

GMM Estimation and Inference with (Asymptotically) Singular Moment Variance

Nicky Grant

University of Manchester

This Draft: September 2018

Abstract

This paper considers efficient GMM estimation from linear strongly identified moment functions with (asymptotically) singular variance at the true parameter, β_0 . The 2-Step GMM estimator with singular variance is shown to converge at rate \sqrt{n} to a non-standard limit distribution, with convergence at rate n in certain directions when the null space of the moment variance matrix does not lie within that of the outer product of the expected first order derivative at β_0 . Many examples of (almost) singular are provided along with cases in which the null space condition does not hold. The implications for inference of some common methods of removing singularities employed in applied research are discussed. A simulation study demonstrates the key results of this paper.

KEYWORDS: GMM, Efficiency, Singular Variance.

1 Introduction

Moments with (almost) singular variance at β_0 are often encountered in econometric research. Common examples include singularity from identification failure in non-linear models (Andrews & Cheng (2012), Andrews & Guggenberger (2015), Grant (2013)), simultaneous equation and dynamic panel models (Arrelano & Bond (1991), Han & Phillips (2010)), and parameter restriction tests under the null, (Andrews (1987), Penaranda & Sentana (2010), Dufour & Valery (2011)). However the general properties of GMM type estimators with singular variance remain largely undeveloped. Some headway has been made in the case where singularity arises from identification failure, for example the properties of the Maximum Likelihood estimator with singularity in non-linear models (Melino (1982), Lee & Chesher (1986), Bottai et al (2003), Andrews & Cheng (2013)). In most cases where singularity does not arise from identification failure estimation is based on some form of regularised GMM estimator that removes moment singularity e.g Doran & Schmidt (2006). As such the properties of the 2-Step GMM estimator with singular variance in such cases remain largely unstudied.

This paper derives the asymptotic properties of the 2-Step GMM estimator in identified linear models where moment variance singularity may occur at the true parameter, though is non-singular for (small) perturbations from β_0 . In this the 2-Step GMM estimator is shown to have a highly nonstandard distribution with convergence at rate n in certain directions when the null space of the moment variance matrix is not a subset or equal to that of the outer product of the expect first order moment derivative at β_0 . The condition that moment variance is non-singular for small perturbations to every element of β_0 is shown to hold in all examples provided in this paper- where it is argued violations would occur only in pathological cases.

To derive the properties of 2-Step GMM with singular variance, eigen-pair expansions of the moment variance matrix around β_0 in Grant (2013)¹ used to study the properties of the Generalised Anderson Rubin Statistic with singular variance are utilised².

The nearest paper to this is Grant (2014) that considers the properties of 2-Step GMM when eigenvalues are small but converge at a slow enough rate such that the GMM estimator when appropriately scaled has a Gaussian distribution. Namely when smallest eigenvalue is non-zero and $O(n^{-\delta})$ for $\delta < 1$ then 2-Step is shown to converge at non-standard rate $n^{(1+\delta)/2}$ in certain directions when the null space condition does not hold to the usual Gaussian limit distribution under strong identification, Hansen (1982).

Other papers have also considered the issue of (almost) singular variance for the properties of mo-

¹These Eigen-pair expansions are based heavily on the results in Kato (1982) and others.

²These expansions are valid when the eigenvalues of the moment variance matrix are simple which restricts the dimension of the null space the be less than or equal to one.

ment based estimators satisfying the identification condition. Doran & Schmidt (2006) consider a regularised GMM estimator based on the method of principal components for dynamic panel GMM with moment variance close to singular. Ruge-Murcia (2007) overcome the issue of (almost) singularities in linearised DSGE models by adding random noise to the moment function termed 'stochastic singularity'. Penaranda & Sentana (2010) consider the case the singularity at β_0 is known under the null and propose a 2-step GMM estimator with weight matrix based on a generalized inverse of the sample moment variance. Such methods are commonly used in empirical research when moments variances are (close to) singular along with other ad-hoc methods such as adding and deleting exogenous/endogenous variables in IV simultaneous equation models. Once the singularity is removed then under the usual identification and further regularity conditions 2-Step GMM has a Gaussian limit distribution with root-n rate of convergence.

Results in this paper show that when singularity occurs at small enough region around β_0 then removing singularities in the moment variance via some form of regularisation or otherwise increases the asymptotic variance of 2-Step GMM and potentially reduces the rate of convergence. This suggests that the method commonly applied in empirical research can lead to a less efficient estimator of the true parameter. Given the highly non-standard limit distribution of 2-Step GMM then inference based on the normal approximation without a regularisation would be asymptotically invalid. As such results in this paper suggest that regularisation does have the benefit of validating standard inference methods. A preferred method may be to use the distributional results in this paper robust to singularity, yielding asymptotically valid inference along with a more efficient estimator.

To highlight the results in this paper general examples of singular variance are provided based on a linear simultaneous IV model. We show how singularity may arise from common shocks in the unobservable when interacted with instruments. Using these examples we show how the methods of regularisation, changing the instrument set and stochastic singularity of Ruge-Murcia (2007) can remove moment singularities.

A simulation study verifies the main results of this paper- along with the properties of 2-Step GMM and the Wald Statistic based on such methods of removing singularities. As predicted by results of this paper removing singularities leads to an increased moment variance, a lower speed of convergence (when the null space condition does not hold) where the small sample distribution of 2-Step GMM is better approximated by a Gaussian distribution as the sample size increases.

This paper considers linear models with a finite number of moments that satisfy the identification condition. Extensions to non-linear models and/or many weak or unidentified instruments would be possible, though to clearly highlight the implications of singular variance in isolation are not considered in this appear for clarity. Correspondingly we assume that the dimension of the null space

of the variance matrix at β_0 is at most 1, i.e. at most one zero (or $O(n^{-1})$) eigenvalue. Again this simplifies the proofs of this paper. The author intends to work on the case multiple zero eigenvalues in future work.

Section 2 discusses singularity and the main results of the paper. Section 3 derives the asymptotic properties of the eigenvalues and eigenvectors of the sample moment variance evaluated at an initial GMM estimator. Using the results in Section 3 the asymptotic properties of 2-Step GMM and related statistics are provided in Section 4. Examples of singular variance and a simulation study are provided in Sections 5 and 6 respectively. Finally some concluding remarks on directions for future research and extensions are detailed in Section 7. Proofs and definitions used throughout the paper are collected in an Appendix.

1.1 Notation and Setup of Paper

For simplicity this paper considers $\{x_t: (t = 1, \dots, T), T \geq 1\}$ where $x \in \mathcal{X} \subseteq \mathbb{R}^k$. Let the known moment function $m \times 1$ moment function $g(\cdot, \cdot) : \mathcal{X} \times \mathbf{B} \mapsto \mathbb{R}$ where $\mathbf{B} \subset \mathbb{R}^p$ is the parameter space

$$E\left[\frac{1}{T} \sum_{t=1}^T g(x_t, \beta_0)\right] = 0 \quad (1)$$

where $\beta_0 \in \text{int}(\mathbf{B})$ which is a standard regularity assumption to avoid a parameter on the boundary issue Andrews (1999). An assumption of linearity is made partly for simplicity and to highlight the unto now properties of 2-Step GMM with singular variance clearly. We make the following definitions,

$$g_T(\beta) = E\left[\sum_{t=1}^T g(x_t, \beta)/T\right] \quad \hat{g}_T(\beta) = \frac{1}{T} \sum_{i=1}^T g_t(\beta) \quad \hat{g} = \hat{g}_T(\beta_0) \quad (2)$$

$$\hat{\Omega}_T(\beta) = \frac{1}{T} \sum_{t=1}^T g_t(\beta)g_t(\beta)' \quad \Omega_T(\beta) = \frac{1}{T} \sum_{t=1}^T E[g_t(\beta)g_t(\beta)'] \quad \Omega(\beta) = \lim_{T \rightarrow \infty} \Omega_T(\beta) \quad (3)$$

where we let $\hat{\Omega}_T = \hat{\Omega}_T(\beta_0)$, $\Omega_T := \Omega_T(\beta_0)$, $\Omega := \Omega(\beta_0)$. Define the first order derivative of the moment function (where under the linearity assumption is not a function of β)

$$G_t = \frac{\partial}{\partial \beta'} g_t(\beta) \quad G_T := E\left[\frac{1}{T} \sum_{t=1}^T G_t\right] \quad \hat{G}_T = \frac{1}{T} \sum_{t=1}^T G_t \quad (4)$$

2 Singular Variance & Identification

We first define Almost Singular Variance (ASV)

ALMOST SINGULAR VARIANCE

$$\Omega_T = P_{+T}\Lambda_{+T}P'_{+T} + P_{0T}\Lambda_{0T}P'_{0T}$$

where Λ_{+T} is an $(m - \bar{m}_T) \times (m - \bar{m}_T)$ diagonal matrix where $[\Lambda_{+T}]_{jj} = \lambda_{+jT}$ where $\inf_{T \geq 1} \lambda_{+jT} > 0$ for all $j = 1, \dots, m - \bar{m}_T$ and Λ_{0T} an $\bar{m}_T \times \bar{m}_T$ where $P_T := (P_{+T}, P_{0T})$ is such that $P'_T P_T = I_m$ and Λ_{0T} is a diagonal $\bar{m}_T \times \bar{m}_T$ matrix where $[\Lambda_{0T}]_{jj} = \bar{\lambda}_{0jT}/T^{\delta_j}$ for all $j = 1, \dots, \bar{m}_T$ where $0 \leq \bar{\lambda}_{0jT} < \infty$, $\delta_j > 0$.

The Almost Singular Variance (ASV) assumption allows $0 \leq \bar{m}_T \leq m$ of the eigenvalues of the moment variance at the true parameter Ω_T converge to or be exactly equal zero. Define (assuming the limits exist) $P_+ = \lim_{T \rightarrow \infty} P_{+T}$ and $\Lambda_+ = \lim_{T \rightarrow \infty} \Lambda_{+T}$ then as $\Lambda_{0T} \rightarrow 0$ and $P'_{0T} P_{0T} = I_{\bar{m}_T}$ so that P_{0T} is bounded then $\Omega_T \rightarrow \Omega$ where $\Omega = P_+ \Lambda_+ P'_+$. Let P_0 be the eigenvector of the $\bar{m} = \lim_{T \rightarrow \infty} \bar{m}_T$ zero eigenvalues of Ω such that $P'_0 P_+ = 0$ and $P'_0 \Omega P_0 = 0$. Under these assumptions we can show that $P_{0T} \rightarrow P_0$ which is crucial in deriving the properties of the 2-Step GMM estimator under ASV.

We allow Nearly Weak Identification similar to the many weak moment setup in Newey & Windmeijer (2009) though we consider for simplicity the case of a finite number of moments. Namely we assume

$$\sqrt{T} S_T^{-1} G'_T \rightarrow G' \tag{5}$$

where G is a $p \times p$ full rank matrix where $S_T = \bar{S}_T \text{diag}(\mu_{1T}, \dots, \mu_{pT})$ and \bar{S}_T is a bounded full rank $p \times p$ matrix where $\mu_{1T} \geq \mu_{2T} \geq \dots \geq \mu_{pT} > 0$ and $\mu_{1T}/\sqrt{T} \rightarrow c$ for some $0 \leq c < \infty$ and $\mu_{pT} \rightarrow \infty$. Define $\bar{G} = \lim_{T \rightarrow \infty} \bar{G}_T$. Under this assumption and the requisite regularity conditions then the first step GMM estimator is consistent and converges in distribution to a Normal distribution at rate less than \sqrt{T} when $\mu_{pT} \rightarrow \infty$.

3 Asymptotic Properties of 2-Step GMM with Singular Variance

ASSUMPTION 1 (A1) : *Eigensystem Asymptotics*

(i) $w_i (i = 1, \dots, n)$ is an i.i.d sequence, (ii) $\mathbb{E}[\|g_i\|^4] < \infty$, (iii) $\frac{1}{n} \sum_{i=1}^n \|G_i(\beta) - G_i(\beta^*)\| \leq \hat{M} \|\beta - \beta^*\| \forall \beta, \beta^* \in \mathbf{B}$ where $\hat{M} = O_p(1)$, (iv) $\mathbb{E}[\|G_i\|^2] < \infty$, (v) $\|\hat{\Omega}(\beta) - \hat{\Omega}(\beta^*)\| \leq \hat{M} \|\beta - \beta^*\| \forall \beta, \beta^* \in \mathbf{B}$ for some $\hat{M} = O_p(1)$, (vi) Eigenvalues Λ are simple, (vii) $\|\Lambda\| \leq K$ for some $K < \infty$, (viii) $m < \infty$.

Under the additional strong identification condition in Assumption 2 in Section 4 then it is straightforward to show $\sqrt{n}(\tilde{\beta} - \beta_0) = \Xi \sqrt{n} \tilde{g} + o_p(1)$ for some bounded $p \times m$ matrix Ξ . Hence $\tilde{\beta} - \beta_0 = O_p(n^{-1/2})$. Define $c_i = n^{1/2} P'_0 g_i$ noting that $c_i = 0$ w.p.1 if Ω_n is singular (i.e $\Lambda_0 = 0$) where $c_i = O_p(1)$ when $\Lambda_0 \neq 0$ since $P'_0 \Omega_n P_0 = P'_0 E[g_i g'_i] P_0 = \Lambda_0/n$. This paper considers WSV with $\delta = 1$ and allowing

for exact singularity. We now set out some definitions to simplify the proof of Theorem 1. Write $G'_i = (G_{i1}, \dots, G_{im})$ where G'_{il} is the l 'th row of G_i . Define $\theta_{jk} := \Xi' E[G_{ik}g_{ij}]$, $v_{jk} := \Xi' E[G_{ik}c_{ij}]$, $\gamma_{jk} = \Xi' E[G_{ij}G'_{ik}] \Xi$ resp. for $j, k = \{1, \dots, m\}$ then define the following matrices

$$\Theta := \begin{bmatrix} \theta_{11} & \cdots & \theta_{1m} \\ \vdots & \ddots & \vdots \\ \theta_{m1} & \cdots & \theta_{mm} \end{bmatrix} \quad (6)$$

$$\Upsilon := \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mm} \end{bmatrix} \quad (7)$$

$$\Gamma := \begin{bmatrix} \text{vec}(\gamma_{11})' & \cdots & \text{vec}(\gamma_{1m})' \\ \vdots & \ddots & \vdots \\ \text{vec}(\gamma_{m1})' & \cdots & \text{vec}(\gamma_{mm})' \end{bmatrix} \quad (8)$$

The asymptotic distribution correspondingly the distribution of 2-Step GMM) when $\bar{m} > 0$ depends on both the distribution of $\sqrt{n}\hat{g}$ as standard but also limit distribution of the first order derivative, i.e the matrix random variable $\sqrt{n}(\hat{G} - G)$. As such we must model potential dependence between both random variables. Under A1(i),(ii),(iii) by Lindberg Levy Multivariate Central limit Theorem

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} g_i \\ \text{vec}(G_i - G) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_g \\ Z_G \end{pmatrix} \stackrel{d}{\sim} N(0_{m(p+1) \times 1}, \Sigma) \quad (9)$$

$$\Sigma := \begin{pmatrix} \Omega & \Omega_{Gg} \\ \Omega'_{Gg} & \Omega_{GG} \end{pmatrix} \quad (10)$$

Where $\Omega_{Gg} = E[\text{vec}(G_i - G)g'_i]$, $\Omega_{GG} = E[\text{vec}(G_i - G)\text{vec}(G_i - G)']$. Then we can show,

$$\sqrt{n}(\hat{G} - G) \xrightarrow{d} \bar{Z}_G \quad (11)$$

$$\bar{Z}_G = (Z_{G1}, \dots, Z_{Gp}) \quad (12)$$

where Z_{Gj} is a $m \times 1$ sub-vector of containing elements $\{(j-1)m+1 : jm\}$ of Z_G .³

Firstly we derive eigensystem asymptotic properties of the eigensystem of $\hat{\Omega}(\tilde{\beta})$ which are critical to the limit distributions for 2-Step GMM shown in Theorem 2 in Section 4. Define $\Omega_+^* = P_+ \Lambda_+^{-1} P_+'$. In

³To show this formally note that the characteristic function (c.f) of $\sqrt{n}(\hat{G} - G)$ (i.e $M_G(T) = \mathbb{E}[\exp(\text{tr}((T' \sqrt{n}(\hat{G} - G))) = \mathbb{E}[\exp(\text{tr}(T' \text{vec}(\sqrt{n}(\hat{G} - G)))]$) where $T \in \mathbb{R}^{p \times m}$) is equivalent to the c.f of $\text{vec}(\sqrt{n}(\hat{G} - G))$ (i.e $M_{\text{vec}(G)}(t) = E[\exp(t' \text{vec}(\sqrt{n}(\hat{G} - G)))]$) where $t = \text{vec}(T)$ using the result $\text{tr}(AB) = \text{vec}(A)' \text{vec}(B)$.

Theorem 2 assumption 2 (A2) is provided in Section 4.

THEOREM 1

$$\begin{aligned} \mathcal{W}_{\Lambda_0} &:= \Lambda_0 + P_0' ((I_m \otimes Z_g') \Upsilon + \Upsilon' (I_m \otimes Z_g)) P_0 \\ &\quad + P_0' (\Gamma(I_m \otimes Z_g \otimes Z_g) - \Theta' (I_m \otimes Z_g) \Omega_+^* (I_m \otimes Z_g') \Theta) P_0 \end{aligned} \quad (13)$$

$$\mathcal{W}_{P_0} \xrightarrow{d} \Omega_+^* E[g_i c_i'] + \Omega_+^* (I_m \otimes Z_g') \Theta P_0 \quad (14)$$

If Ω_n is exactly singular, i.e $\Lambda_0 = 0$, then $c_i = 0$ w.p.1 then $E[g_i c_i'] = 0$ and $\Upsilon = 0_{m \times m}$ so that (13), (14) resp. imply,

$$n \hat{\Lambda}_0(\tilde{\beta}) \xrightarrow{d} P_0' (\Gamma(I_m \otimes Z_g \otimes Z_g) - \Theta' (I_m \otimes Z_g) \Omega_+^* (I_m \otimes Z_g') \Theta) P_0 \quad (15)$$

$$n^{1/2} (\hat{P}_0(\tilde{\beta}) - P_0) \xrightarrow{d} \Omega_+^* (I_m \otimes Z_g') \Theta P_0 \quad (16)$$

In this case the drift terms drop out and the asymptotic distribution of $n \hat{\Lambda}_0(\tilde{\beta})$ is consistently estimable. Hence allowing for $\Lambda_{0n} = O(n^{-1})$ does not appreciably change the results relative to the case of exact singularity $\Lambda_{0n} = 0$. Note that (15) could be used to test for singularity of Ω_n under a strong identification assumption.

4 2-Step GMM with Singular Variance

Define the initial GMM estimator $\tilde{\beta}$ (for simplicity suppressing notion governing potential dependence on the initial estimate depend on W) based on some (potentially data dependent) symmetric (asymptotically) bounded p.s.d. matrix W_n , where $W_n \xrightarrow{p} W$ and $\hat{Q}_W(\beta) := n \hat{g}(\beta)' W_n \hat{g}(\beta)$.

$$\tilde{\beta} = \arg \min_{\beta \in \mathbf{B}} \hat{Q}_W(\beta) \quad (17)$$

Then under A1 ,A2 by Hansen (1982),

$$\tilde{\beta} \xrightarrow{p} \beta_0 \quad (18)$$

$$n^{1/2} (\tilde{\beta} - \beta_0) = \Xi \sqrt{n} \hat{g} + o_p(1) \quad (19)$$

Where $\Xi = (G'WG)^{-1}G'$ and the asymptotic variance of $n^{1/2}(\tilde{\beta} - \beta_0)$ is $\Phi = \Xi \Omega \Xi'$ (suppressing dependence of Ξ on W) where W is such that $G'WG$ is full rank. Note that this result holds whether

or not Ω is non-singular.

Define the 2-Step GMM objective function,

$$\hat{Q}_{opt}(\beta) = n\hat{g}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta) \quad (20)$$

and corresponding 2-Step GMM estimator.

$$\hat{\beta} = \arg \min_{\beta \in \mathbf{B}} \hat{Q}_{opt}(\beta) \quad (21)$$

Non-standard asymptotic results when using $W_n = \hat{\Omega}(\tilde{\beta})^{-1}$ when Ω_n is (weakly) singular.

ASSUMPTION 2 (A2) IDENTIFICATION CONDITIONS

(i) $\text{Rank}(G) = p$, (ii) $\Omega_n(\beta)$ satisfies WSV with $\delta = 1$.

A2(i) assumes first order identification, the standard identification condition, which under the maintained linearity assumption is equivalent to global identification.⁴ A2(i) (along with the regularity conditions A1(i),(ii)) establish \sqrt{n} convergence of the initial GMM estimator. This results holds whether or not Ω_n is singular. Note that A2 (ii) the assumption of WSV with $\delta = 1$ includes the case of non-singularity ($\bar{m} = 0$) and is made as an identification condition given the impact on the rate of convergence and distribution of 2-Step GMM⁵.

A1(ii) encapsulates the standard assumption that Ω_n is full rank when $\bar{m} = 0$. In this case standard strong identified GMM asymptotics holds, e.g Hansen (1982). If $\bar{m} > 0$ then asymptotic of 2-Step GMM under A1(iii) depends on the distribution of $\tilde{\beta}$.

Let B be a bounded full rank $p \times p$ a matrix such that⁶

$$B'G'P_0 := G'_{BP_0} \quad (22)$$

$$G_{BP_0} = \left(G'_{BP_0}, 0_{\bar{m} \times (p-\bar{p})} \right) \quad (23)$$

⁴This result could be extended to allow the many-weak moment setup of Newey & Windmeijer (2009), however for clarity in exposition on the unto now unknown impact of singular variance for inference is left for future research.

⁵Together with the results of Grant (2014) for $0 < \delta < 1$ we could allow for differing rates of convergence of Λ_{0n} , though for clarity of exposition given the complexity of the limit distributions for the singular variance assumed in this paper this result is omitted

⁶This transformation is made- similar to the method in Phillips (1989) to isolate the directions of convergence of $\hat{\beta} - \beta_0$ with different rates- in this case rate n and $n^{1/2}$ since the term in the asymptotic expansion of $\sqrt{n}(\hat{\beta} - \beta_0)$, namely $(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}$ will shrink at rate n^{-1} (since $\hat{\Lambda}_0(\tilde{\beta})^{-1} = O_p(n)$) in certain directions which will yield the faster rate of convergence of $\sqrt{n}(\hat{\beta} - \beta_0)$.

where $G_{BP_0}^{\bar{p}}$ is a full column rank $\bar{p} \times \bar{m}$ matrix for some $0 \leq \bar{p} \leq p$. Note that B is not unique. If $G'P_0 = 0_{p \times \bar{m}}$ then $\bar{p} = 0$ and any B satisfies (23). If $G'P_0$ is full column rank (if $\bar{p} \leq \bar{m}$) then $\bar{p} = p$ and again any full rank matrix B satisfies (23). If $0 < \bar{p} < p$ then we can perform Gauss-Jordan elimination to re-write $G'P_0$ in the form (23).

Similarly define,

$$B'G'P_+ := G'_{BP_+} \quad (24)$$

$$G_{BP_+} = \left(G_{BP_+}^{\bar{p}'}, G_{BP_+}^{p-\bar{p}'} \right) \quad (25)$$

$G_{BP_+}^{\bar{p}}, G_{BP_+}^{p-\bar{p}}$ as the upper $\bar{p} \times (m - \bar{m})$ and lower $(p - \bar{p}) \times (m - \bar{m})$ sub-matrix of G_{BP_+} respectively.

$$\text{Define } B_n = \begin{pmatrix} n^{1/2}I_{\bar{p}} & 0_{\bar{p} \times \bar{p}} \\ 0_{(p-\bar{p}) \times (p-\bar{p})} & I_{p-\bar{p}} \end{pmatrix} B^{-1}, \bar{B} = \lim_{n \rightarrow \infty} B_n^{-1} = B \begin{pmatrix} 0_{\bar{p}} & 0_{\bar{p} \times \bar{p}} \\ 0_{(p-\bar{p}) \times (p-\bar{p})} & I_{p-\bar{p}} \end{pmatrix}.$$

Theorem 1 below shows that the limit distribution of $\sqrt{n}B_n(\hat{\beta} - \beta_0)$ is highly non-standard. B_n will contain \bar{p} elements growing at rate $n^{1/2}$ where \bar{p} is the rank of $G'P_0$. Namely when there exist \bar{p} linearly independent directions such that $\delta'\Omega = 0$ does not imply $\delta'G \neq 0$.

Define the following,

$$\mathcal{C} := (P_0'\bar{Z}_G + \mathcal{W}'_{P_0}G)\bar{B} + G_{BP_0} \quad (26)$$

$$\Phi_1 = \bar{B}'G'_{BP_+}\Lambda_+^{-1}G_{BP_+}\bar{B} + \mathcal{C}'\mathcal{W}_{\Lambda_0}^{-1}\mathcal{C} \quad (27)$$

$$\Phi_2 = \bar{B}G'_{BP_+}\Lambda_+^{-1}P_+' + \mathcal{C}'\mathcal{W}_{\Lambda_0}^{-1}\mathcal{W}_{P_0} \quad (28)$$

ASSUMPTION 3: *Regularity Conditions*

(i) \mathcal{W}_{Λ_0} is full rank w.p.1, (ii) Φ_1 is full rank w.p.1.

A3() is a mild restriction that \mathcal{W}_{Λ_0} is full rank which requires that $\hat{\Omega}(\tilde{\beta})$ exists w.p.1. Only in pathological cases would A3(i) or (iii) break down, where further discussion is provided in the appendix. A3(ii) is a high level condition for the existence of the limit distribution of 2-Step GMM which will depend on singularities the distribution of Z_G .

For simplicity Theorem 1 is provided for the case of exact singularity ($\Lambda_0 = 0$) allowing for general Λ_0 would change the result below by the inclusion of a drift term governed by Λ_0 . Define $\hat{V}(\beta) = (\hat{G}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G}(\beta))^{-1}$.

THEOREM 1: Under A1,A2,A3

$$\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1) \quad (29)$$

$$\begin{aligned}\sqrt{n}B_n(\hat{\beta} - \beta_0) &\xrightarrow{d} \Phi_1^{-1}\Phi_2Z_g \\ \hat{V}(\tilde{\beta})^{-\frac{1}{2}}(\hat{\beta} - \beta_0) &\xrightarrow{d} \Phi_1^{-1/2}\Phi_2Z_g\end{aligned}\tag{30}$$

REMARKS

(i) When $\bar{m} = 0$ (Ω non-singular) then $\bar{p} = 0$ $\mathcal{C} = 0$, $\bar{B} = I_p$ $\Phi_1 = (G'_{BP_+}\Lambda_+^{-1}G_{BP_+} = B'G'\Omega^{-1}GB$ since $\Omega = \Omega_+$ which is full rank when $\bar{m} = 0$. Note also that $\Phi_2 = B'G'\Omega^{-1}$ then finally since $B_n = B$ when $\bar{p} = 0$ then (30) of Theorem 1 collapses to the well known result in the non-singular case $\sqrt{n}B^{-1}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, (B'G'\Omega^{-1}GB)^{-1})$.

(ii) When $\bar{m} > 0$ (Ω singular) both Φ_1 , Φ_2 are non-linear functions of Z_g , Z_G . If $\bar{p} = 0$ then $B_n = B$ and convergence occurs at the standard rate \sqrt{n} . When $\bar{p} > 0$ (i.e $G'P_0 \neq 0$) then convergence of $\hat{\beta} - \beta_0$ occurs at rate n in certain directions governed by B^{-1} .

(iii) If a regularisation was performed using a transformed moment function with a full rank $m - \bar{m}$ variance matrix, then the standard asymptotic result in the identified case holds and $\bar{C} = 0$ which will increase the asymptotic variance of the GMM estimator.

Theorem 1 can be used to derive the limit distribution of the GMM Wald Statistic $\hat{W} = n(\hat{\beta} - \beta_0)' \hat{\Omega}(\tilde{\beta})^{-1}(\hat{\beta} - \beta_0)$

COROLLARY 1 Under A1-A3 as a consequence of Theorem 1,

$$\hat{W} \xrightarrow{d} Z'_g\Phi'_2\Phi_1^{-1}\Phi_2Z_g\tag{31}$$

When $\bar{m} > 0$ then unless a regularisation is performed to remove the singularity in Ω then inference on β_0 should be based on the Wald Statistic and corresponding test statistics with the limit distributions provided in Theorem 1 and Corollary 1. Section 4.1 discusses how to perform such inference which requires estimates of \bar{m} , \bar{p} that converge in with probability 1. To do so we use the results in Theorem 1 that if $\bar{m} = 1$ then $n\hat{\lambda}_j(\tilde{\beta}) = O_p(1)$ for some $j = \{1, \dots, m\}$ and if $\bar{m} = 0$ then $n\hat{\lambda}_j(\tilde{\beta}) \rightarrow \infty$ for all $j = \{1, \dots, m\}$. which allows the estimation of \bar{m} w.p.1 (under the maintained assumption that $\bar{m} \leq 1$).

4.1 Inference Robust to Singularity

[Under Construction]

5 Examples of Singular Moment Variance in Linear Models

This section provides examples of singular variance in linear models. Some discussion is also provided on the impact of some common ad-hoc methods in the applied research to remove (almost) singu-

larities in the moment function. These include changing the instrument set, expanding the included variable set and also the stochastic singularity approach commonly used in linearised DSGE models. A simulation experiment in Section 6 shows the impact on asymptotic properties of efficient GMM estimation when (almost) singular variance is removed from the moment function via some transformation.

EXAMPLE 1: LINEAR IV SIMULTANEOUS EQUATIONS

This section considers a more general example of the linear IV example used in Grant (2014) for Linear IV simultaneous equations. We also provide some discussion on the ad hoc methods of overcoming singularity which seem to be new in the literature.

$$y_j = \theta'_{0j}x_j + \delta_{0j}x + \varepsilon_j \quad (32)$$

For $j = \{1, \dots, J\}$ where $\theta_{0j} \in \mathbb{R}^{p_j}$ where $p = \sum_{j=1}^J p_j$, $\kappa_{0j} \in \mathbb{R}$, where the true unknown partaker vector $\beta_0 = (\theta_0, \kappa_0)'$ where $\theta_0 = (\theta_{01}, \dots, \theta_{0J})$, $\kappa_0 = (\kappa_{01}, \dots, \kappa_{0J})$. This specification allows the inclusion of a common variable x , where high correlation within the moment functions could arise from omission of this variable, i.e if we run

$$y_j = \theta'_{0j}x_j + \epsilon_j \quad (33)$$

Where $\epsilon_j = \varepsilon_j + \delta_{0j}x$ for $j = \{1, \dots, J\}$.

Suppose we have m^* instruments z where for simplicity $z \sim N(0, I_{m^*})$ where x independent of $(z, x_J)'$ where $x_J = (x_1, \dots, x_J)'$ where for simplicity assume $E[x] = 0, E[x^2] = \sigma_x^2$. Define $\epsilon = (\epsilon_1, \dots, \epsilon_J)'$ where $E[\epsilon|z] = 0$ where. Suppose

$$x_j = f_j(z) + \eta_j \quad (34)$$

for some functions $f_j(\cdot)$ and where $E[\eta_j|z] = 0$ and $E[\eta_j^2|z] = \sigma_{\eta}^2$ for $j = \{1, \dots, J\}$. Then the following moment function,

$$g(\beta) = \epsilon(\beta) \otimes \phi(z) \quad (35)$$

with $\beta := (\theta, \delta)'$ where $\theta = (\theta'_1, \dots, \theta'_J)' \in \mathbb{R}^p$ and $\delta = (\delta_1, \dots, \delta_J)' \in \mathbb{R}^J$ where $\phi(\cdot)$ is $m \times 1$ function of the instruments z (commonly polynomials in z). This gives mJ moments, then for $p \leq mJ$ and G full rank then $E[g(\beta)] = 0$ uniquely at $\beta = \beta_0$. Suppose the unobservable residuals in (32) satisfy

$$\varepsilon_j = v_j h_j(z) \quad (36)$$

For some functions $h_j(\cdot)$ allowing general forms of heteroscedasticity of ε_j in the instruments where for simplicity assume $E[v_j|z] = 0$ and $E[v_j^2|z] = 1$ for $j = \{1, \dots, J\}$.

When running regressions based on (33), the moment function at β_0 is,

$$g(\beta_0) = (\varepsilon + \kappa_0 x) \otimes \phi(z) \tag{37}$$

where Grant (2014) considers a special case with $J = 2$, $p_1 = p_2 = 1$, $m^* = 2$ where $\kappa_0 = 0$ with $h_1(z) = z_2$ and $h_2(z) = z_1$ hence

$$\text{cor}(g_1(\beta_0), g_2(\beta_0)) = \text{cor}(\varepsilon_1 z_2, \varepsilon_2 z_1) = \text{cor}(v_1, v_2) := \rho \tag{38}$$

where $\rho \approx 1$ then Ω_n is almost singular since some linear combinations of $g(\beta_0)$ are highly correlated. Under the assumptions made $\text{cor}(\varepsilon_1, \varepsilon_2) = 0$ so that correlations between the moment function at β_0 does not arise from correlation in the residuals of (32). Simulation evidence based on this example in Grant (2014) shows the poor approximation of a Gaussian distribution to the distribution of 2-Step GMM even for large samples for ρ near to 1. Below we use simple examples of Linear IV simultaneous equations to model some common causes of singular variance and detail the theoretical consequences of ad-hoc methods used in practise to remove (almost) singularities.

5.1 Ad-Hoc Methods of Removing Moment Singularity

CHANGING THE INSTRUMENT SET

A common problem faced in dynamic panel and large dynamic linear simultaneous systems is both how to choose the instrument set, namely $\phi(z)$. Often there are potentially an infinite number of moments to choose from, where the many instrument problem is well known in the literature. Commonly the properties of GMM can vary dramatically when altering the instrument set, which is often attributed non-standard properties of GMM with many/weak instruments, e.g Stock & Wright (2000). However as seen in Theorem 1 the distribution of 2-Step GMM is highly non-normal, if changing the instrument set removes moment singularity but the new moment still identified then the limit distribution is asymptotically normal. As such changing the instrument set can lead to distributional changes in the 2-Step GMM even in a strongly identified setting with few instruments.

This issue has largely been left untreated in the theoretical econometric literature, possibly due to the lack of results on the properties of GMM with (almost) singular variance. This example shows that even for small m and J where instruments are strong, the rank of the variance matrix can change when altering which instruments are included. A simulation example shows the quite dramatic change

in even large sample properties of 2-Step GMM.

Consider Example 1 with $J = 2$, $m^* = 3$ where $\kappa_0 = 0$ where we assume $p \leq mJ$ and G is full rank so the model is identified. Suppose $\phi(z) = (z_1, z_2, z_1^2)'$ where $h_1(z) = z_1$ and $h_2(z) = 1$ then

$$g(\beta_0) = \begin{pmatrix} v_1 z_1 \\ v_2 \end{pmatrix} \otimes \begin{pmatrix} z_1 \\ z_2 \\ z_1^2 \end{pmatrix} \quad (39)$$

and $\text{cor}(g_1(\beta_0), g_6(\beta_0)) = \text{cor}(v_1 z_1^2, v_2 z_1^2) = \rho$ so when $\rho = 1$ then Ω_n is exactly singular. Though note given $E[z_1] = 0$ and is independent of v_1, v_2 then $\text{cor}(\epsilon_1, \epsilon_2) = \rho E[z_1] = 0$ so that the residuals from the regression function are uncorrelated. If we use $\phi^*(z) = (z_1, z_2)$ then $g^*(\beta_0) = \epsilon(\beta) \otimes \phi^*(\beta)$ has no perfectly correlated elements, even if $\rho = 1$.

CHANGING THE INCLUDED VARIABLE SET

Below we consider a simple example where (almost) singular variance may also (partly) arise due to a common omitted variable x . Since in practise a common method of removing singularities from the moment is to include variables or factors that are suspected to potentially drive the correlation in the moment function. Simulation evidence in Section 5 demonstrates the implication for inference of changing the variable set to remove singularities in light of result in Theorem 2 and are new in the literature.

Firstly we consider a more general example, providing a special case below, which though unrealistic serves to model the impact of changing the variable set on the asymptotic properties of 2-Step GMM. To model this take the case $J = 2$, $m^* = 2$ where $h_1(z) = h_2(z) = 1$ then noting that $\epsilon_1 z_1 = (v_1 z_1 + \kappa_{01} x z_1)$ and $\epsilon_2 z_1 = (v_2 z_1 + \kappa_{02} x z_1)$ under the assumptions made it is straightforward to show

$$\text{cor}(g_1(\beta_0), g_3(\beta_0)) = \text{cor}(\epsilon_1 z_1, \epsilon_2 z_1) = \frac{\rho + \kappa_{01} \kappa_{02} \sigma_x^2}{\sqrt{(1 + \kappa_{01}^2 \sigma_x^2)(1 + \kappa_{02}^2 \sigma_x^2)}} \quad (40)$$

and since z_1, z_2 are independent then $\text{cor}(g_1(\beta_0), g_2(\beta_0)) = \text{cor}(g_3(\beta_0), g_4(\beta_0)) = 0$. As σ_x^2 , (and/or $|\kappa_{01}|, |\kappa_{02}|$) increase and/or as $|\rho|$ approaches 1 with $|\kappa_{01} - \kappa_{02}|$ approaches zero then $\text{cor}(g_1(\beta_0), g_3(\beta_0))$ approaches 1.

Suppose we construct the moment function including the variable x in y_1 ,

$$\bar{g}(\beta) = \bar{\epsilon}(\beta) \otimes \phi(z) \quad (41)$$

where $\bar{\epsilon}(\beta) := (\bar{\epsilon}_1(\beta), \epsilon_2(\beta))$ for $\bar{\epsilon}_1(\beta) = (y_1 - \theta x_1 - \kappa x)$, where $\bar{\epsilon}_1(\beta_0) = \epsilon_1$ then

$$\text{cor}(\bar{g}_1(\beta_0), \bar{g}_3(\beta_0)) = \text{cor}(\epsilon_1 z_1, \epsilon_2 z_1) = \frac{\rho}{\sqrt{(1 + \kappa_{02}^2 \sigma_x^2)}} \quad (42)$$

Suppose $\rho = 1$ and $|\kappa_{01} - \kappa_{02}| = 0$ where $\kappa_{02} \neq 0$ so moment $g(\beta_0)$ has a singular variance matrix, though $\bar{g}(\beta_0)$ will have a non-singular variance matrix, where correlation between $\bar{g}_1(\beta_0), \bar{g}_3(\beta_0)$ decreases as $\kappa_{02}\sigma_x^2$ increases. Hence this modification to the moment function has removed singularity in the moment variance matrix. This example though somewhat unrealistic, as singular variance arises from perfect residual correlation, allows the formulation of a simple simulation to demonstrate the impact of changing the variable set where omitted variables may be a (part) cause of the singularity of Ω_n . Appendix C provides an extension to this example provided here with $J = 4$ and further unobservable shocks, providing a more realistic situation when common omitted variables can lead to singularity.

STOCHASTIC SINGULARITY

Another method common used in DSGE models is to include further artificial shocks, for example independent random normal variables, in to the the dependent variables and/or directly to the moment function e.g Ruge-Murcia (2007). Again for simplicity consider $J = 2, m^* = 2, h_1(z) = z_2, h_2(z) = z_1, \kappa_0 = 0$.

$$\text{cor}(g_2(\beta_0), g_3(\beta_0)) = \rho \quad (43)$$

Let $(\xi_1, \xi_2)' \stackrel{i.i.d}{\sim} N(0, \sigma_\xi^2)$, then define $\epsilon^*(\beta) = (\epsilon_1^*(\beta), \epsilon_2^*(\beta))'$ where $\epsilon_1^*(\beta) = \epsilon_1(\beta) + \xi_1, \epsilon_2^*(\beta) = \epsilon_2(\beta) + \xi_2$ and the transformed moment function

$$g^*(\beta) = \epsilon^*(\beta) \otimes \phi(z) \quad (44)$$

Then $\Omega_n^* := E[g^*(\beta_0)g^*(\beta_0)']$ which given ξ is i.i.d and independent to all entries then $\Omega_n^* = \Omega_n + \sigma_\xi^2 I_4$ (since $E[\phi(z)\phi(z)'] = I_2$) and $g(\beta_0)^*$ has a non-singular variance matrix at β_0 even when Ω_n is singular. As σ_ξ^2 increased the smallest eigenvalues of Ω_n^* increase. In essence the addition of extra noise serves to regularise Ω_n from within. To see the impact on the correlation between moment functions note that,

$$\text{cor}(g_2^*(\beta_0), g_4^*(\beta_0)) = \text{cor}((\epsilon_1 + \xi_1)z_1, (\epsilon_2 + \xi_2)z_2) = \frac{\rho}{1 + \sigma_\xi^2} \quad (45)$$

Where $\text{cor}(g_1^*(\beta_0), g_2^*(\beta_0)) = \text{cor}(g_3^*(\beta_0), g_4^*(\beta_0)) = 0$.

Note that if $\sigma_\xi^2 > 0$ then $|\text{cor}(g_2^*(\beta_0), g_4^*(\beta_0))| < 1$ and is decreasing in σ_ξ^2 for any configuration of parameters β_0 . The addition of the shock ξ does not enter the expected first order derivative matrix G . When Ω_n is non-singular, then so is Ω_n^* standard asymptotic theory holds where the asymptotic variance is increase in σ_x^2 . However if Ω_n is (almost) singular, then the distribution of

2-Step GMM based on $g(\beta)$ shown Theorem 1 is highly non-standard (with potential non-standard rate of convergence), whereas 2-Step GMM based on $g^*(\beta)$ has a limit normal distribution with the usual asymptotic variance matrix.

If Ω_n is (almost) singular in light of Theorem 1 a Gaussian approximation to the distribution of the 2-step GMM be poor even for large sample sizes, which is demonstrated in Grant (2014) and in the simulation below. We would expect as σ_ξ^2 increases that a Gaussian limit distribution better approximate the small sample distribution of 2-Step GMM.

However two things to note - as σ_ξ^2 increases so does the asymptotic variance. Also the rate of convergence will slow down if $\mathcal{N}(\Omega) \not\subseteq \mathcal{N}(GG')$. Hence we may expect 2-Step GMM based on $g(\beta)$ be closer to the true parameter for any given n relative to that based on $g^*(\beta)$. This modification often used in practise, though serves to render the usual asymptotic theory valid, will to lead inefficient inference and a potentially slower rate of convergence in light of Theorem 1. In sum- taking in to account singularities in to the limit distribution as discussed in Section 4 such that asymptotically inference based on 2-Step GMM using $g(\beta)$ is valid despite potential singularities in Ω_n and does not suffer asymptotic losses of efficiency. This issue is currently unknown in the literature. A simulation experiment studies this in Section 6.

6 Simulation

This section considers a specific example of a linear IV example set out Section 5.

$$y_1 = x_1 + \epsilon_1 \quad y_2 = 0.5x_2 + \epsilon_2$$

$$x_1 = 1.3 * z_1 - 0.8 * z_2 + \eta_1 \quad x_2 = 1.5 * z_2 + \eta_2$$

Where $z = (z_1, z_2)' \stackrel{i.i.d}{\sim} N(0, I_2)$,

$$\epsilon_1 = v_1 z_1, \quad \epsilon_2 = v_2 z_2$$

$$v_1 = \sqrt{\frac{1+\rho}{2}}\zeta_1 + \sqrt{\frac{1-\rho}{2}}\zeta_2, \quad v_2 = \sqrt{\frac{1+\rho}{2}}\zeta_1 - \sqrt{\frac{1-\rho}{2}}\zeta_2$$

such that $cor(v_1, v_2) = \rho$.

$$(\zeta_1, \zeta_2, \eta_1, \eta_2)' | z \stackrel{i.i.d}{\sim} N(0_4, \Pi) \quad \Pi = \begin{pmatrix} 1 & 0 & 0.3 & 0 \\ 0 & 1 & 0.5 & 0 \\ 0.3 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 1 \end{pmatrix}$$

Let $\hat{\beta} := (\hat{\beta}_1, \hat{\beta}_2)'$ be the 2-Step GMM estimator with $W = \hat{\Omega}(\tilde{\beta})^{-1}$ where $\tilde{\beta}$ is a first-step GMM estimator with $W_n = I_{2 \times 2}$.

$\epsilon(\beta) = (y_1 - x_1\beta_1, y_2 - x_2\beta_2)'$ where $\beta := (\beta_1, \beta_2)'$, $\beta_1, \beta_2 \in \mathbb{R}$.

$$g(\beta) = \epsilon(\beta) \otimes z \quad (46)$$

As noted in Section 4- the rate of convergence depends crucially on $P_0'G$. In this example,

$$\Omega = \begin{pmatrix} 3 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 3 \end{pmatrix} \quad G = - \begin{pmatrix} E[x_1z_1] & 0 \\ E[x_1z_2] & 0 \\ 0 & E[x_2z_1] \\ 0 & E[x_2z_2] \end{pmatrix} \quad (47)$$

where $P_0 = (0, -\sqrt{2}, \sqrt{2}, 0)'$ (since in this case the 2 and 3rd moment in $g(\beta_0)$ are perfectly positively correlated so that $G'P_0 = \sqrt{2} \begin{pmatrix} -E[x_1z_2] \\ E[x_2z_1] \end{pmatrix}$). If $E[x_1z_2] = E[x_2z_1] = 0$ then $G'P_0 = 0$ and the 2-step GMM estimator is $O_p(n^{-1/2})$. If either $E[x_1z_1]$, $E[x_2z_1]$ do not equal zero then in some directions $\sqrt{n}(\hat{\beta} - \beta_0)$ converges at rate $n^{1/2}$. We set $E[x_2z_1] = 0$ and let $E[x_1z_2]$ vary so that $B = I_2$.

We consider the case of exact singularity $\rho = 1^7$ for $n = \{500, 5000, 50000, 50000\}$ to verify the results of this paper⁸.

6.1 Simulation Results

The simulation is split in to three parts- firstly we verify the results of Theorem 1 and 2 for the example above. Then the finite sample properties of the 2-step GMM/Wald Statistic are studied when removing singularity by changing the instrument set and stochastic singularity as discussed above.

6.1.1 Simulation 1: Small Sample Properties with Singular Variance

The small sample distribution 2-step GMM estimator $b = \sqrt{n}(\hat{\beta} - \beta_0)$, $b^* = \hat{V}(\tilde{\beta})^{-1/2}b$ and the Wald Statistic $W = b^*b^*$ are studied for and the main results of Theorem 1 and 2 are verified. When Ω is non-singular $b^* = (b_1^*, b_2^*)'$ should be standard normal. Figures 1 and 2 demonstrate that b^* is highly

⁷A related simulation experiment was studied in Grant (2013) for the case of Weakly Singular Variance with $\delta < 1$ and $\Lambda_0 > 0$, demonstrating the non-standard rate of convergence in some directions of $n^{(1+\delta)/2}$ of the 2-Step GMM estimator when $G'P_0 \neq 0$. In this case of WSV 2-Step GMM and the Wald Statistic have the standard normal and χ_m^2 limit distribution which is verified in the simulation.

⁸We also consider $\rho = 1 - n^{-1}$ and find similar results as expected by Theorem 1 given the non-standard rate of convergence of 2-Step GMM and it's limit distribution hold for such sequences also, these results are omitted for brevity.

non-standard as expected by Theorem 1. Finally Figure 3 shows the non-standard limit of the Wald Statistic, which in the case of Ω non-singular is asymptotically χ_2^2 in this example. Notably the small sample distribution of W even for $n = 500000$ is oversized. As such rejection frequencies based on a Wald Test of $\beta_0 = c(1, 0.5)$ based on quantiles of a χ_2^2 are also oversized as noted in Table 1 below.

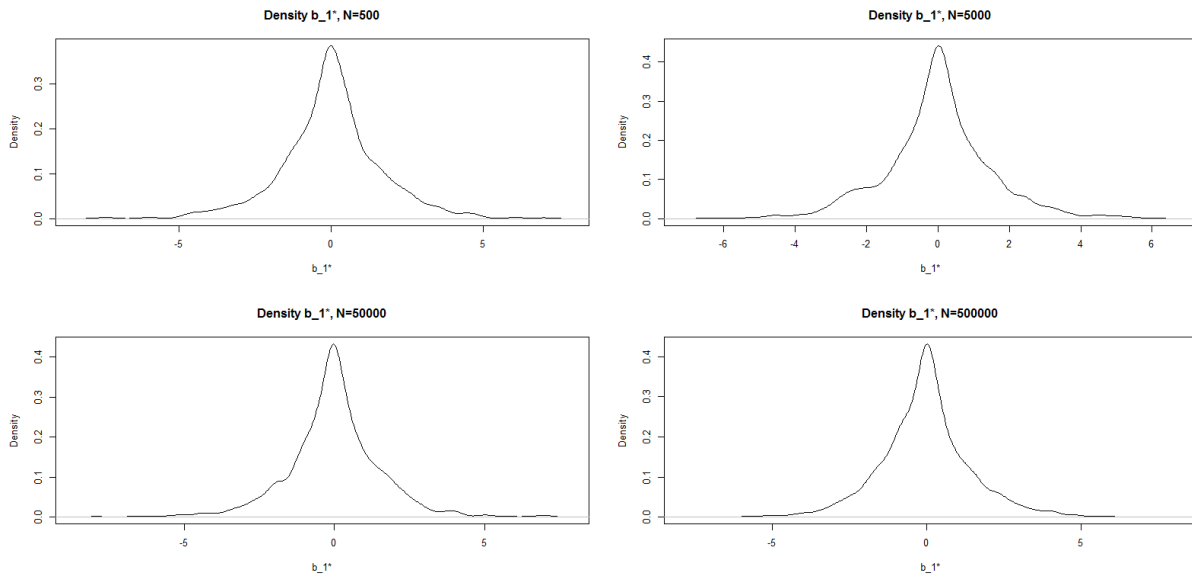


Figure 1: Density of b_1^*

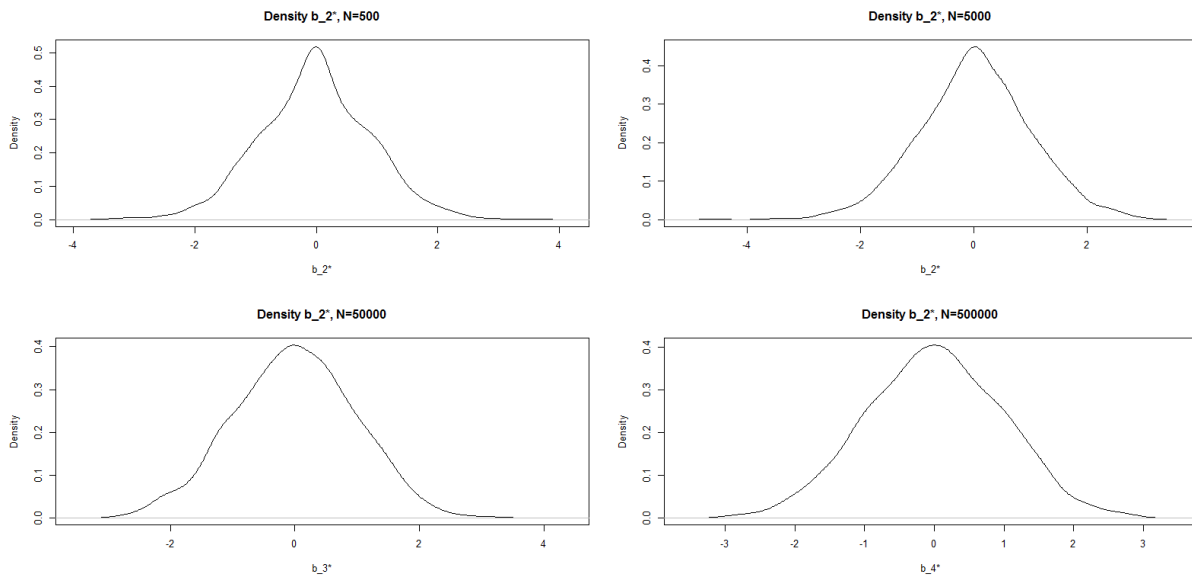


Figure 2: Density of b_2^*

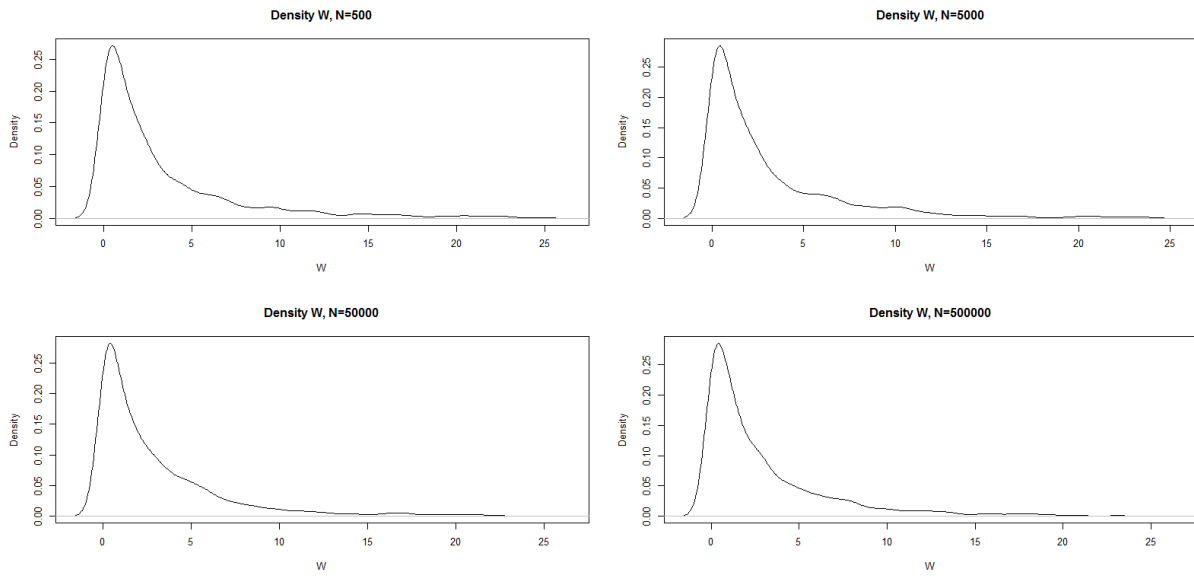


Figure 3: Density of W

6.1.2 Simulation 2: Stochastic Singularity

Simulation 2 considers the impact of introducing a perturbation to the moment $g(\beta)$ by $\sqrt{s}\xi$ where $\xi \stackrel{i.i.d}{\sim} N(0, I_4)$ for $s = \{\sqrt{0.1}\sqrt{0.5}\}$. As noted in Section 5 the transformed moment is now nonsingular, where the smallest eigenvalue increases as s increases.

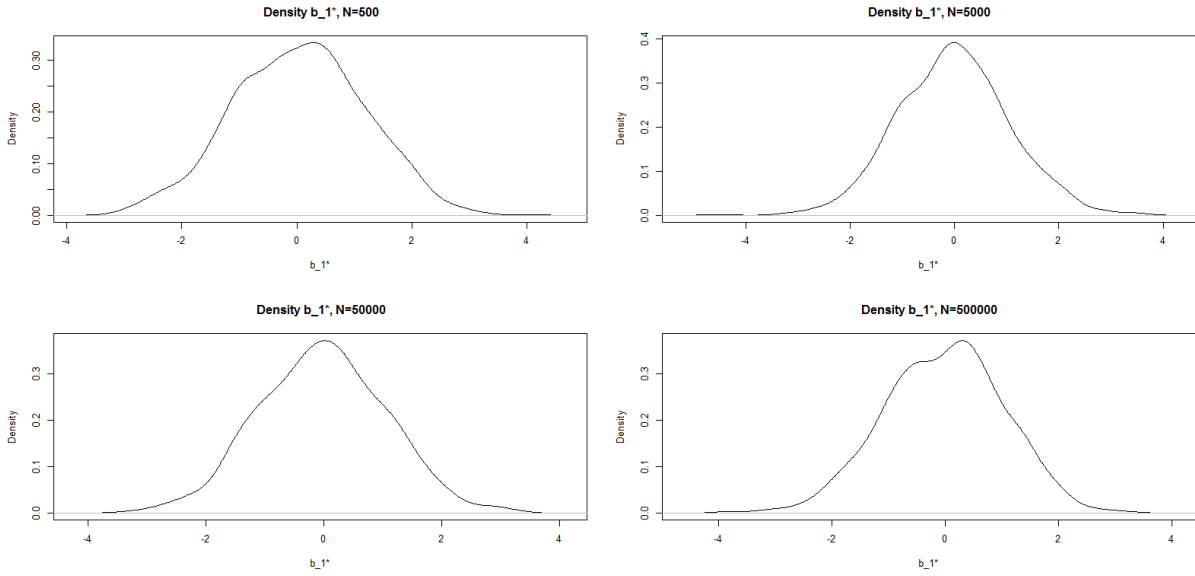


Figure 4: Simulation 2(a): Density of b_1^* $s = \sqrt{0.1}$

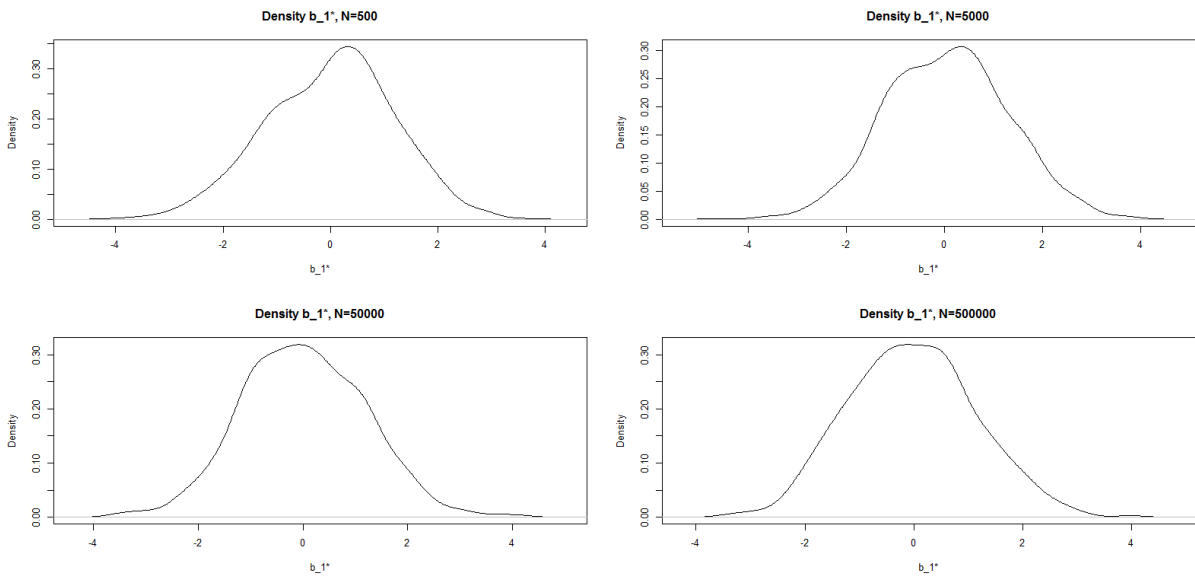


Figure 5: Simulation 2(b): Density of b_1^* $s = \sqrt{0.5}$

Table 1 summarises key results on the distribution of b, b^*, W .

6.1.3 Simulation 3: Changing the Instrument Set

Simulation 3 considers the impact on the small sample distribution of b, b^* and W of removing the singularity by changing the instrument set. Namely we drop the instrument z_1 . In this case the singularity is removed and now moments are just identified. Since G is full rank then b, b^*, W should satisfy standard asymptotics when Ω is non-singular, e.g Hansen (1982). Figure 4 demonstrates that for large n the distribution of b_1 is better approximated by a normal distribution relative simulation 1. For brevity we omit the distributions of b_2^* and W , where summary of the properties of the distribution are provided in Table 1.

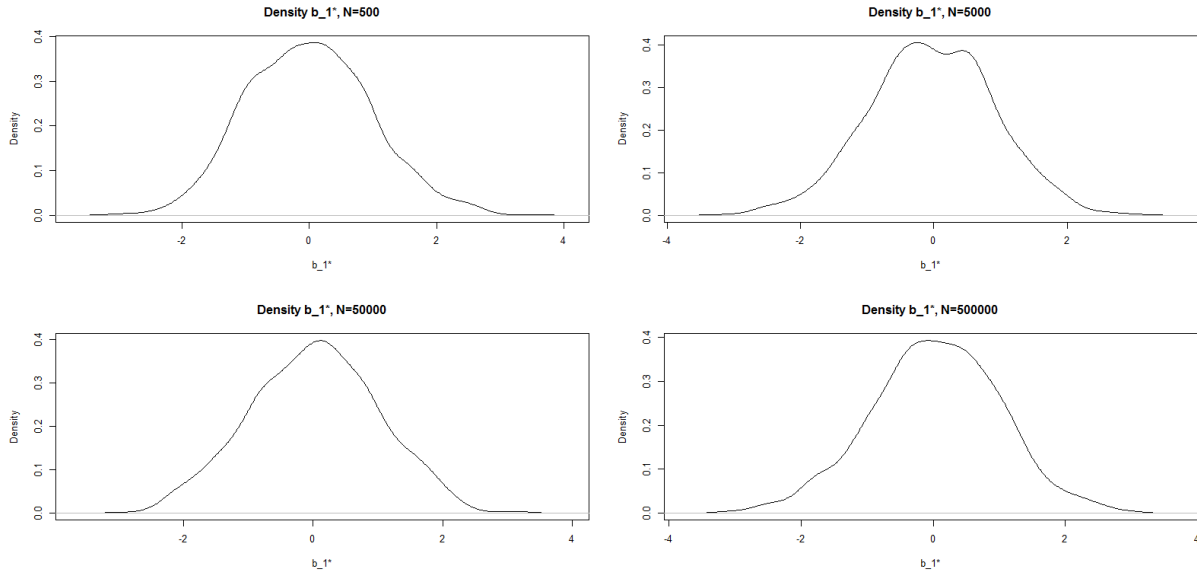


Figure 6: Simulation 2: Density of b_1^* removing instrument z_1 .

Table 1 below summarises the key properties of the 2-Step GMM estimator in the three simulations above. Monte Carlo estimates ($R = 2000$) of the variance of b and the rejection frequencies from a Wald Test that $\beta_0 = (1, 0.5)'$ based on the quantiles of a χ_2^2 .

First to note that in Sim 1 then 2-Step GMM with singular variance is such that the variance of b_1 is $O(n^{-1})$ which is as expected given $n^{1/2}b_1 = O_p(1)$ by Theorem 1 since in this example $B_n = \begin{pmatrix} n^{1/2} & 0 \\ 0 & 1 \end{pmatrix}$ where $b_2 = O_p(1)$ as seen in Sim 1 below. As expected given the non-standard distribution of the Wald Statistic when $\bar{m} > 0$ the size of the Wald Test is oversized even for $n = 500000$.

Both Sim 2 and Sim 3 have performed a modification which remove the singularity in the moment function. As such both b_1, b_2 should be $O_p(1)$ and the Wald Statistic have correct size for n large. The variance of b_1 does not shrink at rate n but is bounded away from zero. Also the variance of b_2 has increased. In both Sim 2 and Sim 3 size of the Wald Test is closer to the nominal level for all n relative to Sim 1.

	Efficiency		Limit Distribution			
	Var(b_1)	Var(b_2)	Wald-Size(90%)	Wald-Size (95%)	Wald-Size(99%)	
Sim 1	$n = 500$	$1.40 \cdot 10^{-2}$	1.18	0.230	0.176	0.102
	$n = 5000$	$1.27 \cdot 10^{-3}$	1.25	0.223	0.168	0.088
	$n = 50000$	$1.18 \cdot 10^{-4}$	1.17	0.222	0.151	0.079
	$n = 500000$	$1.15 \cdot 10^{-5}$	1.15	0.219	0.163	0.081
Sim 2(a)	$n = 500$	0.289	1.23	0.141	0.065	0.016
	$n = 5000$	0.256	1.29	0.125	0.053	0.017
	$n = 50000$	0.269	1.31	0.114	0.070	0.019
	$n = 500000$	0.257	1.16	0.121	0.054	0.018
Sim 2(b)	$n = 500$	0.807	1.52	0.121	0.066	0.014
	$n = 5000$	0.855	1.54	0.132	0.075	0.021
	$n = 50000$	0.816	1.48	0.116	0.07	0.022
	$n = 500000$	0.779	1.57	0.12	0.062	0.015
Sim 3	$n = 500$	1.57	1.32	0.114	0.062	0.014
	$n = 5000$	1.39	1.34	0.082	0.040	0.011
	$n = 50000$	1.53	1.36	0.096	0.049	0.008
	$n = 500000$	1.52	1.34	0.098	0.054	0.013

7 Concluding Remarks

This paper considers efficient GMM estimation in strongly identified linear models when moments have (asymptotically) singular variance. We derive the non-standard limit distribution of 2-Step GMM and related statistics. Discussion is provided which discusses how regularising the variance matrix, a method commonly applied in applied and theoretical research on moment type estimators leads to an estimator with a larger asymptotic variance and a potentially slower rate of convergence. Namely we show that the rate of convergence is shown to be rate- n when the null space of the moment variance matrix is not a subset or equal to the null space of the outer product of the expected first order derivative at β_0 .

A common intuition in the literature is that those linear combination of moments with zero variance provide no informative content in estimation of β_0 , e.g Doran & Schmidt (2006). Results in this paper

show in fact the opposite is true and that regularising the variance matrix can lead to a less efficient estimator. This opens up the question of what is the efficient limiting variance and rate of convergence when we remove the assumption that the moment variance (after a regularisation) is asymptotically non-singular. This paper argues that the assumption on the form and ran of the moment variance be made as an identification condition, similar the the rank assumption on the expected first order moment derivative.

Results in this paper were based on some simplifying assumptions. Relaxing the assumption eigenvalues are simple will allow for the rank of the variance matrix to take any value. The author intends to work on this in future research. Considering the case of weak/lack of identification would also be interesting- given the increase of rate of convergence in certain cases with singular variance in the strongly identified case. This paper served as an initial foray in to the properties of moment estimator with singular variance and extensions to many weak moments, generalised empirical likelihood estimators and non-linear models may shed further light on the impact of singular variance for inference.

Appendix

Appendix A: Proof of Lemmas

LEMMA 1

$$\sqrt{n}\hat{\Omega}_T(\tilde{\beta})P_{0T}\Xi_T \xrightarrow{d} \Psi \quad (48)$$

where $\Psi = (I_m \otimes Z'_g)\Theta P_0$

PROOF OF LEMMA 1

By a Taylor expansion of $\hat{\Omega}(\tilde{\beta})$ around β_0 and under the Linearity assumption,

$$\hat{\Omega}(\tilde{\beta}) = \hat{\Omega} + \frac{1}{n} \sum_{i=1}^n g_i \tilde{\Delta}' G'_i + \frac{1}{n} \sum_{i=1}^n G_i \sqrt{n} \tilde{\Delta} g'_i + \frac{1}{n} \sum_{i=1}^n G_i \tilde{\Delta} \tilde{\Delta}' G'_i \quad (49)$$

Multiply the Left Hand Side (LHS) of (49) by P_0 and define $c_{i,n} = n^{\delta/2} P'_0 g_i$

$$\sqrt{n}\hat{\Omega}(\tilde{\beta})P_0 = n^{1/2}\hat{\Omega}P_0 + \frac{1}{n} \sum_{i=1}^n g_i \sqrt{n} \tilde{\Delta}' G'_i P_0 + \frac{1}{n} \sum_{i=1}^n G_i \tilde{\Delta} c'_{i,n} + \frac{1}{n} \sum_{i=1}^n G_i \sqrt{n} \tilde{\Delta} \tilde{\Delta}' G'_i P_0 \quad (50)$$

Firstly is is straightforward to show

$$\frac{1}{n} \sum_{i=1}^n \sqrt{n} G_i \tilde{\Delta} \tilde{\Delta}' G'_i P_0 \leq \sqrt{n} \|P_0\| \left\| \frac{1}{n} \sum_{i=1}^n \|G_i\|^2 \right\| \|\tilde{\Delta}\|^2 = O_p(n^{-1/2}) \quad (51)$$

$$\frac{1}{n} \sum_{i=1}^n G_i \tilde{\Delta} c'_{i,n} \leq \|\tilde{\Delta}\| \frac{1}{n} \sum_{i=1}^n \|G_i\| \frac{1}{n} \sum_{i=1}^n \|c_{i,n}\| = O_p(n^{-1/2}) \quad (52)$$

$$\sqrt{n} \hat{\Omega}(\tilde{\beta}) P_0 = \frac{1}{n} \sum_{i=1}^n g_i \sqrt{n} \tilde{\Delta}' G'_i P_0 + O_p(n^{-1/2}) \quad (53)$$

Using Theorem $\sqrt{n} \tilde{\Delta} = \Xi \sqrt{n} \hat{g} + o_p(1)$ then the jk 'th element of $\frac{1}{n} \sum_{i=1}^n g_i \sqrt{n} \tilde{\Delta}' G'_i$,

$$\left[\frac{1}{n} \sum_{i=1}^n g_i \sqrt{n} \tilde{\Delta}' G'_i \right]_{jk} = \frac{1}{n} \sum_{i=1}^n g_{ij} \sqrt{n} \hat{g}' \Xi' G_{ik} = \sqrt{n} \hat{g}' \hat{\theta}_{jk} \quad (54)$$

Define $\hat{\theta}_{jk} := \frac{1}{n} \sum_{i=1}^n \Xi' G_{ik} g_{ij}$ hence

$$\frac{1}{n} \sum_{i=1}^n g_i \sqrt{n} \tilde{\Delta}' G'_i = (I_m \otimes \sqrt{n} \hat{g}') \hat{\Theta} \quad (55)$$

Where $\hat{\Theta}$ is defined,

$$\hat{\Theta} := \begin{bmatrix} \hat{\theta}_{11} & \cdots & \hat{\theta}_{1m} \\ \vdots & \ddots & \vdots \\ \hat{\theta}_{m1} & \cdots & \hat{\theta}_{mm} \end{bmatrix} \quad (56)$$

and $\hat{\theta}_{jk} \xrightarrow{p} \theta_{jk}$ where $\theta_{jk} = \lim_{n \rightarrow \infty} E \left[\frac{1}{n} \sum_{i=1}^n \Xi' G_{ik} g_{ij} \right]$ for all $j, k = \{1, \dots, m\}$ by the WLLn for Triangular Arrays.

$$\hat{\Theta} \xrightarrow{p} \Theta \quad (57)$$

and by the Lindberg-Feller Central Theorem under A1(i),(ii) then $\sqrt{n} \hat{g}' \xrightarrow{d} Z'_g$ then together (??), (57) by Slutskys Theorem establishes (16)

$$\frac{1}{n} \sum_{i=1}^n g_i \sqrt{n} \tilde{\Delta}' G'_i P_0 \xrightarrow{d} (I_m \otimes Z'_g) \Theta P_0 \quad (58)$$

Establishing the result,

$$\sqrt{n} \hat{\Omega}(\tilde{\beta}) P_0 \xrightarrow{p} (I_m \otimes Z'_g) \Theta P_0 \quad (59)$$

LEMMA 2 Under Assumption

$$\Xi_T P'_{0T} \hat{\Omega}_T(\tilde{\beta}) P_{0T} \Xi_T \xrightarrow{d} \mathcal{W}$$

PROOF: Taylor expanding $\hat{\Omega}_T(\tilde{\beta}_T)$ around β_0

$$P'_{0T}\hat{\Omega}_T(\tilde{\beta}_T)P_{0T} = P'_{0T}\hat{\Omega}_TP_{0T} + \frac{1}{T}\sum_{t=1}^T c_t\tilde{\Delta}'_TG'_tP_{0T} + \frac{1}{T}\sum_{t=1}^T P'_{0T}G_t\tilde{\Delta}_Tc'_t + \frac{1}{T}\sum_{t=1}^T P'_{0T}G_t\tilde{\Delta}_T\tilde{\Delta}'_TG'_tP_{0T}$$

utilising linearity of the moment function where $c_t = P'_{0T}g_t$ where dependence on T is suppressed and defining $\tilde{\Delta}_T = \tilde{\beta}_T - \beta_0$. Noting that $c_t = \Lambda_{0T}^{1/2}\bar{c}_t$ where $\Lambda_{0T} = \text{diag}(\lambda_{m-\bar{m},T}, \dots, \lambda_{m,T})$ where $E\left[\frac{1}{T}\sum_{t=1}^T \bar{c}_t\bar{c}'_t\right] = I_{\bar{m}}$ and $P'_{0T}\Omega_TP_{0T} = E\left[\frac{1}{T}\sum_{t=1}^T c_t c'_t\right] = \Lambda_{0T}$. Define $\tilde{\Delta}_T^S = S'_T\tilde{\Delta}_T$ where $\tilde{\Delta}_T^S \xrightarrow{d} CZ_g$ where $C = (G'WG)^{-1}G'W\Omega^{1/2}Z_g$

$$\begin{aligned} \Xi_T \frac{1}{T} \sum_{t=1}^T P'_{0T}G_t\tilde{\Delta}_Tc_t &= \mu_{pT}\Xi_T \frac{1}{T} \sum_{t=1}^T P'_{0T}G_t\bar{S}_T^{-1}\text{diag}(\mu_{pT}\mu_{1T}^{-1}, \dots, 1)\tilde{\Delta}_T^S\bar{c}_t\Lambda_{0T}^{1/2}\Xi_T \\ &\xrightarrow{d} \Xi_\mu \frac{1}{T} \sum_{t=1}^T ()\Xi_\lambda \end{aligned}$$

where $\mu_{pT}\Xi_T = \text{diag}(\mu_{pT}\xi_{1T}, \dots, \mu_{pT}\xi_{\bar{m}T}) \rightarrow \text{diag}(\xi_{1,\mu}, \dots, \xi_{\bar{m},\mu})$ and $\xi_{j,\mu} = \lim_{T \rightarrow \infty} \mu_{pT}\xi_{jT}$ for $j = \{1, \dots, \bar{m}\}$ and $\xi_{j,\mu} = \lim_{T \rightarrow \infty} \mu_{pT}/\lambda_{m-\bar{m}-1+j,T}^{1/2}$ for $j = \{1, \dots, k\}$ $\xi_{j,\mu} = 1$ for $j = \{k+1, \dots, \bar{m}\}$ and $\Lambda_{0T}^{1/2}\Xi_T = \text{diag}(\lambda_{m-\bar{m},T}^{1/2}\xi_{1T}, \dots, \lambda_{m,T}^{1/2}\xi_{\bar{m}T}) \rightarrow \text{diag}(\xi_{1,\lambda}, \dots, \xi_{\bar{m},\lambda})$ where $\xi_{j,\lambda} = 1$ for $j = \{1, \dots, k\}$ and $\xi_{j,\lambda} = 0$ for $j = \{k+1, \dots, \bar{m}\}$.

Again using $\sqrt{n}\tilde{\Delta}' = \sqrt{n}\Xi\hat{g}' + o_p(1)$ then $[\frac{1}{n}\sum_{i=1}^n c'_{i,n}\sqrt{n}\tilde{\Delta}'G'_i]_{jk} = \sqrt{n}\hat{g}'\Xi'\frac{1}{n}\sum_{i=1}^n G_{ik}c_{ij} + o_p(1)$.

Define $\hat{v}_{jk} = \Xi'\frac{1}{n}\sum_{i=1}^n G_{ik}c_{ij}$, then

$$\frac{1}{n}\sum_{i=1}^n c'_{i,n}\sqrt{n}\tilde{\Delta}'G'_i = (I_m \otimes \sqrt{n}\hat{g})\hat{\Upsilon} + o_p(1) \quad (60)$$

Where $\hat{\Upsilon}$ is defined similarly to Υ replacing v_{jk} with \hat{v}_{jk} for all $j, k = \{1, \dots, m\}$ where $\hat{\Upsilon} \xrightarrow{p} \Upsilon$ which establishes (??). Finally to show,

$$P'_0 \frac{1}{n} \sum_{i=1}^n G_i n \tilde{\Delta} \tilde{\Delta}' G'_i P_0 \xrightarrow{d} P'_0 \Gamma (I_m \otimes Z_g \otimes Z_g) P_0 \quad (61)$$

Define $\hat{\gamma}_{jk} := \Xi'\frac{1}{n}\sum_{i=1}^n G_{ik}G'_{jk}\Xi$ then note

$$\left[\frac{1}{n}\sum_{i=1}^n G_i n \tilde{\Delta} \tilde{\Delta}' G'_i\right]_{jk} = \frac{1}{n}\sum_{i=1}^n G'_{ij} n \tilde{\Delta} \tilde{\Delta}' G_{ij} = \text{tr}(\hat{\gamma}'_{jk} n \hat{g} \hat{g}') = \text{vec}(\hat{\gamma}_{jk})' \text{vec}(n \hat{g} \hat{g}') \quad (62)$$

Using the results that for two square matrices A,B $\text{tr}(AB) = \text{vec}(A)'\text{vec}(B)$. Noting the fact $a \in \mathbb{R}^q$ $\text{vec}(aa') = a \otimes a$ then

$$\frac{1}{n}\sum_{i=1}^n G_i n \tilde{\Delta} \tilde{\Delta}' G'_i = \hat{\Gamma} (I_m \otimes \sqrt{n}\hat{g} \otimes \sqrt{n}\hat{g}) + o_p(1) \quad (63)$$

Where $\hat{\Gamma}$ is defined similarly to Γ in (8) replacing Γ_{jk} with $\hat{\Gamma}_{jk}$ for $j, k = \{1, \dots, m\}$ where $\hat{\Gamma} \xrightarrow{P} \Gamma$ which together imply,

$$P'_0 \frac{1}{n} \sum_{i=1}^n G_i n \tilde{\Delta} \tilde{\Delta}' G_i' P_0 \xrightarrow{d} P_0 \Gamma(I_m \otimes Z_g \otimes Z_g') P_0 \quad (64)$$

Finally by (??),

$$nP'_0(\hat{\Omega}(\tilde{\beta}) - \hat{\Omega})P_0 \xrightarrow{d} P'_0((I_m \otimes Z_g')\Upsilon + \Upsilon'(I_m \otimes Z_g))P_0 + P'_0\Gamma(I_m \otimes Z_g \otimes Z_g)P_0 \quad (65)$$

$$nP'_0(\hat{\Omega})P_0 \xrightarrow{P} \Lambda_0 \quad (66)$$

Establishing the result

$$nP'_0\hat{\Omega}(\tilde{\beta})P_0 \xrightarrow{d} \mathcal{W} \quad (67)$$

Define

LEMMA 3: Under A1 when $\bar{m} > 0$ then for any $\lambda_1, \lambda_2 \in \mathbb{R}^{k \times m}$ where $k \geq 1$ defining $W_1 = (I - \hat{\Delta}_\Omega M_0)\lambda_1$, $W_2 = (I - \hat{\Delta}_\Omega M_0)\lambda_2$

$$\lambda_1' \hat{\Omega}(\tilde{\beta})^{-1} \lambda_2 = \lambda_1' M_0 \lambda_2 + W_1' \hat{\Pi} W_2 + O_p(n^{-1/2} \|W_1\| \|W_2\|). \quad (68)$$

PROOF OF LEMMA 1

Let A be an $m \times m$ singular matrix and B some $m \times m$ (asymptotically) bounded matrix. Define $A(z) = A + zB$ for some scalar z . By Theorem 2.10 and Theorem 2.11 (pages 22-25) of Avrachenkov et al. (2013) for all $z \rightarrow 0$ then if $A(z)^{-1}$ exists in a punctured neighbourhood around $z = 0$ and B satisfies $\det(V_0' B V_0) \neq 0$ w.p.1. for V_0 such that $AV_0 = 0$

$$A(z)^{-1} = \frac{1}{z} X_0 + X_1 - z X_1 B X_1 + z^2 X_1 B X_1 B X_1 + \dots \quad (69)$$

where $X_0 = V_0 [V_0' B V_0]^{-1} V_0'$, $X_1 = (I - X_0 B) A^* (I - B X_0)$ and A^* is the Moore-Penrose generalised inverse of A .

Applying this result to $\hat{\Omega}(\tilde{\beta})$ where $A = \Omega$ and $V_0 = P_0$ (the expansion is invariant to right multiplication by a full rank matrix), $B = \hat{\Delta}_\Omega$ setting $z = 1$ noting that $\|B\| \xrightarrow{P} 0$. Finally $X_0 = M_0$, $X_1 = M_1$ where $M_1 = (I - M_0 \hat{\Delta}_\Omega) \Omega^* (I - \hat{\Delta}_\Omega M_0)$ and noting that $V_0' B V_0 = P_0' \hat{\Omega}(\theta_n) P_0$ is full rank w.p.1 by A1(vii). Then $\hat{\Omega}(\tilde{\beta})^{-1}$ has a Laurent Series expansion

$$\hat{\Omega}(\tilde{\beta})^{-1} = M_0 + M_1 - M_1 \hat{\Delta}_\Omega M_1 + M_1 \hat{\Delta}_\Omega M_1 \hat{\Delta}_\Omega M_1 + \dots \quad (70)$$

plugging the relevant entries above in to (69).

Define $M_2 = (I - \hat{\Delta}_\Omega M_0) \hat{\Delta}_\Omega (I - M_0 \hat{\Delta}_\Omega) \Omega^*$ then by simple algebraic multiplication we may express

(70)

$$\hat{\Omega}(\tilde{\beta})^{-1} = M_0 + (I - M_0 \hat{\Delta}_\Omega) \Omega^* \sum_{j=0}^{\infty} (-M_2)^j (I - \hat{\Delta}_\Omega M_0) \quad (71)$$

which can be verified by expanding the sum and using the definition of M_1 .

Expanding out M_2 we find

$$M_2 = (\hat{\Delta}_\Omega - 2\hat{\Delta}_\Omega M_0 \hat{\Delta}_\Omega + \hat{\Delta}_\Omega M_0 \hat{\Delta}_\Omega M_0 \hat{\Delta}_\Omega) \Omega^*. \quad (72)$$

Firstly note that since $\hat{\Delta}_\Omega = (\hat{\Omega}(\tilde{\beta}) - \hat{\Omega}) - (\Omega - \hat{\Omega})$ then by T

$$\|\hat{\Delta}_\Omega\| \leq \|\hat{\Omega}(\tilde{\beta}) - \hat{\Omega}\| + \|\hat{\Omega} - \Omega\| \quad (73)$$

where $\|\hat{\Omega}(\tilde{\beta}) - \hat{\Omega}\| = O_p(n^{-1/2})$ by A1(iv) and $\|\hat{\Omega} - \Omega\| = o_p(1)$ by A1(i),(ii) so that since $\|\Omega^*\| \leq \|P_+\|^2 \|\Lambda_+\|^{-1} = O(1)$ as $\|P_+\| = m - \bar{m} < \infty$ and $\Lambda_+ = P_+' \Omega P_+$ where $\|\Omega\| = O(1)$ by A1(ii) then

$$\hat{\Delta}_\Omega \Omega^* = o_p(1). \quad (74)$$

Noting that $M_0 \hat{\Delta}_\Omega M_0 = M_0$

$$\hat{\Delta}_\Omega M_0 \hat{\Delta}_\Omega M_0 \hat{\Delta}_\Omega = \hat{\Delta}_\Omega M_0 \hat{\Delta}_\Omega \quad (75)$$

so together (75) and (74) plugged in to (72) imply

$$-M_2 = \hat{\Delta}_\Omega M_0 \hat{\Delta}_\Omega \Omega^* + o_p(1) \quad (76)$$

$$= n^{1/2} \hat{\Omega}(\theta_n) P_0 [n P_0' \hat{\Omega}(\tilde{\beta}) P_0]^{-1} n^{1/2} P_0' \hat{\Omega}(\tilde{\beta}) \quad (77)$$

where by Lemma 1 and Lemma 2 and that Γ is invertible by A1(viii) an application of CMT establishes

$$n^{1/2} \hat{\Omega}(\theta_n) P_0 [n P_0' \hat{\Omega}(\theta_n) P_0]^{-1} n^{1/2} P_0' \hat{\Omega}(\theta_n) = \hat{\Psi}' \hat{\mathcal{W}}^{-1} \hat{\Psi} \quad (78)$$

which plugged in to (76) implies

$$M_2 = -\hat{\Psi}' \hat{\Gamma} \hat{\Psi} \Omega^* + o_p(1). \quad (79)$$

Finally multiplying (71) on the left and right by λ_1, λ_2 respectively and plugging in $-M_2 = \Psi' \Gamma \Psi \Omega^* + o_p(1)$ and using by definition that $W_1 = (I - \hat{\Delta}_\Omega M_0) \lambda_1, W_2 = (I - \hat{\Delta}_\Omega M_0) \lambda_2$

$$\lambda_1' \hat{\Omega}(\theta_n)^{-1} \lambda_2 = \lambda_1' M_0 \lambda_2 + W_1' \Omega^* \sum_{j=0}^{\infty} (\hat{\Psi}' \hat{\mathcal{W}} \hat{\Psi} \Omega^* + o_p(1))^j W_2 \quad (80)$$

$$= \lambda_1' M_0 \lambda_2 + W_1' \hat{\Pi} W_2 + o_p(W_1' W_2) \quad (81)$$

where (81) follows since

$$\Omega^* \sum_{j=0}^{\infty} (\hat{\Psi}'\Gamma\Psi\Omega^* + o_p(1))^j = \Omega^* + \sum_{j=1}^{\infty} \Omega^* (\Psi'\Gamma^{-1}\Psi\Omega^*)^j + o_p(1) \quad (82)$$

$$= \Omega^* + \Omega^* \Psi' \left(\Gamma^{-1} \sum_{j=0}^{\infty} (\Psi'\Omega^*\Psi\Gamma^{-1})^j \right) \Psi\Omega^* + o_p(1) \quad (83)$$

$$= \Pi + o_p(1) \quad (84)$$

and (84) follows since by repeated application of the Woodbury Formula,

$$\Phi^{-1} = \Gamma^{-1} \sum_{j=0}^{\infty} (\Psi'\Omega^*\Psi\Gamma^{-1})^j \quad (85)$$

since Γ and $\Phi = \Gamma - \Psi'\Omega^*\Psi$ are full rank by A1(viii) and $\Pi = \Omega^* + \Omega^*\Psi'\Phi^{-1}\Psi\Omega^*$ by definition. Finally since $W_1'W_2 \leq \|W_1\|W_2\|$ by CS which together with (81) establishes (68).

Appendix B: Proofs of Main Theorems

To show (30) we use the usual Taylor expansion of the GMM first order condition around β_0 . By definition $\hat{\beta}$ solves the first order condition

$$\hat{G}'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta}) = 0 \quad (86)$$

Then performing a Mean Value Expansion around β_0

$$\sqrt{n}(\hat{\beta} - \beta_0) = -(\hat{G}'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}(\tilde{\beta})^{-1}\sqrt{n}\hat{g}(\beta_0) \quad (87)$$

Define $\hat{\Phi}_1 = (\hat{G}'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G})^{-1}\hat{G}'$, $\hat{\Phi}_2 = \hat{G}'\hat{\Omega}(\tilde{\beta})^{-1}\sqrt{n}\hat{g}(\beta_0)$ then by (87)

$$\sqrt{n}(\hat{\beta} - \beta_0) = \hat{\Phi}_1^{-1}\hat{\Phi}_2 \quad (88)$$

. We can rewrite $\hat{\Phi}_1^{-1}\hat{\Phi}_2 = B_n^{-1}(B_n'^{-1}\hat{\Phi}_1 B_n^{-1})^{-1}B_n'^{-1}\hat{\Phi}_2$ where,

$$B_n\sqrt{n}(\hat{\beta} - \beta_0) = (B_n'^{-1}\hat{\Phi}_1 B_n^{-1})^{-1}B_n'^{-1}\hat{\Phi}_2 \quad (89)$$

We now show that both $B_n'^{-1}\hat{\Phi}_1 B_n^{-1} \xrightarrow{d} \Phi_1$ and $B_n'^{-1}\hat{\Phi}_2$ which establishes (30). By Lemma 3 for $\lambda_1 = \lambda_2 = B_n'^{-1}\hat{G}'$ when W_1 and W_2 are $O_p(1)$ and by Lemma 3

$$\begin{aligned} B_n'^{-1}\hat{\Phi}_1 B_n^{-1} &= B_n'^{-1}\hat{G}'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G}B_n^{-1} \\ &= B_n'^{-1}\hat{G}'M_0\hat{G}B_n^{-1} + B_n'^{-1}\hat{G}'(I - \hat{\Delta}_\Omega M_0)\hat{\Pi}(I - M_0\hat{\Delta}_\Omega)\hat{G}B_n^{-1} + o_p(1) \end{aligned}$$

$$B_n'^{-1}\hat{G}'M_0\hat{G}B_n^{-1} = n^{1/2}B_n'^{-1}\hat{G}'P_0[nP_0'\hat{\Delta}_\Omega P_0]^{-1}n^{1/2}P_0'\hat{G}B_n^{-1} \quad (90)$$

Where $[nP_0'\hat{\Delta}_\Omega(\tilde{\beta})P_0]^{-1} \xrightarrow{d} \Gamma^{-1}$ by CMT and

$$n^{1/2}B_n'^{-1}\hat{G}'P_0 = n^{1/2}B_n'^{-1}(\hat{G} - G)'P_0 + n^{1/2}B_n'^{-1}G'P_0 \quad (91)$$

$$\xrightarrow{d} \begin{pmatrix} 0 & 0 \\ 0 & I_{p-\bar{p}} \end{pmatrix} B'Z_G'P_0 + \begin{pmatrix} I_{\bar{p}} & 0 \\ 0 & 0 \end{pmatrix} B'G'P_0 \quad (92)$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & B'_{p-\bar{p}} \end{pmatrix} Z_G'P_0 + \begin{pmatrix} B'_{\bar{p}} & 0 \\ 0 & 0 \end{pmatrix} G'P_0 \quad (93)$$

$$:= \mathcal{C} \quad (94)$$

where (92) holds since $B'G'P_0 = \begin{pmatrix} G_{BP_0} \\ 0 \end{pmatrix}$ by definition of B and (93) holds noting

$$\begin{pmatrix} I_{\bar{p}} & 0 \\ 0 & 0 \end{pmatrix} B' = \begin{pmatrix} B'_{\bar{p}} & 0 \\ 0 & 0 \end{pmatrix} \text{ where } B'_{\bar{p}} \text{ is the upper } \bar{p} \times \bar{p} \text{ sub-matrix of } B'$$

Together by the CMT establishes

$$B_n'^{-1}\hat{G}'M_0\hat{G}B_n^{-1} \xrightarrow{p} \mathcal{C}'\Gamma^{-1}\mathcal{C} \quad (95)$$

$$\begin{pmatrix} 0 & 0 \\ 0 & I_{p-\bar{p}} \end{pmatrix} B' = \begin{pmatrix} 0 & 0 \\ 0 & B'_{p-\bar{p}} \end{pmatrix} \text{ where } B'_{p-\bar{p}} \text{ is the lower } (p-\bar{p}) \times (p-\bar{p}) \text{ sub-matrix of } B'.$$

We now show that

$$B_n'^{-1}\hat{G}'(I - M_0\hat{\Delta}_\Omega)\Pi(I - \hat{\Delta}_\Omega M_0)\hat{G}B_n^{-1} = \quad (96)$$

$$B_n'^{-1}\hat{G}'(I - M_0\hat{\Delta}_\Omega) = B_n'^{-1}\hat{G}' - B_n'^{-1}\hat{G}'M_0\hat{\Delta}_\Omega \quad (97)$$

Where

$$B_n'^{-1}\hat{G} \xrightarrow{p} \begin{pmatrix} 0 & 0 \\ 0 & B'_{p-\bar{p}} \end{pmatrix} G' \quad (98)$$

$$= \bar{B}'G' \quad (99)$$

$$B_n'^{-1}\hat{G}'M_0\hat{\Delta}_\Omega = n^{1/2}B_n'^{-1}\hat{G}'P_0[nP_0'\hat{\Delta}_\Omega P_0]^{-1}n^{1/2}P_0'\hat{\Delta}_\Omega \quad (100)$$

$$\xrightarrow{d} \mathcal{C}'\Gamma^{-1}\Psi' \quad (101)$$

Hence

$$B_n'^{-1} \hat{\Phi}_1 B_n^{-1} \xrightarrow{d} C' \Gamma^{-1} C + (G\bar{B} - \Psi' \Gamma^{-1} C)' \Pi (G\bar{B} - \Psi' \Gamma^{-1} C) \quad (102)$$

Appendix C: Examples of Singular Variance

The example used in Section X to demonstrate the impact of common omitted variables was somewhat unrealistic as it required the residuals from the regression equations be perfectly correlated. As the number of equations, J increases and with further omitted shocks- both observable and unobservable, singularity may arise in situations residuals in the regression functions are not perfectly correlated. Take the case $J = 4$, $m = 2$ where with $\sum_{j=1}^4 p_j \leq mJ = 8$ and assume G is full rank.

$$\varepsilon_1 = \varepsilon_3 + \bar{\varepsilon}_1 \quad (103)$$

$$\varepsilon_2 = \varepsilon_4 + \bar{\varepsilon}_2 \quad (104)$$

Where $E[\bar{\varepsilon}_1|z] = E[\bar{\varepsilon}_2|z] = 0$, $E[\bar{\varepsilon}_1^2|z] = E[\bar{\varepsilon}_2^2|z] = 1$ and $E[\bar{\varepsilon}_1 \bar{\varepsilon}_2|z] = \rho_{\bar{\varepsilon}}$. In this case a common shock enters equations 1 and 2, and also 3 and 4.

$$g_1(\beta_0) - g_5(\beta_0) = (\bar{\varepsilon}_1 + (\kappa_{01} - \kappa_{03})x)z_1 \quad (105)$$

$$g_2(\beta_0) - g_6(\beta_0) = (\bar{\varepsilon}_2 + (\kappa_{01} - \kappa_{03})x)z_1 \quad (106)$$

$$g_3(\beta_0) - g_7(\beta_0) = (\bar{\varepsilon}_1 + (\kappa_{02} - \kappa_{04})x)z_2 \quad (107)$$

$$g_4(\beta_0) - g_8(\beta_0) = (\bar{\varepsilon}_2 + (\kappa_{02} - \kappa_{04})x)z_2 \quad (108)$$

So that

$$\text{cor}(g_1(\beta_0) - g_5(\beta_0), g_2(\beta_0) - g_6(\beta_0)) = \frac{\rho_{\bar{\varepsilon}} + (\kappa_{01} - \kappa_{03})^2 \sigma_x^2}{1 + (\kappa_{01} - \kappa_{03})^2 \sigma_x^2} \quad (109)$$

$$\text{cor}(g_3(\beta_0) - g_7(\beta_0), g_4(\beta_0) - g_8(\beta_0)) = \frac{\rho_{\bar{\varepsilon}} + (\kappa_{02} - \kappa_{04})^2 \sigma_x^2}{1 + (\kappa_{02} - \kappa_{04})^2 \sigma_x^2} \quad (110)$$

When $\rho_{\bar{\varepsilon}} = 1$ then Ω_n is exactly singular. As $(\kappa_{01} - \kappa_{03})^2$ and $(\kappa_{02} - \kappa_{04})^2$ increase (also as σ_x^2 increase), then $\text{cor}(g_1 - g_5, g_2 - g_6)$ and $\text{cor}(g_3(\beta_0) - g_7(\beta_0), g_4(\beta_0) - g_8(\beta_0))$ approach 1 respectively, even when $\rho_{\bar{\varepsilon}} < 1$. For example if $\kappa_{01} = 2$, $\kappa_{02} = -2$ and $\sigma_x^2 = 2$ and $\rho_{\bar{\varepsilon}} = 0.75$ then $\text{cor}(g_1 - g_5, g_2 - g_6) = 0.992$. Note in this setting the composite error terms ε_j for $j = 1, \dots, 4$ may not be perfectly correlated, and will allow quite general correlation patterns, even when Ω_n is exactly singular. Since $\varepsilon_1 = \varepsilon_3 + \bar{\varepsilon}_1 + \kappa_{01}x$, $\varepsilon_2 = \varepsilon_4 + \bar{\varepsilon}_2 + \kappa_{02}x$, $\varepsilon_3 = \varepsilon_3 + \kappa_{03}x$, $\varepsilon_4 = \varepsilon_4 + \kappa_{04}x$ even if $\rho_{\bar{\varepsilon}} = 1$ none of the composite error terms from the regression functions are necessarily perfectly correlated as $\text{cor}(\varepsilon_3, \varepsilon_4)$ can take any value and similarly κ_{0j} for $j = \{1, \dots, 4\}$.

If $\rho_\varepsilon = 1$ then Ω_n is singular, though if we include x in regression equation 1 and 2 then Ω_n will not be exactly singular when $\kappa_{03}, \kappa_{02} \neq 0$. Also when $\rho_\varepsilon < 1$ though say either $(\kappa_{01} - \kappa_{03})^2$ and σ_x^2 are high then $\text{cor}(g_1(\beta_0) - g_5(\beta_0), g_2(\beta_0) - g_6(\beta_0))$ is close to 1. Including x in any (or all) of the regression equations will reduce the correlation between linear combinations of moments. When we consider the case of multiple omitted variables, observable and unobservable with many equations J then singular may arise in many different ways, where changing any/all of the variables included can alter the rank of Ω_n .

Existence of $\hat{\Omega}(\tilde{\beta})^{-1}$ when Ω_n is Singular

When $\Lambda_{0n} = 0$, i.e $\Lambda_0 = 0_{\bar{m}}$ then in order for 2-Step GMM to exist we require $\Omega_n(\tilde{\beta})$ is full rank on $\beta \in \mathbb{B}(\beta_0, \epsilon)/\beta_0$ for some $\epsilon > 0$ (a ball around epsilon). This assumption is quite natural.

Take Example 1 with $J = 2$, $p_1 = p_2 = 1$, $\kappa_0 = 0$ $h_1(z) = h_2(z) = 1$ so that singularity in the moment function is the cause of singular moment variance. Then $\Omega_n(\beta_0 + \Delta)$ for some $\|\Delta\| > 0$ noting in the case of 2-Step GMM Δ is a random variable.

If $|\text{cor}(x_1, x_2)| < 1$ then $\epsilon_1(\beta_0 + \Delta) = \epsilon_1 + \delta_1 x_1, \epsilon_2(\beta_0 + \Delta) = \epsilon_2 + \delta_2 x_2$ then $\text{cor}(\epsilon_1(\beta_0 + \Delta), \epsilon_2(\beta_0 + \Delta)) < 1$ for all $\|\Delta\| > 0$ where $\Pr\{\tilde{\beta} = \beta_0\} = 0$ given $\tilde{\beta}$ is a continuous random variable. Likewise even if $x_1 = x_2 = x$ so that then $\epsilon_1(\beta_0 + \Delta) = \epsilon_1 + \delta_1 x, \epsilon_2(\beta_0 + \Delta) = \epsilon_2 + \delta_2 x$ then $\text{cor}(\epsilon_1(\beta_0 + \Delta), \epsilon_2(\beta_0 + \Delta)) = 1$ iff $\delta_1 = \delta_2$. The probability $\tilde{\beta}_1 - \beta_{01} = \tilde{\beta}_2 - \beta_{02}$ is equal to zero. In fact in linear models especially it is difficult to construct an example where this assumption will not hold. There would need to be a plane such that $\Omega_n(\beta_0 + \Delta)$ were singular for all $\Delta_1 \in (-\varsigma_1, \varsigma_2)$ and $\Delta_2 \in (-\zeta_1, \zeta_2)$ for some $\varsigma_1 > 0$ and/or $\varsigma_2 > 0$ and $\zeta_1 > 0$ and/or $\zeta_2 > 0$. In this case the probability that $\tilde{\beta}$ lies in this space may be greater than zero. Even in non-linear models, for example $y = \beta x^\gamma + \epsilon$ when $\beta_0 = 0$ then the moment variance matrix is singular for β_0 and $\gamma = \gamma_0 + \delta$ for any δ . Again the probability that the GMM estimator lies in this space is zero. We would require also that moment variance were singular for some β in some small neighborhood around β_0 . It is difficult to think of such examples. As such the assumption at first seemingly restrictive conditions is in fact in a relatively innocuous assumption

Bibliography

- Andrews, D.W.K., 1987. Asymptotics results for generalized Wald tests. *Econometric Theory* 3, 347-358.
- Andrews, D.W.K., 2001. Testing When a Parameter Is on the Boundary of the Maintained Hypothesis. *Econometrica* 69, 683-734.
- Andrews, D. W. K., and Cheng, X. 2012: Estimation and Inference With Weak, Semi-Strong, and Strong Identification, *Econometrica*, 80, 2151-2211.
- Andrews, Donald W.K. and Cheng, Xu, 2013. Maximum likelihood estimation and uniform inference with sporadic identification failure, *Journal of Econometrics*, Elsevier, vol. 173(1), pages 36-56.
- Andrews, D. W. K., and Guggenberger, P., 2015. Identification- and Singularity-Robust Inference for Moment Condition Models, Cowles Foundation Discussion Paper No. 1978.
- Arellano, M., and Bond, S., 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, Wiley Blackwell, vol. 58(2), pages 277-97.
- Bosq, D., 2000. *Linear Processes in Function Spaces*. Springer, New York.
- Bottai, M., Rotnitzky, A., D. R. Cox, , and J. Robins, 2000. Likelihood-based inference with singular information matrix, *Bernoulli* 6 , 243-284.
- Choi, I., and Phillips, P C. B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics*, 51, pages 113-150.
- Doran, H. E., and Schmidt, P., 2006. GMM estimators with improved finite sample properties using principal components of the weighting matrix, with an application to the dynamic panel data model. *Journal of Econometrics* 133, 387-409.
- Dufour, J.M., and Valéry, P., 2011. Wald-type tests when rank conditions fail: a smooth regularization approach. Unpublished Working Paper.
- Grant, N.L., 2013. Identification Robust Inference with Singular Variance, Manchester University Economics Discussion Paper Series, EDP-1315.
- Han, C., and Phillips, Peter C. B., 2010. GMM Estimation For Dynamic Panels With Fixed Effects And Strong Instruments At Unity, *Econometric Theory*, Cambridge University Press, vol. 26(01), pages 119-151.
- Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-54.
- Kato, T. 1982: 'A short introduction to the perturbation theory of linear operators,' Springer-Verlag 1982.

- Lee, L., and Chesher, A., 1986. Specification testing when score test statistics are identically zero, *Journal of Econometrics*, Elsevier, vol. 31(2), pages 121-149.
- Melino, A. 1982. Testing for sample selection bias. *Review of Economic Studies*. 49: 151-153.
- Newey, W.N., and Windmeijer, F., 2009. Generalized Method of Moments With Many Weak Moment Conditions. *Econometrica* 77, 687-719.
- Peñaranda, F. and Sentana, E., 2010. A unifying approach to the empirical evaluation of asset pricing models. *Economics Working Papers 1229*, Department of Economics and Business, Universitat Pompeu Fabra.
- Ruge-Murcia, Francisco J., 2007. Methods to estimate dynamic stochastic general equilibrium models, *Journal of Economic Dynamics and Control*, Elsevier, vol. 31(8), pages 2599-2636.
- Staiger, D., and Stock, J.H., 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica* 65, 557-586.
- Stock, J.H. and Wright, J., 2000. GMM with Weak Identification. *Econometrica* 68, 1055-1096.