\quad \times \mathbb{E}_{\mathcal{P}_0}[g_i \delta' G_i'] P_0 (P_0' \mathbb{E}_{\mathcal{P}_0}[G_i \delta\delta' G_i'] P_0)^{-1} n^{1/2} P_0' (\hat{G}_n - G)\delta \\ &\quad + o_p(1) \\ &= n^{1/2} \delta' (\hat{G}_n - G)' P_0 \Sigma(\delta)^{-1} n^{1/2} P_0' (\hat{G}_n - G)\delta + o_p(1). \end{aligned} \tag{B.3}$$

Therefore, combining eqs. (B.1), (B.2) and (B.3), up to an $o_p(1)$ term,

$$\begin{aligned} \hat{T}_n(\theta_n) &= n \hat{g}_n' \Omega^- \hat{g}_n \\ &\quad + n(P_0'((\hat{G}_n - G)\delta - \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \Omega^- \hat{g}_n))' \Sigma(\delta)^{-1} (P_0'((\hat{G}_n - G)\delta - \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \Omega^- \hat{g}_n)). \end{aligned}$$

Now, since $n^{1/2}\hat{g}_n \xrightarrow{d} N(0, \Omega)$ by CLT, $n\hat{g}_n' \Omega^- \hat{g}_n \xrightarrow{d} \chi^2_{r_\Omega}$. Similarly,

$$n(P_0'((\hat{G}_n - G)\delta - \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \Omega^- \hat{g}_n))' \Sigma(\delta)^{-1} (P_0'((\hat{G}_n - G)\delta - \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \Omega^- \hat{g}_n)) \xrightarrow{d} \chi^2_{\bar{r}_\Omega}$$

as, by CLT, $n^{1/2}(P_0'((\hat{G}_n - G)\delta - \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \Omega^- \hat{g}_n)) \xrightarrow{d} N(0, \Sigma(\delta))$. Theorem 3.1(a) is then immediate since $P_0'((G_i - G)\delta - \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \Omega^- g_i)$ and $\Omega^- g_i$ are uncorrelated, i.e., $\mathbb{E}_{\mathcal{P}_0}[P_0'((G_i - G)\delta - \mathbb{E}_{\mathcal{P}_0}[G_i \delta g_i'] \Omega^- g_i) g_i' \Omega^-] = 0$ from the definition of a Moore-Penrose generalized inverse, i.e., $\Omega^- \Omega \Omega^- = \Omega^-$.

(b) The proof, being almost identical to that of part (a), is omitted.■

Proof of Theorem 3.2. **(a)** First, $\hat{G}_n(\theta_n^*) = (\hat{G}_n(\theta_n^*) - \hat{G}_n) + \hat{G}_n = G + o_p(1)$, noting

$\hat{G}_n(\theta_n^*) - \hat{G}_n = O_p(n^{-\varepsilon})$ and $\hat{G}_n - G = O_p(n^{-1/2})$ from Assumption 3.1(d) and CLT. Then, since Assumption 3.2(c) no longer holds, i.e., $P_0' G\delta \neq 0$,

$$P_0'\hat{G}_n(\theta_n^*)\delta_n = P_0' G\delta + o_p(1).$$

Secondly, from the Proof of Theorem 3.1(a), from eq. (B.1), by CLT,

$$\begin{aligned}\hat{T}_n^1/n &= \hat{g}_n'\hat{\Omega}_n(\theta_n)^{-1}\hat{g}_n \\ &= \hat{g}_n'\Omega^-(\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0\Sigma(\delta)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'])\Omega^-\hat{g}_n + o_p(1) \\ &= o_p(1). \end{aligned} \tag{B.4}$$

Next, from Lemmas A.1(a) and A.2(a),

$$\begin{aligned} n^{-\varepsilon}M_{P_0}(\partial\hat{\Omega}_n)'\hat{G}_n(\theta_n^*)\delta_n &= (n^{-\varepsilon}I_{d_g} - n^{\varepsilon}\hat{\Omega}_n(\theta_n)P_0(n^{2\varepsilon}P_0'\hat{\Omega}_n(\theta_n)P_0)^{-1}P_0'\hat{G}_n(\theta_n^*))\delta_n \\ &= -\mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0)^{-1}P_0' G\delta + o_p(1). \end{aligned}$$

Hence,

$$\begin{aligned}\hat{T}_n^2/n &= 2n^{-\varepsilon}\hat{g}_n'\hat{\Omega}_n(\theta_n)^{-1}\hat{G}_n(\theta_n^*)\delta_n \\ &= -2\hat{g}_n'\Omega^-(\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0\Sigma(\delta)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'])\Omega^- \\ &\quad \times\mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0)^{-1}P_0' G\delta + o_p(1) \\ &= -2\hat{g}_n'\Omega^-\mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0\Sigma(\delta)^{-1}P_0' G\delta + o_p(1) \\ &= o_p(1). \end{aligned} \tag{B.5}$$

Finally,

$$\begin{aligned}\hat{T}_n^3/n &= n^{-2\varepsilon}\delta_n'\hat{G}_n(\theta_n^*)'\hat{\Omega}_n(\theta_n)^{-1}\hat{G}_n(\theta_n^*)\delta_n \\ &= \delta' G' P_0(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0)^{-1}P_0' G\delta \\ &\quad +\delta' G' P_0(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'] \\ &\quad \times\Omega^-(\Omega + \mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0\Sigma(\delta)^{-1}P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i'])\Omega^- \\ &\quad \times\mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i']P_0(P_0'\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i']P_0)^{-1}P_0'(\hat{G}_n - G)\delta \\ &\quad +o_p(1) \\ &= \delta' G' P_0\Sigma(\delta)^{-1}P_0' G\delta + o_p(1). \end{aligned} \tag{B.6}$$

Combining eqs. (B.4), (B.5) and (B.6), gives the required result.

(b). The only alteration necessary to the Proof of Theorem 4.1(a) is the substitution of $\tilde{\Sigma}(\delta) = P_0'(\mathbb{E}_{\mathcal{P}_0}[G_i\delta\delta' G_i'] - G\delta\delta' G' - \mathbb{E}_{\mathcal{P}_0}[G_i\delta g_i']\Omega^-\mathbb{E}_{\mathcal{P}_0}[g_i\delta' G_i'])P_0$ for $\Sigma(\delta)$ in the Proof of Theorem 3.1(a).■

Proof of Theorem 4.1. The proof demonstrates that the events $\{\theta_n \in \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))\}$ and

$\{\theta_0 \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))\}$ are equivalent for large enough n, for any θ_0, i.e.,

$$\lim_{n\to\infty} \mathcal{P}_0\{\theta_0 \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))\} = \lim_{n\to\infty} \mathcal{P}_0\{\theta_n \in \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))\} = \alpha$$

where the second equality follows by Theorem 3.1 for $\theta_n \in \Theta_n^0(\Delta_b \cap \Delta_c)$.

By definition $\{\theta_n \in \hat{\mathcal{C}}_n(\chi^2_{d_g}(\alpha))\}$ implies $\{\theta_n \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))\}$; see (4.4). Consider $\theta_n \in \Theta_n$ where Θ_n is defined in (4.3) with $\varepsilon = \kappa - 1/2$. Hence, as $\kappa > v + 1/2$ and $v > 1/2$ by hypothesis, $\kappa > 1$ and thus $\varepsilon > v > 1/2$. Therefore, because, by definition, $\|\theta_n - \theta_0\| \leq n^{-\varepsilon}\|\delta_n\| \leq Cn^{-\varepsilon}$ and $\|\theta - \theta_n\| \leq Cn^{-v}$ for all $\theta \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$, since $\varepsilon > v$, $\theta_0 \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ for all n large enough.

To show the reverse, first note that $\theta_0 \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ implies $\hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ contains θ_0 and all $\theta \in \Theta$ such that $d_H(\theta, \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))) \leq Cn^{-v}$. Let B_n denote a ball of radius Cn^{-v} centred at θ_0. Hence B_n includes $\theta \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$ such that $\|\theta - \theta_0\| \leq Cn^{-v}$. Therefore, since $\varepsilon > v$, for sufficiently large n, there exists a sequence $\theta_n = \theta_0 + n^{-\varepsilon}\delta_n$ such that $\theta_n \in \hat{\mathcal{C}}_n^0(\chi^2_{d_g}(\alpha))$.■

References

Anderson, T. W. and Rubin, H. (1949), 'Estimators of the parameters of a single equation in a complete set of stochastic equations', *Annals of Mathematical Statistics* **21**, 570–582.

Andrews, D. W. K. and Cheng, X. (2012), 'Estimation and inference with weak, semi-strong, and strong identification', *Econometrica* **80**, 2151–2211.

Andrews, D. W. K. and Cheng, X. (2013), 'Maximum likelihood estimation and uniform inference with sporadic identification failure', *Journal of Econometrics* **173**, 36–56.

Andrews, D. W. K. and Cheng, X. (2014), 'GMM estimation and uniform subvector inference with possible identification failure', *Econometric Theory* **30**, 287–333.

Andrews, D. W. K. and Guggenberger, P. (2015), 'Identification- and singularity-robust inference for moment condition models', Cowles Foundation Discussion Paper No. 1978R.

Andrews, D. W. K. and Marmer, V. (2008), 'Exactly distribution-free inference in instrumental variables regression with possibly weak instruments', *Journal of Econometrics* **142**, 183–200.

Andrews, D. W. K., Moreira, M. J. and Stock, J. H. (2007), 'Performance of conditional Wald tests in IV regression with weak instruments', *Journal of Econometrics* **139**, 116–132.

Avrachenkov, K. E., Filar, J. A. and Howett, P. G. (2013), *Analytic Perturbation Theory and its Applications*, SIAM: Philadeplhia PA.

Bottai, M. (2003), 'Confidence regions when the Fisher information is zero', *Biometrika* **90**, 73–84.

Cheng, X. (2014), 'Uniform inference in nonlinear models with mixed identification strength', PIER Working Paper Archive 14-018.

Chernozhukov, V., Hansen, C. and Jansson, M. (2009), 'Admissible invariant similar tests for instrumental variables regression', *Econometric Theory* **25**, 806–818.

Dufour, J. M. and Valéry, P. (2016), 'Wald-type tests when rank conditions fail: a smooth regularization approach', Unpublished Working Paper.

Grant, N. L. and Smith, R. J. (2018), 'Supplementary material to "Generalized Anderson-Rubin statistic based inference in the presence of a singular moment variance matrix" '.

Guggenberger, P., Kleibergen, F. R., Mavroeidis, S. and Chen, L. (2012), 'On the asymptotic sizes of subset Anderson-Rubin and Lagrange Multiplier tests in linear instrumental variables regression', *Econometrica* **80**, 2649–2666.

Guggenberger, P., Ramalho, J. S., and Smith, R. J. (2012), 'GEL statistics under weak identification', *Journal of Econometrics* **170**, 331–349.

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

Guggenberger, P., and Smith, R. J. (2008), 'Generalized empirical likelihood tests in time series models with potential identification failure', *Journal of Econometrics* **142**, 134–161.

Hansen, L. P. and Scheinkman, J. A. (1995), 'Back to the future: generating moment implications for continuous-time Markov processes', *Econometrica* **63**, 767–804.

Horn, R. A. and Johnson, C. R. (2013), *Matrix Analysis*, 2nd. Ed., Cambridge University Press: Cambridge.

Jagannathan, R., Skoulakis, G. and Wang, G. (2002), 'Generalized method of moments: Applications in finance', *Journal of Business and Economic Statistics* **20**, 470–481.

Jorgenson, D. W. and Laffont, J. (1974), 'Efficient estimation of nonlinear simultaneous equations with additive disturbances', *Annals of Economic and Social Measurement* **3**, 615–640.

Kleibergen, F. R. (2002), 'Pivotal statistics for testing structural parameters in instrumental variables regression', *Econometrica* **70**, 1781–1803.

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

Kleibergen, F. R. and Mavroeidis, S. (2009), 'Weak instrument robust tests in GMM and the New Keynesian Phillips curve', *Journal of Business and Economic Statistics* **27**, 293–311.

Magnusson, L. M. (2010), 'Inference in limited dependent variable models robust to weak identification', *Econometrics Journal* **13**, 56–79.

Mikusheva, A. (2010), 'Robust confidence sets in the presence of weak instruments', *Journal of Econometrics* **157**, 236–247.

Moreira, M. J. (2003), 'A conditional likelihood ratio test for structural models', *Econometrica* **71**, 1027–1048.

Newey, W. K. (1993), 'Efficient estimation of models with conditional moment restrictions', *Handbook of Statistics* Vol. 11. Elsevier: Amsterdam.

NEWEY, W. K. AND MCFADDEN, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of Econometrics* Vol. 4, 2111–2245. North Holland: New York.

NEWEY, W.K. AND WINDMEIJER, F. (2009), 'Generalized method of moments with many weak moment conditions', *Econometrica* **77**, 687–719.

PEÑARANDA, F. AND SENTANA, E. (2012), 'Spanning tests in return and stochastic discount factor mean-variance frontiers: A unifying approach', *Journal of Econometrics* **170**, 303–324.

ROTNITZKY, A., COX, D. R., BOTTAI, M. AND ROBINS, J. (2000), 'Likelihood-based inference with singular information matrix', *Bernoulli* **6**, 243–284.

STOCK, J. H. AND WRIGHT, J. (2000), ' GMM with weak identification', *Econometrica* **68**, 1055–1096.

Supplement E: Examples for 'Generalised Anderson-Rubin Statistic Based Inference In The Presence Of A Singular Moment Variance Matrix'

Nicky L. Grant[†]
University of Manchester
nicky.grant@manchester.ac.uk

Richard J Smith[‡]
cemmap, U.C.L and I.F.S.
University of Cambridge
University of Melbourne
ONS Economic Statistics Centre of Excellence
rjs27@econ.cam.ac.uk

This Draft: January 2019

Example E.1 Bivariate Linear IV Regression Model with a Common Shock

To provide an initial, albeit stylised, illustration, consider the bivariate linear regression model with a common shock component ε

$$y_j = \theta_{0j} x_j + \varepsilon, \ (j = 1, 2).$$

Here $\theta_0 = (\theta_{01}, \theta_{02})'$ with $d_\theta = 2$. The 2×1 covariate vector $x = (x_1, x_2)'$ is generated according to $x_j = \pi_j w_1 + \eta_j$ where $\mathbb{E}_{\mathcal{P}_0}[\eta_j|w] = 0$, $\mathbb{E}_{\mathcal{P}_0}[\eta_j^2|w] = 1$, $(j = 1, 2)$, and $\mathbb{E}_{\mathcal{P}_0}[\eta_1\eta_2|w] = 0$. Here $w = (w_1, w_2)'$ denotes a 2×1 vector of instruments.

In this example y_j, $(j = 1, 2)$, are both subject to the common shock ε. It is assumed for simplicity that the instruments wj, $(j = 1, 2)$, are independent standard normal variates, $\mathbb{E}_{\mathcal{P}_0}[\varepsilon|w] = 0$, $\mathbb{E}_{\mathcal{P}_0}[\varepsilon^2|w] = 1$ and $\mathbb{E}_{\mathcal{P}_0}[\varepsilon\eta_j|w] = \rho_{\varepsilon\eta}$ where $0 \leq \rho_{\varepsilon\eta}^2 \leq 1/2$ which follows since $\det(\mathrm{Var}_{\mathcal{P}_0}(\varepsilon, \eta_1, \eta_2)) = 1 - 2\rho_{\varepsilon\eta}^2 > 0$.

[*]Address for correspondence: N.L. Grant, Arthur Lewis Building, Department of Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom.

[†]Arthur Lewis Building, Department of Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom.

[‡]Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK.

Define $\varepsilon_j(\theta) = y_j - \theta_j x_j$, $j = 1, 2$, and $\theta = (\theta_1, \theta_2)'$. The moment indicator function $g(z, \theta)$ comprises the elements

$$\begin{aligned} g_1(z,\theta) &= \varepsilon_1(\theta) w_1, \\ g_2(z,\theta) &= \varepsilon_2(\theta) w. \end{aligned}$$

Hence, $d_g = 2$. Moreover, $g_1(\theta)$ and the first element of $g_2(\theta)$ are perfectly correlated at $\theta = \theta_0$.

In terms of the notation of Section 3

$$\Omega = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad G = -\begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \\ 0 & 0 \end{pmatrix}.$$

Hence,

$$P_0' G \delta = -\pi_1 \frac{\Delta_1}{\sqrt{2}} + \pi_2 \frac{\Delta_2}{\sqrt{2}}$$

where $\delta = (\delta_1, \delta_2)'$. Therefore,
$Delta_c = \{\delta \in \mathbb{R}^2 : -\pi_1 \delta_1 + \pi_2 \delta_2 = 0\}$ and $\Delta_c = \mathbb{R}^2$ if and only if $\pi_1 = \pi_2 = 0$, i.e., $P_0' G = 0'$.

To derive those $\delta \in \Delta_c$ that also belong to Δ_b, first note that $P_0' G_i \delta = P_0'(G_i - G)\delta = \frac{1}{\sqrt{2}}(-\eta_{1i} w_{1i} \delta_1 + \eta_{2i} w_{1i} \delta_2)$ for $\delta \in \Delta_c$. Thus,

$$\begin{aligned} \mathrm{Var}(P_0' G_i \delta) &= \Upsilon(\delta) \\ &= \frac{1}{2} \mathbb{E}_{\mathcal{P}_0}[w_1^2(\delta_1^2 \eta_1^2 + \delta_2^2 \eta_2^2 - 2\delta_1 \delta_2 \eta_1 \eta_2)] \\ &= \frac{1}{2}(\delta_1^2 + \delta_2^2) > 0 \text{if } \delta_1 \neq 0 \text{ or } \delta_2 \neq 0 \text{ on } \Delta_c. \end{aligned}$$

Hence, $\Delta_b \cap \Delta_c = \{\delta \in \mathbb{R}^2 : -\pi_1 \Delta_1 + \pi_2 \Delta_2 = 0,\ \delta_1 \neq 0 \text{ or } \delta_2 \neq 0\} \neq \emptyset$.

We now establish that $\mathrm{rk}(\Phi(\delta)) = \bar{r}_\Omega$ on $\Delta_b \cap \Delta_c$ where $\Phi(\delta) = \Upsilon(\delta) - \Psi(\delta)\Omega_+ \Psi(\delta)'$ is defined in **eq.?**.

This result follows from Assumption 3.2(b), i.e., $\Delta_b \neq \emptyset$. Recall $\Theta_n^0(\Delta_b \cap \Delta_c)$ is the set of sequences on which Theorem 3.1 holds. We could establish this result more generally on $\Delta_b = \{\delta \in \mathbb{R}^2 :\ \delta_1 \neq 0 \text{ and/or } \delta_2 \neq 0\}$ where $\Delta_b \cap \Delta_c \subseteq \Delta_b$, but this is omitted for brevity.

Define $K_i(\delta) = P_0' G_i \delta g_i' P_+ \Lambda_+^{-1/2}$. Then $\Psi(\delta)\Omega_+ \Psi(\Delta)' = \mathbb{E}_{\mathcal{P}_0}[K_i(\delta)] \mathbb{E}_{\mathcal{P}_0}[K_i(\delta)']$. Noting

$$P_+ = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1 \end{pmatrix}, \quad \Lambda_+ = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},$$

$g_i' P_+ \Lambda_+^{-1/2} = \varepsilon_i(w_{1i}, w_{2i})$. Hence, since $P_0' G_i \delta = \frac{1}{sqrt2} w_{1i}(x_{2i}\delta_2 - x_{1i}\delta_1)$, $\mathbb{E}_{\mathcal{P}_0}[K_i(\delta)] = \frac{1}{\sqrt{2}}(\rho_{\varepsilon\eta}(\delta_2 -$

$\delta_1), 0)$, and

$$\begin{aligned}\Phi(\Delta) &= \frac{1}{2}(\delta_1^2 + \delta_2^2 - \rho_{\varepsilon\eta}^2(\delta_1 - \delta_2)^2) \\ &\geq \frac{1}{2}(\delta_1^2 + \delta_2^2 - \frac{1}{2}(\delta_1 - \delta_2)^2) \\ &> 0 \text{ on } \Delta_b \cap \Delta_c\end{aligned}$$

as $\rho_{\varepsilon\eta}^2 \leq 1/2$ and $\delta_1^2 + \delta_2^2 - \frac{1}{2}(\delta_1 - \delta_2)^2 = \frac{1}{2}(\delta_1 + \delta_2)^2 > 0$ unless $\delta_1 = \delta_2 = 0$.

Example E.2 Nonlinear Regression Model with First Order Moment Singularity

Consider the following nonlinear regression model

$$y = \alpha_0 + \pi_0(1 + \kappa_0 x)^{-1} + \varepsilon,\ x \geq 0, \mathbb{E}_{\mathcal{P}_0}[\varepsilon|x] = 0.$$

Here the parameter vector of interest is $\theta_0 = (\alpha_0, \pi_0, \kappa_0)'$, i.e., $d_\theta = 3$, and the moment indicator vector is defined by

$$g(z, \theta) = \varepsilon(\theta)(x, (1 + \kappa x)^{-1}, -\pi x(1 + \kappa x)^{-2})',$$

where $\varepsilon(\theta) = y - \alpha - \pi(1 + \kappa x)^{-1}$ and $\theta = (\alpha, \pi, \kappa)'$, i.e., $d_g = 3$.

There are two cases when the moment indicator variance matrix Ω has deficient rank 2. First, if $\pi_0 = 0$, then κ_0 is unidentified. Moment functions with this form of singular variance matrix have been studied previously in Andrews and Cheng (2012, 2013, 2014) although they consider, more generally, weak and semi-strong identification in which a particular transformation of the moment variance matrix is full rank; see Andrews and Cheng (2012, Assumption D2, p.25). Secondly, if $\kappa_0 = 0$, then the first and third element of the moment indicator vector are perfectly correlated. In this case, since $P_0' g(z, \theta_0) = 0$, $P_0 = \frac{1}{\sqrt{1+\pi_0^2}}(\pi_0, 0, 1)'$ and, thus,

$$P_0' G_i \delta = \frac{\varepsilon_i}{\sqrt{1 + \pi_0^2}}(-x_i \delta_2 + 2\pi_0 x_i^2 \delta_3).$$

Hence,

$$\Delta_b = \{\delta \in \mathbb{R}^3 :\ \delta_2 \neq 0 \text{ or } \delta_3 \neq 0\} \neq \emptyset.$$

where $\delta = (\delta_1, \delta_2, \delta_3)'$. Therefore $\mathrm{Var}_{\mathcal{P}_0}(P_0' G_i \delta) > 0$ on Δ_b since $P_0' G = 0$ noting $\mathbb{E}_{\mathcal{P}_0}[\varepsilon_i | x_i] = 0$. Consequently, $P_0' G \delta = 0$ on $\Delta_c = \mathbb{R}^3$ and, thus, $\Delta_b \cap \Delta_c = \Delta_b \neq \emptyset$.

Example E.3 Interest Rate Dynamics

Suppose that the interest rate r is generated by the process

$$r - r_{-1} = a_0(b_0 - r_{-1}) + \varepsilon \sigma_0 r_{-1}^{\gamma_0},$$

where ε is stationary and r_{-1} denotes the first lag of r. Here, $\theta_0 = (a_0, b_0, \gamma_0, \sigma_0)'$ and $z = (r, r_{-1})'$.

Consider the moment indicator vector function

$$g(z, \theta) = \begin{pmatrix} a(b-r)r^{-2\gamma} - \gamma\sigma^2 r^{-1} \\ a(b-r)r^{-2\gamma+1} - (\gamma - \frac{1}{2})\sigma^2 \\ (b-r)r^{-a} - \frac{1}{2}\sigma^2 r^{2\gamma-a-1} \\ a(b-r)r^{-\sigma} - \frac{1}{2}\sigma^3 r^{2\gamma-\sigma-1} \end{pmatrix},$$

where $\theta = (a, b, \gamma, \sigma)'$, which has zero mean at $\theta = \theta_0$; see Jagannathan et al. (2002, p.479).

If $\sigma_0 = a_0$, $\gamma_0 = 1/2(a_0 + 1)$ and/or $\gamma_0 = 1/2(\sigma_0 + 1)$, then Ω is singular. Consider the case when $\sigma_0 = a_0$, then the third and fourth elements of $g(z, \theta_0)$ are perfectly correlated, assuming for simplicity that $\gamma_0 \neq 1/2(a_0 + 1)$. Hence, $\bar{r}_\Omega = 1$ and $P_0 = \frac{1}{\sqrt{2}}(0, 0, 1, -1)'$. Without loss of generality, set $\sigma_0 = a_0 = b_0 = 1$ and $\gamma_0 = 0$, i.e., $r = 1 + \varepsilon$. Write $\delta = (\delta_a, \delta_b, \delta_\gamma, \delta_\sigma)'$ and set $\delta_b = \delta_\gamma = 0$ for simplicity. Then

$$\begin{aligned} P_0' G_i \delta &= \left(\frac{1}{2} r^{-2} - (1-r)r^{-1}\log r - \frac{1}{2} r^{-2}\log r\right)\frac{\delta_\sigma}{\sqrt{2}} \\ &\quad + \left(-(1-r)r^{-1}(1 + \log r) + \frac{1}{2} r^{-2}\log r\right)\frac{\delta_a}{\sqrt{2}}. \end{aligned}$$

Hence, $\text{Var}_{\mathcal{P}_0}(P_0' G_i \delta) > 0$ on $\Delta_b \cap \Delta_c$ if $\delta_a \neq 0$ and/or $\delta_\sigma \neq 0$ and so $\Delta_b \neq \emptyset$. Taking expectations, $P_0' G \delta = \mu_\sigma \delta_\sigma + \mu_a \delta_a$, where $\mu_\sigma = \frac{1}{\sqrt{2}}\mathbb{E}_{\mathcal{P}_0}[\frac{1}{2} r^{-2} - (1-r)r^{-1}\log r - \frac{1}{2} r^{-2}\log r]$ and $\mu_\alpha = \frac{1}{\sqrt{2}}\mathbb{E}_{\mathcal{P}_0}[-(1-r)r^{-1}(1+\log r) + \frac{1}{2} r^{-2}\log r]$. Then Δ_c includes all δ such that $\mu_\sigma \delta_\sigma + \mu_a \delta_a = 0$ which has a solution if $\delta_\sigma \neq 0$ and/or $\delta_a \neq 0$. Hence $\Delta_b \cap \Delta_c \neq \emptyset$.

Example E.4 Bivariate Linear Simultaneous Equations Model with Polynomial Instruments

Consider bivariate linear simultaneous equations model

$$y_j = \theta_{0j} x_j + \varepsilon_j,$$

where $\varepsilon_j = \upsilon \exp(-\zeta_j w/2)$, $(j = 1, 2)$, with common shock υ. Here, $z = (y_1, y_2, x_1, x_2, w)'$ with w a scalar instrument, $\theta = (\theta_1, \theta_2)'$ and $d_\theta = 2$. It is assumed that $\mathbb{E}_{\mathcal{P}_0}[\upsilon|w] = 0$, $\mathbb{E}_{\mathcal{P}_0}[\upsilon^2|w] = 1$, $\mathbb{E}_{\mathcal{P}_0}[x_1|w] = \pi(1+w)$ and $\mathbb{E}_{\mathcal{P}_0}[x_2|w] = \pi(1+w^2)$. Therefore, $\mathbb{E}_{\mathcal{P}_0}[\varepsilon_j|w] = 0$, $\mathbb{E}_{\mathcal{P}_0}[\varepsilon_j^2|w] = \exp(-\zeta_j w)$, $(j = 1, 2)$, and $\mathbb{E}_{\mathcal{P}_0}[\varepsilon_1\varepsilon_2|w] = \exp(-(\zeta_1 + \zeta_2)w/2)$. The instrument w is distributed as a standard normal variate.

The residual function is then defined by

$$\rho(z, \theta) = \begin{pmatrix} y_1 - \theta_1 x_1 \\ y_2 - \theta_2 x_2 \end{pmatrix}.$$

Hence $d_\rho = 2$. Let the d_ψ-vector of instruments $\psi(w) = (1, w, \ldots, w^{d_\psi - 1})'$; cf. section 5.2. Define

$\bar{p} = (p_1', -p_2')'$, where

$$p_j = (1, \frac{\zeta_j}{2}, \dots, \frac{(\zeta_j/2)^{d_\psi - 1}}{(d_\psi - 1)!})', \ (j = 1, 2).$$

Then $\mathbb{E}_{\mathcal{P}_0}[(p_j'\psi(w) - \exp(\zeta_j w/2))^2] \to 0$, $(j = 1, 2)$, as $d_\psi \to \infty$.

Now, writing $\Omega_w(\theta) = \mathbb{E}_{\mathcal{P}_0}[\rho(z,\theta)\rho(z,\theta)'|w] \otimes \psi(w)\psi(w)'$,

$$\Omega_w(\theta) = \begin{pmatrix} \exp(-\zeta_1 w) & \exp(-(\zeta_1 + \zeta_2)w/2) \\ \exp(-(\zeta_1 + \zeta_2)w/2) & \exp(-\zeta_2 w) \end{pmatrix} \otimes \psi(w)\psi(w)'.$$

Therefore

$$\begin{aligned} \bar{p}'\Omega_w(\theta)\bar{p} &= \exp(-\zeta_1 w)(p_1'\psi(w))^2 + \exp(-\zeta_2 w)(p_2'\psi(w))^2 \\ &\quad -2\exp(-(\zeta_1 + \zeta_2)w)(p_1'\psi(w))(p_2'\psi(w)) \\ &\to 0 \text{ as } d_\psi \to \infty. \end{aligned}$$

Hence, $\bar{p}'\Omega\bar{p} \to 0$ as $d_\psi \to \infty$. However, since

$$\mathbb{E}_{\mathcal{P}_0}[\bar{p}'G_i|w] = -\pi \begin{pmatrix} (1+w)p_1'\psi(w) & -(1+w^2)p_2'\psi(w) \end{pmatrix},$$

it follows that $\lim_{d_\psi \to \infty} \bar{p}_0'G \neq 0'$. Therefore, in the limit, as $d_\psi \to \infty$, there exists $\delta \in \mathbb{R}^2$ that satisfy Assumption 3.2(b) but violate Assumption 3.2(c). From Theorem 3.2, the GAR statistic $\hat{T}(\theta_n)$ is asymptotically unbounded for sequences θ_n for such δ. Note that $\bar{p}$ is not an eigenvector corresponding to a zero eigenvalue of $\Omega_w(\theta)$ but rather a vector which annihilates $\Omega_w(\theta)$ as $d_\psi \to \infty$.

To illustrate, consider the case when $d_\psi = 2$, i.e., $\psi(w) = (1, w)'$, and $\zeta_1 = \zeta_2 = 0$. Then

$$G = -\pi \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}, \ P_0' = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Hence,

$$P_0'G\delta = -\frac{1}{\sqrt{2}} \begin{pmatrix} \delta_1 \\ \delta_1 + 2\delta_2 \end{pmatrix}.$$

By Theorem 3.2, $\hat{T}(\theta_n)$ is asymptotically unbounded for sequences $\theta_n \in \Theta_n^0(\Delta_b \cap \bar{\Delta}_c)$ if $\delta_1 \neq 0$ or $\delta_1 \neq -2\delta_2$. Simulation S.2 in Supplement S, which considers a specific case of this example, indicates that confidence regions based on the GAR statistics $\hat{T}(\theta_n)$ are oversized for large n as d_ψ increases for sequences $\theta_n \in \Theta_n^0(\Delta_b \cap \bar{\Delta}_c)$, i.e., for δ such that $\lim_{d_\psi \to \infty} P_0'G\delta \neq 0$.

References

Andrews, D. W. K. and Cheng, X. (2012), 'Estimation and inference with weak, semi-strong, and strong identification', *Econometrica* **80**, 2151–2211.

Andrews, D. W. K. and Cheng, X. (2013), 'Maximum likelihood estimation and uniform inference with sporadic identification failure', *Journal of Econometrics* **173**, 36–56.

Andrews, D. W. K. and Cheng, X. (2014), 'GMM estimation and uniform subvector inference with possible identification failure', *Econometric Theory* **30**, 287–333.

Supplement S: Additional Simulation Evidence for 'Generalised Anderson-Rubin Statistic Based Inference In The Presence Of A Singular Moment Variance Matrix'

Nicky L. Grant[†]
University of Manchester
nicky.grant@manchester.ac.uk

Richard J Smith[‡]
cemmap, U.C.L and I.F.S.
University of Cambridge
University of Melbourne
ONS Economic Statistics Centre of Excellence
rjs27@econ.cam.ac.uk

This Draft: January 2019

Simulation S.1 Nonlinear Regression Model with First Order Moment Singularity

Consider the following nonlinear regression model of Example E.2 in Supplement E

$$y = \alpha_0 + \pi_0(1 + \kappa_0 x)^{-1} + \varepsilon,\ x \geq 0, \mathbb{E}_{\mathcal{P}_0}[\varepsilon|x] = 0.$$

Here the true value of the parameter vector of interest is $\theta_0 = (\alpha_0, \pi_0, \kappa_0)' = (1, 2, \kappa_0)'$ with $\kappa_0 = 0.00$, 0.05 and 0.50, i.e.,

$$y = x + 2(1 + \kappa_0 x)^{-1} + \varepsilon,$$

with x and ε distributed, respectively, as the absolute value of and, given x, standard normal variates. Recall from Example E.2 in Supplement E that if $\kappa_0 = 0.00$, then the first and third element of the moment indicator vector are perfectly correlated, i.e., $\text{rk}(\Omega(\kappa_0)) = 2$; $\text{rk}(\Omega(\kappa_0)) = 3$ if $\kappa_0 \neq 0$.

[*]Address for correspondence: N.L. Grant, Arthur Lewis Building, Department of Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom.

[†]Arthur Lewis Building, Department of Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom.

[‡]Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK.

Table 1 plots the rejection probabilities $\mathcal{P}_0\{\hat{T}(\theta_n) \geq \chi^2_3(0.9)\}$, i.e., level $\alpha = 0.1$, estimated using $R = 5000$ replications for a GAR-based test of the null hypothesis $\mathrm{H}_0 : \theta = \theta_n$ against the alternative hypothesis $\mathrm{H}_1 : \theta \neq \theta_n$ for $\theta_n = \theta_0 + n^{-1}(0,1,1)'$, i.e., $\delta_n = \delta = (0,1,1)'$, for sample sizes $n = 100$, 500, 1000, 5000 and 50000. Recall $\Delta_b = \{\delta \in \mathbb{R}^3 : \delta_2 \neq 0 \text{ or } \delta_3 \neq 0\} \neq \emptyset$ and $\Delta_c = \mathbb{R}^3$. Hence, $\Delta_b \cap \Delta_c = \Delta_b \neq \emptyset$ and Assumption 3.2 is satisfied.

	$\kappa_0 = 0.00$	$\kappa_0 = 0.05$	$\kappa_0 = 0.50$
$n = 100$	0.088	0.088	0.091
$n = 500$	0.099	0.100	0.100
$n = 1000$	0.102	0.099	0.100
$n = 5000$	0.098	0.106	0.099
$n = 50000$	0.094	0.097	0.109

Table 1: GAR Rejection Probabilities: Nonlinear Regression with First Order Moment Singularity.

Table 1 corroborates Theorem 3.1. For the larger sample sizes n, and for all values of κ_0, the 0.9 quantile of $\hat{T}(\theta_n)$ is well approximated by that of the χ^2_3 distribution. Not reported here, in other experiments, this observation appeared to be robust for $\delta_n = o(n^{-1/2})$ with $\delta_2 \neq 0$ or $\delta_3 \neq 0$. If the discretisation Θ_n of Θ is sufficiently fine so that $d_H(\Theta_n, \theta_0) = o(n^{-1/2})$, then the GAR-based confidence region eq. (4.4) covers all sequences $\theta_n \in \Theta_n$ with correct asymptotic size.

Simulation S.2 Bivariate Linear Simultaneous Equations Model with Polynomial Instruments

This simulation is based on a specific case of Example E.4 in Supplement E; *viz.*

$$\begin{aligned} y_1 &= 1.0x_1 + \varepsilon_1,\ x_1 = \pi(1+w) + \eta_1,\ \varepsilon_1 = \upsilon_1 \exp(-\zeta_1 w/2), \\ y_2 &= 0.5x_2 + \varepsilon_2,\ x_2 = -\pi(1+w^2) + \eta_2,\ \varepsilon_2 = \upsilon_2 \exp(-\zeta_2 z/2). \end{aligned}$$

The innovation vector $(\upsilon_1, \upsilon_2, \eta_1, \eta_2)'$ is distributed conditional on w as $N(0, \Xi)$ where

$$\Xi = \begin{pmatrix} 1.0 & \rho & 0.3 & 0.0 \\ \rho & 1.0 & 0.5 & 0.0 \\ 0.3 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 1.0 \end{pmatrix}.$$

We consider values $\rho = 0.999500$, 0.999995 and 1.000000, $\pi = 0.0$, 0.1, 0.5, $(\zeta_1, \zeta_2) = (0.0, 0.0)$, $(0.0, 0.5)$, $(0.0, 1.0)$. The instrument vectors are $(1, w)'$, $(1, w, w^2)'$ and $(1, w, w^2, w^3)'$, i.e., polynomial instruments with dimensions $d_\psi = 2$, 3 and 4 respectively. Hence, since $d_\rho = 2$, $d_g = 2d_\psi$.

GAR statistic rejection probabilities ($\alpha = 0.1$) for $R = 5000$ replications are tabulated for $\theta_n = (1, 0.5)' + n^{-1}\delta$ with $\delta = (1,1)'$ for sample sizes $n = 100, 500, 1000, 5000, 50000$. This choice for δ does not satisfy $\bar{p}_0'G\delta \to 0$; see Example E.4 in Supplement E with $\zeta_1 = \zeta_2 = 0$ for $d_\psi = 2$.[1]

Table 2 displays the GAR-based test rejection probabilities for $\pi = 0.1$.[2] When $\rho = 1.0$ and $\zeta_1 = \zeta_2 = 0$, i.e., Ω is singular, the GAR-based test rejection probabilities converge to 1 as n increases as expected from Theorem 3.2 since $\hat{T}(\theta_n)$ is asymptotically unbounded in this case for any d_g. For both values $\rho = 0.999995$ and $\rho = 0.999500$, the rejection probabilities for any n and d_g are smaller when compared with those for $\rho = 1.000000$ but are still oversized in small samples.

As ζ_2 increases, and the residuals become less correlated conditional on w, the rejection probabilities decrease for any value of ρ and n. Table 3 indicates that the GAR-based test rejection probabilities increase as d_ψ increases, corroborating the intuition in Example E.4 in Supplement E that with many polynomial instruments $\hat{T}(\theta_n)$ diverges when $\Omega_w(\theta_0)$ is nearly singular a.s.(w) and $\bar{p}_0'G\delta \nrightarrow 0$.

This pattern is also observed in Table 3 with stronger instruments, i.e., when $\pi = 0.5$. In this case the GAR-based test rejection probabilities are, in general, relatively more oversized for any n, d_g, ρ and ζ_2 than those when $\pi = 0.1$. In this case $\bar{p}_0'G\delta$ is further from zero so that the rejection probabilities are relatively more oversized than when $\bar{\pi} = 0.1$.

[1] It may be demonstrated that $\bar{p}_0'G\delta \nrightarrow 0$ as $d_\psi \to \infty$ but the proof is omitted for brevity.

[2] The value $\pi = 0.0$ was also considered when $\Delta_c = \mathbb{R}^2$ and Theorem 3.1 holds in all cases. Rejection probabilities around 0.1 were found for all n, ρ, ζ_1, ζ_2 and d_g but these results are not reported here for brevity.

			$\rho = 0.999500$			$\rho = 0.999995$			$\rho = 1.000000$		
			$d_g = 4$	$d_g = 6$	$d_g = 8$	$d_g = 4$	$d_g = 6$	$d_g = 8$	$d_g = 4$	$d_g = 6$	$d_g = 8$
$\zeta_1 = 0.0$	$\zeta_2 = 0.0$	$n = 100$	0.135	0.123	0.198	0.428	0.380	0.800	0.492	0.421	0.867
		$n = 500$	0.110	0.114	0.132	0.727	0.724	0.998	0.995	0.996	1.000
		$n = 1000$	0.106	0.100	0.120	0.628	0.641	0.992	1.000	1.000	1.000
		$n = 5000$	0.091	0.103	0.095	0.251	0.253	0.599	1.000	1.000	1.000
		$n = 50000$	0.092	0.100	0.108	0.117	0.110	0.150	1.000	1.000	1.000
$\zeta_1 = 0.0$	$\zeta_2 = 0.5$	$n = 100$	0.117	0.118	0.360	0.204	0.292	0.800	0.218	0.329	0.850
		$n = 500$	0.102	0.107	0.267	0.124	0.542	1.000	0.119	0.954	1.000
		$n = 1000$	0.106	0.103	0.194	0.105	0.461	1.000	0.104	1.000	1.000
		$n = 5000$	0.105	0.098	0.109	0.106	0.196	0.986	0.094	1.000	1.000
		$n = 50000$	0.103	0.105	0.103	0.099	0.107	0.278	0.095	0.676	1.000
$\zeta_1 = 0.0$	$\zeta_2 = 1.0$	$n = 100$	0.080	0.107	0.521	0.089	0.234	0.739	0.076	0.263	0.764
		$n = 500$	0.094	0.099	0.623	0.086	0.247	1.000	0.094	0.314	1.000
		$n = 1000$	0.087	0.099	0.420	0.099	0.162	1.000	0.095	0.199	1.000
		$n = 5000$	0.098	0.088	0.150	0.093	0.102	0.972	0.102	0.096	1.000
		$n = 50000$	0.101	0.096	0.095	0.099	0.095	0.230	0.104	0.098	1.000

Table 2: GAR rejection probabilities $\pi = 0.1$.

			$\rho = 0.999500$			$\rho = 0.999995$			$\rho = 1.000000$		
			$d_g = 4$	$d_g = 6$	$d_g = 8$	$d_g = 4$	$d_g = 6$	$d_g = 8$	$d_g = 4$	$d_g = 6$	$d_g = 8$
$\zeta_1 = 0.0$	$\zeta_2 = 0.0$	$n = 100$	0.927	0.893	0.999	1.000	1.000	1.000	1.000	1.000	1.000
		$n = 500$	0.495	0.477	0.939	1.000	1.000	1.000	1.000	1.000	1.000
		$n = 1000$	0.317	0.286	0.706	1.000	1.000	1.000	1.000	1.000	1.000
		$n = 5000$	0.145	0.144	0.222	1.000	1.000	1.000	1.000	1.000	1.000
		$n = 50000$	0.104	0.102	0.110	0.530	0.544	0.964	1.000	1.000	1.000
$\zeta_1 = 0.0$	$\zeta_2 = 0.5$	$n = 100$	0.761	0.895	1.000	0.988	1.000	1.000	0.992	1.000	1.000
		$n = 500$	0.292	0.480	1.000	0.690	1.000	1.000	0.713	1.000	1.000
		$n = 1000$	0.193	0.283	1.000	0.404	1.000	1.000	0.425	1.000	1.000
		$n = 5000$	0.106	0.125	0.642	0.148	1.000	1.000	0.148	1.000	1.000
		$n = 50000$	0.100	0.106	0.150	0.102	0.340	1.000	0.107	1.000	1.000
$\zeta_1 = 0.0$	$\zeta_2 = 1.0$	$n = 100$	0.171	0.707	1.000	0.200	1.000	1.000	0.194	0.996	1.000
		$n = 500$	0.101	0.277	1.000	0.097	0.996	1.000	0.089	0.955	1.000
		$n = 1000$	0.096	0.182	1.000	0.091	0.936	1.000	0.098	0.349	1.000
		$n = 5000$	0.088	0.109	0.978	0.094	0.306	1.000	0.092	0.102	1.000
		$n = 50000$	0.095	0.105	0.220	0.102	0.108	1.000	0.091	0.102	1.000

Table 3: GAR rejection probabilities $\pi = 0.5$.