

Economics
Discussion Paper Series
EDP-1417

GMM with Weakly Singular Variance

Nicky Grant

October 2014

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL


GMM with Weakly-Singular Variance

Nicky Grant

University of Manchester

August 8, 2014

Abstract



This paper considers 2-Step GMM when moment functions may be highly correlated at β_0 . When moments have Weakly-Singular variance (WSV), defined as a subset of eigenvalues of the moment variance local-to-zero, GMM is consistent, converging to a Gaussian limit distribution at rate faster than $n^{1/2}$ in certain directions. WSV is shown to be related to Nearly-Singular Design, an assumption employed in Caner (2008), where results in this paper are demonstrated to be incorrect. Simulation evidence verifies the results of this paper, highlighting also the potential poor small sample Gaussian approximation of the distribution of GMM when moments are highly correlated.

KEYWORDS: GMM, Singular Variance, Efficiency.

JEL CLASSIFICATION: C10, C13.

*Would like to thank Alastair Hall and Peter C.B. Phillips for useful comments and discussion when writing this paper. Any remaining errors are my own.

[†]Nicky Grant, Arthur Lewis Building, Faculty of Economics, Bridgeford Street, M13 9PL, Manchester, UK. 07930861907. nicky.grant@manchester.ac.uk.

1 Introduction

This paper derives the asymptotic properties of the 2-Step GMM estimator when moments may have a variance matrix close to singular at the true parameter, β_0 . This setting was initially considered in linear models by Knight & Fu (2000) and Knight (2008) which model this almost singularity by the ‘Nearly-Singular Design’ (NSD) assumption. Caner (2008) and Caner & Yildiz (2012) employ NSD for general moment type estimators. This paper demonstrates the results in Caner (2008) are incorrect, in part due to an inadequacy shown with the NSD assumption for deriving asymptotic properties of the GMM estimator¹.

The issues with the NSD assumption are detailed, and a related assumption termed ‘Weakly-Singular Variance’ (WSV) is provided that overcome these problems. WSV models a subset of eigenvalues of the moment variance matrix as local-to-zero, so that asymptotically the moment variance matrix may be singular. This modeling framework is directly analogous to the method with which weak identification is modeled, which allows the expected first order derivative matrix at β_0 to shrink to zero at some rate (e.g Staiger & Stock (1997), Newey & Windmeijer (2009)). This paper considers the strongly identified moment setting, where results could be extended with relevant modification and regularity conditions to the many weak moment asymptotics of Newey & Windmeijer (2009)².

In contrast to the results in Caner (2008), showing that convergence to a Gaussian limit distribution is slower than $n^{1/2}$ under NSD, when moments have WSV (which in Section 2 under some mild conditions is shown to be equivalent to NSD when moments have asymptotically singular variance) con-

¹This paper does not consider the results in Knight & Fu (2000) and Knight (2008), where results in both papers pertaining to NSD are also incorrect, though in ways that differ to Caner (2008). Showing this is beyond the scope of this paper.

²Caner & Yildiz (2012) consider the many weak moment setup of Newey & Windmeijer (2009) with NSD. Results in this paper suggest these results may also be incorrect or incomplete, given the paucity of NSD to derive general asymptotics. This issue is beyond the scope of this paper.

vergence occurs at a rate faster than $n^{1/2}$ in certain directions. When the null space of the variance matrix lies within that of the outer product of the expectation of the moment derivative at β_0 convergence is shown at the standard $n^{1/2}$ rate. Interestingly this is the condition required for the Generalized Anderson Rubin (GAR) statistic to not diverge near the true parameter when moments have singular variance, Grant (2013). When this condition does not hold it is shown in certain directions, when eigenvalues converge slowly enough to zero, the 2-Step GMM estimator converges at rate faster than $n^{1/2}$ to a Gaussian distribution of the standard form when moments are non-singular. A heuristic explanation of this result is provided along with formal results and proofs along with the relevant regularity conditions.

Though 2-Step GMM potentially converges at a non-standard rate, the Hansen (1982) over-identification test statistic is shown to converge at the standard rate to a chi-squared distribution. Correspondingly the Wald Statistic for testing linear restrictions is also shown to have first order asymptotics not impacted by asymptotic singularity of the moment variance matrix (when eigenvalues converge slowly enough to zero).

In order to derive the asymptotic properties of 2-Step GMM the eigen-system expansions of the sample moment variance around β_0 of Grant (2013) when moments are exactly singular are extended to the case when moments have WSV. These results may be of interest in their own right in establishing asymptotic properties of other statistics related to estimation from moment functions with WSV.

The issue of singular variance for inference in general has been considered elsewhere in the literature. Peñaranda & Sentana (2010) provide asymptotic results for GMM when those linearly redundant combinations at the true parameter are known a priori. A transformed moment function that removes these linearly redundant moment conditions such that this transformed moment has non-singular variance is proposed. Standard results on GMM then

go through under the usual identification and regularity conditions on the transformed moment function.

Grant (2013) studies the link between identification and singular variance-showing for a wide class of moment functions lack of identification and singular moment variance are equivalent. Results are provided on the GAR statistic dropping the full rank variance assumption and general conditions provided when the GAR statistic has a chi-squared limit distribution at points near to the true parameter when moments may have singular variance.

Another strand of the literature focuses on the issues of Wald Type Tests where under the null the of restricted moment variance is singular. Andrews (1987) provide general conditions when the use of a Generalized Inverse will yield valid inference when utilizing standard asymptotics. Dufour & Valéry (2011) consider this issue further, developing a regularized Wald Statistic and the corresponding limit distribution to allow for inference methods in more general settings when the conditions in Andrews (1987) for regular asymptotics to work may break down. Other papers have also considered this singularity issue in Wald Tests, however none provide properties of the 2-Step GMM estimator with (asymptotically) singular variance.

Outside of the NSD literature- the issue of singular variance for the properties of moment estimators seems to have garnered little attention. Those that have considered this case in any way tend to remove the (asymptotically) singular variance issue via some transformation. Common methods assume those linear combinations of redundant moment at β_0 are known, or a regularized inverse is used to eradicate the singularity issue. As such the general properties of 2-Step GMM with asymptotically singular variance are unknown.

Singular or almost singular variance matrices are also commonly met in applied research, especially in dynamic simultaneous equation or panel type models where moments with almost singular variance in the finite sample are not uncommon, e.g Doran & Schmidt (2006). The methods practitioners use to

remove this (almost) singularity are often be ad hoc with little formal theoretical justification for procedures employed to remove singularities. Singularities also commonly arise in large factor models, for example Aguilar (2009) who proposes a regularization approach to overcome the issue, with other methods include adding and deleting variables until moment singularity is eradicated or a change the instrument set. The implications of these ad hoc methods can alter the rate of convergence of GMM when moments have WSV as shown in this paper.

Simulation evidence shows that the distribution of the the GMM estimator is poorly approximated by a Gaussian distribution the closer the moment variance is to singularity, even for potentially very large sample sizes. Theoretical justification for this result is provided in Grant (2014), who considers 2-Step GMM where moments are exactly singular (or eigenvalues converge fast enough to zero) such that the limit distribution when appropriately scaled is highly non-standard. Though inference based on the Gaussian approximation to the distribution of GMM is likely to be more accurate when removing singularities, the estimator may not be as efficient, given it may reduce the rate of convergence. Results in this paper provide some initial insights on the theoretical implications of singular variance for inference from the efficient GMM estimator.

Section 2 details the issue with NSD and defines WSV, showing the link between the two assumptions. Section 3 derives asymptotic properties of 2-Step GMM with WSV and discusses the implications in respect to the conventional notion of efficient estimation and identification. The Wald Test and Hansen's J-statistic are shown to satisfy standard asymptotics when moments have WSV when the eigenvalues converge slowly enough to zero. Section 4 lays out the eigen-system expansions of the sample variance matrix around β_0 under WSV. Section 5 gives examples of singular variance with Section 6 providing a simulation experiment that verifies all key results of this paper.

Finally Section 7 concludes, with definitions and proofs of the main theorems contained in an Appendix.

1.1 Notation of Paper

For simplicity this paper considers $\{w_i: (i = 1, \dots, n), n \geq 1\}$ is and i.i.d sequence where $w \in \mathbb{R}^k$. $\beta \subseteq \mathbf{B} \subset \mathbb{R}^p$. Let the known moment function $m \times 1$ moment function $g(\cdot, \cdot) : \mathbb{R}^k \times B \mapsto \mathbb{R}$ which satisfies,

$$E[g(w_i, \beta_0)] = 0 \tag{1}$$

where $\beta_0 \in \mathbf{B}$ is the unknown true parameter where we maintain the standard $\beta_0 \in \text{int}(\mathbf{B})$ to avoid the parameter on the boundary problem, Andrews (2001). Define $g_i(\beta) = g(w_i, \beta)$, $\hat{\Omega}(\beta) = \frac{1}{n} \sum_{i=1}^n g_i(\beta)g_i(\beta)'$ $\Omega_n(\beta) = E[g_i(\beta)g_i(\beta)']$, $\Omega(\beta) = \lim_{n \rightarrow \infty} \Omega_n(\beta)$. $\Omega_n(\beta)$ is written as potentially a function of n which is used when modeling the eigenvalues of $\Omega_n(\beta_0)$ as local-to-zero in the definition of WSV in Section 2. Define $\hat{\Omega} = \hat{\Omega}(\beta_0)$, $\Omega_n := \Omega_n(\beta_0)$, $\Omega := \Omega(\beta_0)$ ³. Define the sample moment function $\hat{g}(\beta) = \frac{1}{n} \sum_{i=1}^n g_i(\beta)$ and $g(\beta) := E[g_i(\beta)]$ and the first order derivative of the moment function $G_i(\beta) = \partial/\partial\beta g_i(\beta)$, $G(\beta) := E[G_i(\beta)]$, $\hat{G}(\beta) = \frac{1}{n} \sum_{i=1}^n G_i(\beta)$ with $G := G(\beta_0)$. Further statistical and matrix definitions used throughout this paper are detailed in Appendix A.

2 Nearly- Singular Design & Weakly Singular Variance

The Nearly-Singular Design (NSD) assumption was originally developed by Knight & Fu (2000) to model the properties of estimators in linear models with highly correlated regressors. The assumption was also used to derive further results in Knight (2008). Caner (2008) employs the NSD assumption to study the properties of GMM-type estimators of moment functions with almost

³See Appendix A1 for definitions of eigenvalues/vectors of the sample and population variance matrix throughout this paper.

singular variance, with results extended in Caner & Yildiz (2012) to many weak moment setting of Newey & Windmeijer (2009). This paper discusses the issues with the results in Caner (2008). Firstly we define NSD as employed in Caner (2008) and then formally define WSV. NSD and WSV are shown to overlap in the case when moments are asymptotically singular, as such the results in this paper on the asymptotic properties of 2-Step GMM should coincide with corresponding results in Caner (2008). This is show not to be the case as Theorem 5 of Caner (2008) shows GMM converges at a rate slower than $n^{1/2}$ to a distribution of a Gaussian form with a non-standard limiting variance $(G'D^{-1}G)^{-1}$. Results in this paper show when moments are asymptotically singular then under WSV convergence occurs at rate (in certain directions) faster than $n^{1/2}$ to a Gaussian limit distribution with asymptotic variance of a form analogous to the standard case where moments have full rank variance.

The following i the NSD is made in Assumption 1 of Caner (2008, pg. 513).

NEARLY-SINGULAR DESIGN

For $a_n = n^\kappa$ where $0 < \kappa < 1$,

$$a_n(\hat{\Omega}(\beta) - \Omega(\beta)) \xrightarrow{p} D(\beta) \tag{2}$$

where for all $\beta \in \mathbf{B}$ $D(\beta)$ is a bounded $m \times m$ matrix where $u'D(\beta)u > 0$ for all $u \in \mathcal{N}(\Omega(\beta))$ and $\text{Rank}(\Omega(\beta)) < m$.

The moment variance $E[g_i g_i']$ is not modelled as a function of n in Caner (2008) and others utilising the NSD assumption. However this has to be the case, otherwise NSD is a theoretical impossibility. To show this take the i.i.d case and focus on NSD at $\beta = \beta_0$, define $\Omega := E[g_i g_i']$ for any n, then $P'_{0j} \Omega P_{0j} = E[(P'_{0j} g_i)^2] = 0$ for $j = 1, \dots, \bar{m}$ hence $P'_{0j} g_i = 0$ $j = 1, \dots, \bar{m}$ w.p.1. Therefore $\text{Rank}(\hat{\Omega}) \leq m - \bar{m}$ w.p.1. However NSD (at $\beta = \beta_0$) implies

$a_n P_0'(\hat{\Omega} - \Omega)P_0 \xrightarrow{p} P_0' D P_0$ where $P_0' D P_0$ is full rank. Since $P_0'(\hat{\Omega} - \Omega)P_0 = 0$ w.p.1 then the NSD assumption is an impossibility, unless Ω is non-singular for finite n and hence is modelled as a function of n . Hence why we define $\Omega_n := E[g_i g_i']$ to allow the moment variance to be a function of n where $\Omega = \lim_{n \rightarrow \infty} \Omega_n$. Standard asymptotics hold when $\Omega_n = \Omega$ and Ω is full rank. The NSD assumptions could only hold if Ω_n converges to a singular matrix in certain directions at some rate. This is exactly the WSV assumption- by modeling (a subset of) eigenvalues as shrinking to zero at some rate, defined below.

WEAKLY SINGULAR VARIANCE

$$\Omega_n(\beta) = P_+(\beta)\Lambda_+(\beta)P_+(\beta)' + P_0(\beta)\Lambda_{0n}(\beta)P_0(\beta)' \quad (3)$$

For every $\beta \in \mathbf{B}$ where $\Lambda_+(\beta)$ is an $(m - \bar{m}(\beta)) \times (m - \bar{m}(\beta))$ matrix and $\Lambda_0(\beta)$ an $\bar{m}(\beta) \times \bar{m}(\beta)$ matrix such that $\Lambda(\beta) = \text{diag}(\Lambda_+(\beta), \Lambda_0(\beta))$ is full rank and $\Lambda_{0n}(\beta) = \Lambda_0(\beta)/b_n(\beta)$ for $0 \leq \bar{m}(\beta) \leq m$ for $b_n(\beta) = n^{\delta(\beta)}$ where $0 < \delta(\beta) \leq 1$.

Note that WSV allows the rank of $\Omega(\beta)$ to vary with β and includes the case where $\Omega(\beta)$ be full rank ($\bar{m}(\beta) = 0$) and of singularity ($\bar{m}(\beta) > 0$). WSV does assume that $\Omega_n(\beta)$ is non-singular for finite n across β . This assumption could be weakened such that WSV holds only for a neighborhood around β_0 , so that $\hat{\Omega}(\beta)$ at a point near to β_0 will exist w.p.a.¹⁴.

WSV assumes that a subset of eigenvalues shrink towards zero at some rate. This assumption covers the case when moments have non-singular variance with $\bar{m} = 0$, and hence $\Lambda = \Lambda_+$ and $P = P_+$, unlike NSD which assumes Ω is singular. Define $\delta(\beta_0) := \delta$. This paper considers $\delta < 1$ and Λ_0 full rank.

This assumption, under the strong identification and further regularity con-

¹⁴Note that this assumption is implicitly made in any research on 2-Step GMM otherwise the objective function would not exist for finite n . Even in the standard case when $\Omega = \Omega_n$ and is full rank, then unless $\Omega_n(\beta)$ is non-singular in some small region around β_0 the 2-Step GMM estimator may not exist. This issue is discussed further in Grant (2014).

ditions, guarantee the asymptotic distribution (when appropriately scaled) of the GMM estimator is Gaussian. A companion paper Grant (2014) considers the case where $\delta = 1$ where Λ_0 is full rank and $\Lambda_0 = 0$ (i.e exact singularity). In both cases the limit distribution converges at rate n in certain directions to a non-standard limit distribution. The asymptotic properties of GMM depend on the size of the eigenvalues and this issue is discussed in Section 4.

Focussing on both NSD and WSV at $\beta = \beta_0$ (the point critical for the asymptotic properties of GMM) then unless we restrict Ω to be singular then WSV is more general than NSD, since NSD does not allow for the case of Ω non-singular. It is straightforward to show NSD (when $\kappa < 1/2$, as in principle it could only ever hold for these values since $\hat{\Omega} = \Omega_n + O_p(n^{-1/2})$) and WSV for $\delta < 1/2$ are equivalent when Ω is singular and $\hat{\Omega} - \Omega_n = O_p(n^{-1/2})$ which holds under mild regularity conditions under an i.i.d assumption⁵. As such the results found in this paper should be equivalent to those in Caner (2008) for these cases, though the results of this paper are more general in not placing the undesirable assumption that Ω be singular⁶.

LEMMA 1: *NSD (for $\kappa < 1/2$) and WSV (for $\delta < 1/2$) are equivalent at β_0 where Ω is singular and $\hat{\Omega} - \Omega_n = O_p(n^{-1/2})$.*

Note that $\Omega_n = E[g_i g_i']$ is the moment variance and $\hat{\Omega} - \Omega_n = O_p(n^{-1/2})$ holds under the maintained i.i.d assumption along with $E[g_i^4] < \infty$. Given the equivalence of NSD and WSV for certain restrictions on κ and δ when Ω singular then the results in Caner (2008) should coincide with the results in this paper derived under WSV under singularity. This turns out not to be the case due to a mistake in Caner (2008).

The NSD assumption is not sufficient to derive asymptotic results on the

⁵Equivalence is shown only at $\beta = \beta_0$ for brevity, similar results can be extended for $\beta \in \mathbf{B}$ under relevant regularity conditions.

⁶Note that WSV is more general than NSD, though for $\delta < 1/2$ and Ω singular the two are equivalent

GMM estimator since it relies on an assumption moments have singular variance. Also the form of the assumption is not practical in terms of deriving the general asymptotic properties of GMM. NSD does not allow the study of the interaction of the eigenvalues/eigenvectors of the estimated optimal weight matrix with the moment function $\hat{g}(\beta)$ which as discussed in Section 3 turns out to be critical in determining asymptotic properties of the 2-Step GMM estimator. The WSV assumption overcomes both issues, expressing the almost singularity as the eigenvalues of the moment variance matrix potentially shrinking to zero and incorporating also the standard full-rank assumption as a special case.

3 GMM With Weakly-Singular Variance

In this section asymptotic results are provided for 2-Step GMM when moments have Weakly- Singular Variance. To highlight the issues arising solely from Weakly-Singular Variance this paper considers the strongly identified case, with extensions to weak identification setup of Newey & Windmeijer (2009) following with modification and further regularity conditions.

When moments have WSV it is shown that the 2-Step GMM estimator converges at rate faster than $n^{1/2}$ (in some directions) when there exist $\omega \in \mathbb{R}^m$ such that $\omega' \Omega = 0$ does not imply $\omega' G = 0$. If $\mathcal{N}(\Omega) \subseteq \mathcal{N}(GG')$ then convergence is at rate $n^{1/2}$. Note that this is exactly the condition required for the GAR statistic not to diverge, Grant (2013).

For simplicity this paper considers WSV for the case where $\delta < 1$ so that the limit distribution is of a standard Gaussian form. If $\delta = 1$, or moments were exactly singular, then the limit distributions of 2-Step GMM become highly non-standard as the estimation error in the initial estimator of β_0 does not drop out of first order asymptotics. This case though interesting introduces substantial theoretical complications worthy of a full research piece in itself and

as such is considered in Grant (2014). The discontinuity in asymptotic properties of GMM around $\delta = 1$ is analogous to the linear Instrumental Variable (IV) weak-identification setup, modeling the first stage parameter $\Pi_n = C/n^\mu$ where C is $p \times m$ matrix. If C is full column rank and $0 \leq \mu < 1/2$ the IV estimator converges at rate $n^{1/2-\mu}$ to a Gaussian limit distribution, e.g Newey & Windmeijer (2009). If $\mu = 1/2$ IV is inconsistent with a non-standard limit distribution e.g Staiger & Stock (1997). For the exactly unidentified case, i.e $C = 0$, IV is again inconsistent with a non-standard limit distribution but without the drift term C appearing in asymptotic results, e.g Choi & Phillips (1992).

Section 3.1 derives asymptotic properties of 2-Step GMM with WSV. Critical to the proof of both Theorems 1- 3 are the asymptotic eigen-system expansions of $\hat{\Omega}(\beta)$ around β_0 provided in Theorem 4 under Assumption 2⁷.

3.1 2-Step GMM with WSV

We consider the implications of WSV on first order asymptotics for 2-Step GMM under the strong identification condition and regularity conditions laid out in Assumption 1.

ASSUMPTION 1 (A1): IDENTIFICATION CONDITIONS.

(i) G is a full rank $p \times m$ matrix, (ii) $n^{1/2}\|\hat{g}(\beta)\| \leq Cn^{1/2}\|\beta - \beta_0\| + \hat{M}$ for some $0 < C < \infty$ and $\hat{M} = O_p(1)$ for all $\beta \in \mathbf{B}$.

A1(i) assumes first order identification where (ii) is similar to the identification assumption of Newey & Windmeijer (2009). Assumption 1 (ii) follows from the global identification assumption $\|g(\beta)\| \leq C\|\beta - \beta_0\|$ where $\|g(\beta)\| = 0$ if

⁷The proof of Theorems 1-3 can be read taking as given the expansions in Theorem 4. The reader interested in the derivation of such expansions may wish to read Section 4 prior to Section 3.1.

and only if $\beta = \beta_0$ and also convergence of $n^{1/2}(\hat{g}(\beta) - g(\beta)) = O_p(1)$ with sufficient conditions being w_i i.i.d (along with $E[\|g_i(\beta)\|^4] < \infty$ across \mathbf{B}). A1 (ii) could be weakened to hold for β in a neighborhood around β_0 . This would not change the results of this paper and A1 (ii) is made for simplicity. A1 will guarantee $n^{1/2}$ convergence of the initial GMM estimator (under some further assumptions in A2) to a standard limit distribution.

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.d matrix W_n , where $W_n \xrightarrow{p} W$,

$$\tilde{\beta} = \arg \min_{\beta \in \mathbf{B}} \hat{Q}_W(\beta) \quad (4)$$

and define $\hat{Q}_W(\beta) := n\hat{g}(\beta)'W_n\hat{g}(\beta)$. Then if W_n is p.d under A1, A2, then by Hansen (1982),

$$\tilde{\beta} \xrightarrow{p} \beta_0 \quad (5)$$

$$n^{1/2}(\tilde{\beta} - \beta_0) \xrightarrow{d} N(0, \Psi(W)) \quad (6)$$

where $\Psi(W) = (G'WG)^{-1}G'\Omega G(G'WG)^{-1}$ when W is such that $G'WG$ is full rank. Note that this result holds whether or not Ω is full-rank.

Define the 2-Step GMM objective function,

$$\hat{Q}_{opt}(\beta) = n\hat{g}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta) \quad (7)$$

and the corresponding 2-Step GMM estimator,

$$\hat{\beta} = \arg \min_{\beta \in \mathbf{B}} \hat{Q}_{opt}(\beta) \quad (8)$$

non-standard asymptotics may arise when using $W_n = \hat{\Omega}(\tilde{\beta})^{-1}$ when Ω_n is

weakly-singular and Ω is singular. In this case, as noted in Caner (2008) $\hat{\Omega}(\tilde{\beta})^{-1}$ diverges, though this is the case, when $\hat{\Omega}(\tilde{\beta})^{-1}$ is multiplied by $\hat{g}(\beta)$ to form $\hat{Q}_{opt}(\beta)$ will not diverge for β near to β_0 as we demonstrate heuristically below. For simplicity we highlight the intuition for linearity case for simplicity, i.e $\hat{G}(\beta) = \hat{G}(\beta_0) := \hat{G}$ for all $\beta \in \mathbf{B}$. Analogous results under regularity conditions made in this paper can be established similarly for the non-linear case.

Using the eigenvalue decomposition re-write $\hat{\Omega}(\tilde{\beta})^{-1}$,

$$\hat{\Omega}(\tilde{\beta})^{-1} = \hat{P}_+(\tilde{\beta})\hat{\Lambda}_+(\tilde{\beta})^{-1}\hat{P}_+(\tilde{\beta})' + \hat{P}_0(\tilde{\beta})\hat{\Lambda}_0(\tilde{\beta})^{-1}\hat{P}_0(\tilde{\beta})' \quad (9)$$

defining $\hat{Q}_{opt}^+(\beta) = n\hat{g}(\beta)'\hat{P}_+(\tilde{\beta})\hat{\Lambda}_+(\tilde{\beta})^{-1}\hat{P}_+(\tilde{\beta})'\hat{g}(\beta)$,

$\hat{Q}_{opt}^0(\beta) = n\hat{g}(\beta)'\hat{P}_0(\tilde{\beta})\hat{\Lambda}_0(\tilde{\beta})^{-1}\hat{P}_0(\tilde{\beta})'\hat{g}(\beta)$. such that the objective function may be expressed,

$$\hat{Q}_{opt}(\beta) = \hat{Q}_{opt}^+(\beta) + \hat{Q}_{opt}^0(\beta) \quad (10)$$

where if $\hat{Q}_{opt}^+(\beta)$ satisfies standard asymptotics (i.e $\bar{m} = 0$) then $\hat{Q}_{opt}(\beta) = \hat{Q}_{opt}^+(\beta)$ and the 2-Step GMM estimator will converge at rate $n^{1/2}$ to a Gaussian limit distribution. When $\bar{m} > 0$ then $\hat{Q}_{opt}^0(\beta)$ will not diverge for β close enough to β_0 , even though $\hat{\Lambda}_0(\tilde{\beta})^{-1}$ diverges at rate n^δ since $\hat{P}_0(\tilde{\beta})'n^{1/2}\hat{g}(\beta_0) = O_p(n^{-\delta/2})$ as shown in proof of Theorem 1 below in Appendix C.

We now heuristically demonstrate why $\hat{Q}_{opt}(\beta)$ remains bounded for β close enough to β_0 , even when moments have WSV.

Taylor expand $\hat{g}(\beta)$ (and using the linearity assumption),

$$\hat{Q}_{opt}^0(\beta) = \sum_{j=1}^{\bar{m}} \left(n^{\delta/2}\hat{P}_{0j}(\tilde{\beta})'n^{1/2}\hat{g}(\beta_0) + n^{\delta/2}\hat{P}_{0j}(\tilde{\beta})'\hat{G}n^{1/2}(\beta - \beta_0) \right)^2 / n^\delta \hat{\lambda}_{0j}(\tilde{\beta})$$

if $P'_{0j}G = 0$ for all $j = \{1, \dots, \bar{m}\}$ then since $n^{\delta/2}\hat{P}_{0j}(\tilde{\beta}) = n^{\delta/2}P_0 + o_p(1)$ and $n^{\delta/2}\hat{P}_{0j}(\tilde{\beta}) = n^{\delta/2}P_0 + o_p(1)$ by Theorem 4 (since $\delta < 1$) then since \hat{G} is

bounded by A1(i), A2(iv) it is straightforward to show, $\hat{P}_{0j}(\tilde{\beta})'\hat{G} = P'_{0j}(\hat{G} - G)$ where $(\hat{G} - G) = O_p(n^{-1/2})$ under A1,A2, such that it can be shown that for all $j = \{1, \dots, \bar{m}\}$,

$$n^{\delta/2}\hat{P}_{0j}(\tilde{\beta})'\hat{G}n^{1/2}(\beta - \beta_0) = O_p(n^{(\delta-1)/2})O_p(n^{1/2}\|\beta - \beta_0\|) \quad (11)$$

hence for $\beta = \beta_0 + O_p(n^{-1/2})$ then $\hat{Q}_{opt}^0(\beta)$ will be bounded when $\delta < 1$. Since the 2-Step GMM estimator minimizes this function it too will be $O_p(n^{-1/2})$. If any $P'_{0j}G \neq 0$ then $n^{\delta/2}\hat{P}_{0j}(\tilde{\beta})'\hat{G}n^{1/2}(\beta - \beta_0)$ will have a similar expansion to (11) with a term $n^{\delta/2}P'_{0j}Gn^{1/2}(\beta - \beta_0)$ appearing on the left hand side For this statistic not to diverge then in some directions $n^{1/2}(\beta - \beta_0)$ will have to converge to at rate $n^{\delta/2}$ so that $\hat{Q}_{opt}^0(\beta)$ remains bounded. Since the 2-Step GMM estimator minimizes this function it will converge at rate faster than $n^{(1+\delta)/2}$ in certain directions in this case.

Theorem 1 shows that the limit distribution of $n^{1/2}B_n(\hat{\beta} - \beta_0)$ for a certain $p \times p$ matrix B_n has a Gaussian limit distribution with asymptotic variance of a similar form to the standard case where Ω is full rank (i.e $(G'\Omega^{-1}G)^{-1}$). When there exist directions such that $\omega'\Omega = 0$ does not imply $\omega'G \neq 0$ then B_n will have some elements diverging to infinity at a certain rate.

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

$$G_{BP_0} =: B'G'P_0 = \begin{pmatrix} G_{BP_0}^{\bar{p}} \\ 0_{(p-\bar{p}) \times \bar{m}} \end{pmatrix} \quad (12)$$

where $G_{BP_0}^{\bar{p}}$ is full 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 (12). 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 (12). If $0 < \bar{p} < p$ then we can perform Gauss-Jordan elimination to re-write $G'P_0$ in the form (12).

Define $B_n = I_n B^{-1}$, where $I_n := \begin{pmatrix} n^{\delta/2} I_{\bar{p}} & 0_{\bar{p} \times \bar{p}} \\ 0_{(p-\bar{p}) \times (p-\bar{p})} & I_{p-\bar{p}} \end{pmatrix}$,

where I_n is such that $I_n^{-1} G_{BP_0} = O(n^{-\delta/2})$ which is crucial in the proof of

Theorem 1. Define $\Lambda = \text{diag}(\Lambda_+, \Lambda_0)$, $G_{BP} := \begin{pmatrix} G_{BP_+}^{\bar{p}} & 0_{\bar{p} \times \bar{m}} \\ 0_{(p-\bar{p}) \times (m-\bar{m})} & G_{BP_0}^{p-\bar{p}} \end{pmatrix}'$,

where $G_{BP_+}^{p-\bar{p}}$ is the lower $(p-\bar{p}) \times (m-\bar{m})$ sub-matrix of $B'G'P_+ := G_{BP_+}$.

THEOREM 1 Under A1 and A2,

$$n^{1/2}(\hat{\beta} - \beta_0) = O_p(1) \quad (13)$$

$$n^{1/2} B_n(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, (G'_{BP} \Lambda^{-1} G_{BP})^{-1}) \quad (14)$$

where G_{BP} will always be full rank when G is full rank, given B and P are both full rank matrices.

REMARKS

(i) When $\bar{m} = 0$ (and hence $\bar{p} = 0$) and Ω is full rank (i.e $\Omega = P_+ \Lambda_+ P_+'$) where $G_{BP} = G_{BP_+} = B'G'P_+$ s.t $G'_{BP} \Lambda G_{BP} = B'G'P_+ \Lambda_+^{-1} P_+' G B$ and hence Theorem 1 contains the standard case (for strongly identified moments) when Ω is full rank.

(ii) Note also when $\bar{p} = 0$ i.e $G'P_0 = 0$ then $B_n = B^{-1}$ and hence the GMM estimator converges at the standard root-n rate, even when moments have asymptotically singular variance, i.e $\bar{m} > 0$. Note that $G'P_0 = 0$ is exactly the condition that $\mathcal{N}(\Omega) \subseteq \mathcal{N}(GG')$.

At the other extreme if $\bar{p} = p$ (i.e $G'P_0$ is full rank) then $B_n = n^{\delta/2} B^{-1}$ and $(\hat{\beta} - \beta_0)$ is $O_p(n^{-(1+\delta)/2})$. In the case $0 < \bar{p} < p$ then in some directions $(\hat{\beta} - \beta_0)$ is $O_p(n^{-(1+\delta)/2})$ and in others $O_p(n^{-1/2})$. In both cases the condition $\mathcal{N}(\Omega) \subseteq \mathcal{N}(GG')$ does not hold.

(iii) Simulation evidence in Section 6 demonstrates the poor Gaussian approx-

imation to the distribution of $n^{1/2}(\hat{\beta} - \beta_0)$ in small samples when Ω_n is close to singular. This follows from the results in Grant (2014) where the limit distribution of 2-Step GMM becomes non-standard when $\delta = 1$ (or Ω_n is exactly singular).

(iv) By Lemma 1 NSD and WSV are equivalent when Ω is singular and $\kappa, \delta < 1/2$, hence in this case the results in Caner (2008) for GMM should be the same as those found in Theorem 1 under WSV. However Theorem 5 of Caner (2008, pg 517) states that,

$$\frac{n^{1/2}}{a_n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, (G'D^{-1}G)^{-1}) \quad (15)$$

holds ‘on the null space of Ω ’. This statement is made in all proofs and theorems in the paper, however is unclear what this means as no restriction is made to any null space. Two things to note (a) Theorem 1 shows the rate of convergence is rate $n^{1/2}$ or faster when Ω singular and $\mathcal{N}(\Omega) \not\subseteq \mathcal{N}(GG')$ whereas Theorem 5 of Caner (2008) shows the rate of convergence always decreases, (b) the limiting variance $(G'D^{-1}G)^{-1}$ does not equal the limiting variance in Theorem 1. Note from Lemma 1 WSV and NSD are equivalent for $D = P_0\Lambda_0P_0'$ and $a_n = b_n$ though in this case D is not full rank (barring the pathological degenerate case where $\bar{m} = m$)⁸.

Theorem 2 derives the limit distribution of $(\hat{\beta} - \beta_0)$ normalized by an estimator of the square root of the asymptotic variance. This statistic is commonly used in practise when making inference on β_0 . Theorem 2 shows convergence at the standard root-n rate to a $N(0, I_p)$ distribution. Define $\hat{V}(\beta) = \hat{G}(\beta)' \hat{\Omega}(\beta)^{-1} \hat{G}(\beta)$.

⁸Similar problems also hold for results on Generalized Empirical Likelihood and all other theorems in this paper which are built on the limiting distribution of moment type estimators. The mistakes rest largely in the proofs, along with the fact at any rate NSD is not sufficient to derive general asymptotics, only when written in the form of WSV where the interaction of the eigenvectors and the moment function are crucial as seen in Theorem 1.

THEOREM 2 For any $\hat{\beta}^*$ s.t $n^{1/2}(\hat{\beta}^* - \beta_0) = O_p(1)$ under A1,A2.

$$\hat{V}(\hat{\beta}^*)^{-1/2}n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, I_p) \quad (16)$$

REMARKS

(i) Theorem 2 is in some ways surprising given the non-standard rate of convergence when $\mathcal{N}(\Omega) \not\subseteq \mathcal{N}(GG')$. Though in this case $n^{1/2}(\hat{\beta} - \beta_0)$ shrinks to zero in some directions, the square root of the variance $\hat{V}(\hat{\beta}^*)^{-1/2}$ diverges at an equal and opposite direction such that the statistic is non-degenerate with a $N(0, I_p)$ limiting distribution.

(ii) Weakly singular variance (such that eigenvalues do not converge too quickly to zero) does not impact inference based using standard asymptotics on $\hat{V}(\hat{\beta}^*)^{-1/2}n^{1/2}(\hat{\beta} - \beta_0)$ when moments are strongly identified. This result could also follow be shown to hold for the many weak moment setup of NW with for example $G = G/\mu_n$ for some $\mu_n = o(n^{1/2})$ under appropriate rate restrictions on δ .

(iii) Simulation evidence in Section 6 also demonstrate the poor small sample approximation of a $N(0, I_p)$ to the distribution of $\hat{V}(\hat{\beta}^*)^{-1/2}n^{1/2}(\hat{\beta} - \beta_0)$ the closer Ω_n is to singularity.

(vi) The limiting distribution is not impacted by whether $\hat{\beta}^* = \tilde{\beta}$ or $\hat{\beta}$ is used to form the statistic $\hat{V}(\hat{\beta}^*)^{-1/2}n^{1/2}(\hat{\beta} - \beta_0)$ as is also the case when Ω is full rank.

Define the Wald Statistic to test the restriction that $\beta = \beta_0$ ⁹,

$$\hat{W} = n(\hat{\beta} - \beta_0)' \hat{V}(\tilde{\beta})^{-1}(\hat{\beta} - \beta_0) \quad (17)$$

then Theorem 2 implies the asymptotic distribution of the Wald Statistic is not impacted by WSV under A1,A2.

⁹Note it is straightforward to extend the result to cover general linear restrictions, namely $R\beta_0 = c$ for some bounded full rank $R \in \mathbb{R}^{q \times p}$ for $1 \leq q \leq p$ and $c \in \mathbb{R}^q$.

COROLLARY 1 By Theorem 2 under A1,A2.

$$\hat{W} \xrightarrow{d} \chi_p^2 \tag{18}$$

PROOF: Note that $\hat{W} = (\hat{V}(\hat{\beta}^*)^{-1/2}n^{1/2}(\hat{\beta} - \beta_0))'\hat{V}(\hat{\beta}^*)^{-1/2}n^{1/2}(\hat{\beta} - \beta_0)$ which by the Continuous Mapping Theorem converges in distribution to χ_p^2 since $\hat{V}(\hat{\beta}^*)^{-1/2}n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, I_p)$ by Theorem 2.

Theorem 3 establishes the overidentifying restrictions test statistic (the Hansen J-Statistic) is asymptotically χ_{m-p}^2 when moments have WSV under A1, A2.

THEOREM 3 Under A1,A2,

$$\hat{Q}_{opt}(\hat{\theta}) \xrightarrow{d} \chi_{m-p}^2 \tag{19}$$

REMARKS

- (i) Despite potential singularity of Ω J-statistic which is a function $\hat{\Omega}(\tilde{\beta})^{-1}$ does not diverge, as detailed heuristically above, possessing the standard χ_{m-p}^2 limit distribution. Hence WSV does not lead to first order asymptotic size distortions to tests of overdiversifying restrictions.
- (ii) The limit distribution of $\hat{Q}_{opt}(\hat{\theta})$ does not depend on whether $\mathcal{N}(\Omega) \subseteq \mathcal{N}(GG')$ unlike the distribution of the 2-Step GMM estimator.
- (iii) Also unlike the distribution of the 2-step GMM estimator the distribution of the over-identification test statistic in the simulation in Section 6 is largely insensitive to the size of the eigenvalues of Ω_n in small samples. Theoretical justification for this result are beyond the scope of this paper, though an interesting avenue for future research.

3.2 Efficiency & Weakly Singular Variance

Theorem 1 highlights an interesting issue with regard to both the notion of efficiency in moment condition models. All of the literature concerning both efficient inference in general overidentified moment settings maintain that (some transformation of) moments have full rank (asymptotic) variance. Peñaranda & Sentana (2010) show efficiency of a modified GMM procedure when there exist known linearly redundant combinations at the true parameter. In this framework the directions of moment function causing singularity are removed and then 2-step GMM performed on the transformed moments with full rank variance. As such standard results in the efficiency and identification literature apply to this transformed moment function. No literature (outside of the invalid results on NSD) exists on the properties of general moment type estimators, and hence on the issue of efficient estimation or identification using 2-Step GMM when moments have asymptotically singular variance where no regularization is made to remove singularities in Ω .

Much literature is devoted to the issue of efficient inference and identification of GMM when Ω is full rank. Hansen (1982) shows that in the i.i.d setting under strong identification (along with the Ω full rank) and further regularity conditions that 2-Step GMM is $n^{1/2}$ consistent with a limiting variance matrix that attains the Semi Parametric Lower Bound (SPLB). Further results have been derived in more general settings by West, Wong and Anatolyev (2009) and others, though all this literature maintains the full rank variance assumption.

The results in this paper show that the notion of efficiency and identification are not limited solely to global and first-order identification, but inextricably linked to the properties of Ω_n (and the limit Ω) and its relationship with G . If $\mathcal{N}(\Omega) \subseteq \mathcal{N}(GG')$ then even when moments are weakly-singular and Ω is not full rank, 2-Step GMM is $n^{1/2}$ consistent with a Gaussian Limit of a similar form to the case when Ω is non-singular. If $\mathcal{N}(\Omega) \not\subseteq \mathcal{N}(GG')$ Theorem

1 shows that in some directions $n^{1/2}(\hat{\beta} - \beta_0)$ converges in some directions to the same Gaussian Limit at rate $n^{(1+\delta)/2}$. The analysis of efficiency and identification seem to be restricted by the assumption Ω is full rank and do not allow a general set of conditions when 2-Step GMM converges at the fastest rate. If $\mathcal{N}(\Omega) \not\subseteq \mathcal{N}(GG')$ and a regularization is performed such that the transformed moment has non-singular variance, then the rate of convergence will slow down. This seems to suggest the notion of efficient estimation and identification span further than the usual identification assumptions. This paper provides initial results on the the implications for inference of the methods used in the literature to overcome moments with almost singular variance.

4 Asymptotic Eigenvalue Expansions with Weakly-Singular Variance

This section provides asymptotic eigen-system expansions under the assumption moments are weakly singular. These results provide an extension to those provided in Grant (2013) for the case where \bar{m} eigenvalues equal zero whilst the remaining eigenvalues are bounded away from zero for all n . Without loss of generality in the following the first $m - \bar{m}$ eigenvalues are assumed to be bounded away from zero. Between the results in this paper and Grant (2013), asymptotic expansions are provided for zero, non-zero, and local-to-zero eigenvalues.

ASSUMPTION 2 : *Eigen-system Asymptotics*

(i) $w_i (i = 1, \dots, n)$ is an *i.i.d* sequence, (ii) $\mathbb{E}[\|g_i\|^4] < \infty$, (iii) $\frac{1}{n} \sum_{i=1}^n \|G_i(\beta) - G_i(\beta^*)\| \leq \hat{M} \|\beta - \beta^*\| \forall \beta, \beta^* \in \Theta$ 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) $\Omega_n(\beta)$ satisfies WSV for $\delta < 1$, (vii) $\|\Lambda\| \leq K$ for some $K < \infty$, (viii) $m < \infty$.

Let $\beta_n = \beta_0 + \Delta_n$ where $\Delta_n = O_p(n^{-1/2})$ ¹⁰.

¹⁰Eigen-system expansions for $\Delta_n = O_p(n^{-1/2})$ are considered since in the 2-Step GMM objective function we evaluate $\hat{\Omega}(\beta)^{-1}$ at $\beta = \tilde{\beta}$ where $\tilde{\beta} = \beta_0 + O_p(n^{-1/2})$ under A1,A2.

THEOREM 4 Under A2.

$$\hat{P}_+(\beta_n) = P_+ + O_p(n^{-1/2}) \quad (20)$$

$$\hat{\Lambda}_+(\beta_n) = \Lambda_+ + O_p(n^{-1/2}) \quad (21)$$

$$\hat{P}_0(\beta_n) = P_0 + O_p(n^{-1/2}) \quad (22)$$

$$n^\delta \hat{\Lambda}_0(\beta_n) = \Lambda_0 + o_p(1) \quad (23)$$

The expansions in Theorem 4 are critical to the derivation of the asymptotic properties of the GMM estimator with WSV in Section 3.

5 Examples of Weakly-Singular Variance

This section reproduces an example similar to that of Example 3 in Grant (2014) for linear IV simultaneous equations. This example is used in a simulation in Section 6 to demonstrate Theorem 1-3.

$$y_1 = x_1\beta_{01} + \epsilon_1 \quad (24)$$

$$y_2 = x_2\beta_{02} + \epsilon_2 \quad (25)$$

With a set of instruments $z = (z_1, z_2)'$ where $z \sim N(0, I_2)$ where β_{01} and β_{02} are the (scalar) true parameters and $\beta_0 := (\beta_{01}, \beta_{02})'$. Then if $E[\epsilon|z] = 0$ where $\epsilon = (\epsilon_1, \epsilon_2)'$ any function of the moments z will serve as valid instruments when interacted with the residual functions $\epsilon_1(\beta) = y_1 - x_1\beta_1$, $\epsilon_2(\beta) = y_2 - x_2\beta_2$ where $\beta := (\beta_1, \beta_2)'$, $\beta_1, \beta_2 \in \mathbb{R}$.

$$x_1 = f_1(z) + \eta_1 \quad (26)$$

$$x_2 = f_2(z) + \eta_2 \quad (27)$$

For some measurable functions $f_1(\cdot), f_2(\cdot)$ where $E[\eta_1|z] = E[\eta_2|z] = 0$ and $E[\eta_1^2|z] = E[\eta_2^2|z] = 1$.

$$g(\beta) = \begin{pmatrix} \epsilon_1(\beta)\phi_1(z) \\ \epsilon_2(\beta)\phi_2(z) \end{pmatrix} \quad (28)$$

With $\beta := (\beta_1, \beta_2)$ where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are resp. $m_1 \times 1, m_2 \times 1$ vector functions of the instruments z . Suppose the unobservable residuals in (24), (25) satisfy,

$$\epsilon_1 = v_1 h_1(z) \quad (29)$$

$$\epsilon_2 = v_2 h_2(z) \quad (30)$$

for some measurable functions $h_1(\cdot), h_2(\cdot)$ which allow potentiality of heteroscedasticity of ϵ_1, ϵ_2 in the instruments. For simplicity assume $E[v_1|z] = E[v_2|z] = 0$ and $E[v_1^2|z] = E[v_2^2|z] = 1$ and $E[v_1 v_2|z] = \rho$. Then almost singular variance may arise if v_1, v_2 are highly correlated conditional on z , where $\phi_1(z)$ and $\phi_2(z)$ are such that when interacted with $\epsilon_1(\beta) = y_1 - x_1\beta_1, \epsilon_2(\beta) = y_2 - x_2\beta_2$ at $\beta = \beta_0$, some linear combination of $\epsilon_1\phi_1(z), \epsilon_2\phi_2(z)$ become almost redundant. Detailed examples were given in the general case in Grant (2014), when v_1, v_2 are perfectly or near perfectly correlated, when $\phi_1(z), \phi_2(z)$ may be approximating functions (e.g polynomials in z).

Take the simple case where $h_1(z) = z_2$ and $h_2(z) = z_1$. Then if $\phi_1(z) = (z_1, z_2)$ and $\phi_2(z) = z_2$, $\epsilon_1 z_1 = v_1 z_1 z_2$ and $\epsilon_2 z_2 = v_2 z_1 z_2$ such that $cov(\epsilon_1 z_1, \epsilon_2 z_2) = E[v_1 v_2 (z_1 z_2)^2] = E[v_1 v_2] E[z_1^2] E[z_2^2]$ by the Law of Iterated Expectations under the assumption $E[v_1 v_2|z] = \rho$. Hence under the assumptions made,

$$cor(\epsilon_1 z_1, \epsilon_2 z_2) = \rho \quad (31)$$

where if $\rho = 1 - n^{-\delta}$ for some $0 < \delta < 1$ then it is straightforward to show that the smallest eigenvalue of Ω_n is $O(n^{-\delta})$. Notice that although the variance of the moment function is weakly singular, this places no restriction on the correlation in the residuals in (24), (25). In this example the residuals ϵ_1, ϵ_2 are uncorrelated since $E[\epsilon_1 \epsilon_2] = E[v_1 v_2] E[z_1 z_2] = 0$ since $E[z_1 z_2] = 0$. 2-Step GMM will converge at rate faster than $n^{1/2}$ when there exist $\delta' \Omega = 0$ such that $\delta' G \neq 0$ where $\Omega = \lim_{n \rightarrow \infty} \Omega_n$. Under the assumptions made on the data generating process then as $\rho \rightarrow 1$ and $E[z_2^4] = 3$.

$$\Omega = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad (32)$$

Then in this example $\bar{m} = 1$ and $\Omega = P_+ \Lambda_+ P_+'$ where $P_+ = \begin{pmatrix} 0 & \sqrt{2} \\ -1 & 0 \\ 0 & -\sqrt{2} \end{pmatrix}$ and $\Lambda_+ = \text{diag}(3, 2)$ where $P_0 = (\sqrt{2}, 0, -\sqrt{2})'$. In this Linear IV setup then,

$$G = - \begin{pmatrix} E[x_1 z_1] & 0 \\ E[x_1 z_2] & 0 \\ 0 & E[x_2 z_2] \end{pmatrix} \quad (33)$$

so that $G' P_0 = -\sqrt{2} \begin{pmatrix} E[x_1 z_1] \\ -E[x_2 z_2] \end{pmatrix}$. Take for example the case $f_1(z) = 2(z_1 + z_2)$, $f_2(z) = 2z_2$ then $E[x_1 z_1] = 2$ and $E[x_2 z_2] = E[z_1 z_2] = 0$. Hence $G' P_0$ is already in Gauss-Jordan form, so that $B = I_{2 \times 2}$ will satisfy (12). Hence $\begin{pmatrix} n^{\delta/2} & 0 \\ 0 & 1 \end{pmatrix} n^{1/2}(\hat{\beta} - \beta_0)$ satisfies Theorem 1 and has a Gaussian limit distribution. So that the distribution of $n^{1/2}(\hat{\beta}_1 - \beta_{01})$ will be degenerate unless scaled by $n^{\delta/2}$. This example is used in the simulation in the next section to demonstrate Theorem 1-3.

6 Simulation

This section demonstrates the main results of the paper using a simulation based on the example highlighted in Section 5.

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

$$x_1 = 2z_2 + \eta_1 \quad x_2 = 2(z_1 + z_2) + \eta_2$$

Where $z = (z_1, z_2)' \sim N(0, I_2)$,

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

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

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

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

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 $W_n = I_{2 \times 2}$.

Consider $\delta = \{0, 0.25, 0.5, 0.75, 0.9\}$ corresponding to varying degrees of correlation between the first and third moment for

$n = \{500, 5000, 10000, 50000, 100000, 250000\}$. Table 1 tabulates the mean absolute deviation (MAD) of $b_1 := n^{1/2}(\hat{\beta}_1 - 1)$, $b_2 := n^{1/2}(\hat{\beta}_2 - 0.5)$. Also the rejection probabilities of the Wald-Statistic testing that $\beta_0 = (1, 0.5)'$ based on a χ_2^2 distribution as governed by Corollary 1 and and the J-Statistic are also tabulated based on a χ_1^2 distribution governed by Theorem 3.

As seen in Table 1 the mean absolute deviation of b_1 for any n and is decreasing

in n , whereas b_2 is bounded as predicted by Theorem 1 in light of the discussion in Section 5 for this example. The Wald Statistic is better approximated by a χ_2^2 distribution as n increases, with the approximation poorer the nearer δ is to 1 (i.e the closer ρ is to 1). Even for sample size $n = 250000$ the rejection probabilities based on a χ_2^2 distribution are undersized. For $\delta = \{0, 0.25, 0.5, 0.75\}$ as n increases the Wald Statistic is approximated well by a χ_2^2 distribution. By Theorem 2 eventually the Wald Statistic will approach a χ_2^2 however not for $\delta = 0.9$ for even the large sample sizes considered in this simulation. Interestingly the J-Statistic is well approximated by a χ_1^2 for even small values of n for all values of δ . This suggests potentially the finite sample distribution of the J-Statistic is not impacted as severely as the Wald Statistic for ρ close to 1.

This paper made assumptions that the eigenvalues were shrinking fast enough so that first order asymptotics (once appropriately scaled by B_n) are not impacted by the fact Ω_n is asymptotically singular. If $\delta = 1$ or moments had exactly singular variance then both the rate of convergence and the limit distribution become non-standard. This issue is beyond the scope of this paper, being considered in Grant (2014), however the simulation highlights this non-standard distribution of b_1 , b_2 and the Wald Statistic for moments close to singular.

7 Conclusion

This paper considers the properties of the efficient 2-step GMM estimator when moments have variance that is close to singular at the true parameter β_0 . To model this situation the Weakly-Singular Variance assumption is proposed, where (a subset) of the the eigenvalues of the moment variance matrix converge to zero at some rate. This assumption is shown in some cases to be equivalent to the Nearly-Singular Design assumption employed in Caner (2008), where the results in this paper are shown to be incorrect.

When moments have WSV the 2-Step GMM estimator when appropriately scaled converges in distribution to a Gaussian limit distribution when the smallest eigenvalue of the moment variance matrix converge slowly enough to zero, namely when they are $O(n^{-\delta})$ for some $0 \leq \delta < 1$. The rate of convergence is shown to be the standard $n^{1/2}$ rate when the null space of the moment variance matrix is a subset or equal to the null space of the outer product of the expected first order derivative of the moment function of β_0 . When this condition is violated the convergence occurs at rate $n^{(1+\delta)/2}$ in certain directions. Results in this paper show singularity of the moment variance can in certain cases lead to an estimator closer to the true parameter given the increased rate of convergence.

Commonly when (almost) singular variance is encountered in practise some form of regularisation or otherwise is performed to remove the singularity, e.g Doran& Schmidt (2006). The methods employed are often ad hoc with little theoretical grounding. The main results of this paper show these methods could lead to an estimator with poorer efficiency properties given a potential reduction in the rate of convergence.

Simulation evidence demonstrates that the distribution of the 2-Step GMM estimator is poorly approximated by a Gaussian distribution for δ close to 1 and hence the closer to singularity the moment variance matrix even for very large sample sizes. Showing theoretically why this is occurs is beyond the scope of this paper, being covered in Grant (2014). This paper shows in this case the eigenvalues of the moment variance matrix converge fast enough to zero, the limit distribution of the 2-Step GMM estimator when appropriately scaled is highly non-standard.

The results in this paper are the first to give the theoretical implications of the moment variance being close to singular when performing efficient inference in overidentified moment conditon models. The link between singular variance and identification and efficiency is discussed, where further research in the link

between the two is an interesting avenue for further research.

Appendix

Appendix A1: Moment Variance Eigensystem Definitions

By construction both $\Omega_n(\beta)$, $\hat{\Omega}(\beta)$ along with $\Omega(\beta) = \lim_{n \rightarrow \infty} \Omega_n(\beta)$ are p.s.d and symmetric hence the following decompositions can be made for all $\beta \in \mathbf{B}$. Let the $m \times m$ matrix $P(\beta)$ be the matrix of population eigenvalues Such that $P(\beta)'P(\beta) = I_{m \times m}$. Define the rank of $\Omega(\beta)$ as $m - \bar{m}(\beta)$ where $0 \leq m(\beta) \leq m$. Express $P(\beta) = (P_+(\beta), P_0(\beta))$ and $\Lambda_n(\beta) = \begin{pmatrix} \Lambda_+(\beta) & 0 \\ 0 & \Lambda_{0n}(\beta) \end{pmatrix}$ where $\Lambda_+(\beta)$ is an $(m - \bar{m}(\beta)) \times (m - \bar{m}(\beta))$ diagonal matrix with the non-zero eigenvalues of $\Omega_n(\beta)$ on the diagonal with corresponding $m \times (m - \bar{m}(\beta))$ eigenvector matrix $P_+(\beta)$. Λ_{0n} is the $\bar{m}(\beta) \times \bar{m}(\beta)$ matrix containing the asymptotically zero eigenvalues as defined in WSV in Section 2 with corresponding $m \times \bar{m}(\beta)$ eigenvector matrix $P_0(\beta)$. Note that the eigenvectors are not modelled as a function of n , doing so would not change the results of this paper, where only the size of the eigenvalues and hence the rank of $\Omega_n(\beta)$ impact the asymptotic properties of GMM. Performing an eigenvalue decomposition re-write $\Omega_n(\beta)$ as,

$$\Omega(\beta) = P_+(\beta)\Lambda_+(\beta)P_+(\beta)' + P_0(\beta)\Lambda_{0n}(\beta)P_0(\beta)'$$

performing a similar decomposition for $\hat{\Omega}(\beta)$

$$\hat{\Omega}(\beta) = \hat{P}_+(\beta)\hat{\Lambda}_+(\beta)\hat{P}_+(\beta)' + \hat{P}_0(\beta)\hat{\Lambda}_0(\beta)\hat{P}_0(\beta)'$$

where $\hat{P}_+(\beta)$ is an $(m - \bar{m}(\beta)) \times (m - \bar{m}(\beta))$ matrix of sample eigenvector estimates of $P_+(\beta)$ with corresponding sample eigenvalue $\hat{\Lambda}_+(\beta)$. $\hat{P}_0(\beta)$ and $\hat{\Lambda}_0(\beta)$ are similarly the sample estimates of $P_0(\beta)$ and $\Lambda_0(\beta)$ respectively letting $\hat{P}(\beta) := (\hat{P}_+(\beta), \hat{P}_0(\beta))$.

Define $\Omega_n = \Omega_n(\beta_0)$, $\Omega = \Omega(\beta_0)$, $\hat{\Omega} = \hat{\Omega}(\beta_0)$ and $\bar{m}(\beta_0) := \bar{m}$ for notational simplicity throughout and let the eigenvalues/vector matrices of Ω_n , Ω , $\hat{\Omega}$ be defined without β_0 , for example $P := P(\beta_0)$, $\hat{P} := \hat{P}(\beta_0)$ and so on. Finally $\Omega_+ = P_+ \Lambda_+ P_+' and $\Omega_+^* = P_+ \Lambda_+^{-1} P_+'.$$

Appendix A2: Notation and Definitions

For any random variables y, x $\mathbb{E}[y], E[y|x]$, refers to the mathematical expectation of y and expectation of y with respect to the density of x respectively. Denote $\xrightarrow{p}, \xrightarrow{d}$, as convergence in probability and convergence in distribution respectively. For any deterministic sequence a_n and constant b then $a_n \rightarrow b$ denotes b as the deterministic limit of a_n . $\stackrel{d}{\sim}$ is shorthand for ‘is distributed as’ and ‘w.p.a.1’ denotes ‘with probability approaching 1’. We use $o_p(a)$ to refer to a variable that converges to zero w.p.a.1 when divided by a and similarly $O_p(a)$ a variable bounded in probability when divided by a . Let $\mathbb{R}^{r \times s}$ refer to the space of all matrices of dimension $r \times s$. Define $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. Let A refer to any arbitrary matrix then define $\mathcal{N}(\cdot)$ as null space of A were $\dim(A)$ denotes the dimension of a matrix A . Let $\text{int}(C)$ refer to the interior of the set C . For any $a > 0$ $I_{a \times a}$ refers to the $a \times a$ Identity Matrix. Define p.d and p.s.d as positive definite and positive semi definite respectively. Let M,CS,T refer to the Markov, Chebychev and Triangle Inequality respectively (resp.). Also let CMT refer to the Continuous Mapping Theorem. Define KWLLN as the ‘Khinchine Weak Law of Large Numbers and LLCT refer to the Lindberg-Lévy Central Limit Theorem.

Appendix B: Proofs of Lemmas

PROOF OF LEMMA 1:

Firstly show WSV (for $\delta < 1/2$) implies NSD when Ω is singular (focussing only at the point $\beta = \beta_0$ for simplicity. Results could be extended to show across $\beta \in \mathbf{B}$).

Take WSV for $\delta < 1/2$ when Ω is singular and hence $\bar{m} > 0$. By definition this implies $\Omega_n = P_+\Lambda_+P_+ + P_0\Lambda_{0n}P_0'$ where $\Lambda_{0n} = \Lambda_0/n^\delta$ and $\Omega = P_+\Lambda_+P_+' since by WSV $\Lambda_{0n} \rightarrow 0$ as $\delta > 0$. Then,$

$$n^\delta(\hat{\Omega} - \Omega) = n^\delta(\Omega_n - \Omega) + n^\delta(\hat{\Omega} - \Omega_n) = P_0\Lambda_0P_0' + O_p(n^{-1/2})n^\delta \quad (34)$$

since by assumption $\hat{\Omega} - \Omega_n = O_p(n^{-1/2})$. Hence for $\delta < 1/2$ then WSV implies,

$$n^\delta(\hat{\Omega} - \Omega) \xrightarrow{p} P_0\Lambda_0P_0' \quad (35)$$

which is NSD with $a_n = n^\delta$ and $D = P_0\Lambda_0P_0'$ where $P_0'DP_0 = \Lambda_0$ which by WSV is assumed full rank.

Now to show NSD implies WSV under the rate restrictions in Lemma 1 when Ω is singular. Take NSD for $a_n = n^\kappa$ for $\kappa < 1/2$ NSD implies $a_n(\hat{\Omega} - \Omega) \xrightarrow{p} D$ where $P_0'DP_0$ is full rank.

Then $a_n(\hat{\Omega} - \Omega) = D + o_p(1)$ such that,

$$P_+' \hat{\Omega} P_+ = P_+' \Omega P_+ D + o_p(a_n^{-1}) \quad (36)$$

where by assumption $\hat{\Omega} = \Omega_n + O_p(n^{-1/2})$ which implies,

$$P_+' \hat{\Omega} P_+ = P_+' \Omega_n P_+ + O_p(n^{-1/2}) \quad (37)$$

where (36), (37) imply that $P_+' \Omega_n P_+ = P_+' \Omega P_+ = \Lambda_+$ w.p.a.1 Since Ω_n and Ω are non-stochastic then $P_+' \Omega_n P_+ = P_+' \Omega P_+ = \Lambda_+$. Now to show that NSD implies $P_0'\Omega_n P_0 = H_n$ where $a_n H_n = H$ where H is full rank and p.d. This

establishes WSV for $H = \Lambda_0$.

By (36) then,

$$P'_0 a_n \hat{\Omega} P_0 = P'_0 D P_0 + o_p(1) \quad (38)$$

since $P'_0 \Omega P_0 = 0$ under singularity, again using $\hat{\Omega} = \Omega_n + O_p(n^{-1/2})$ where $\|P_0\|^2 = \bar{m} = O(1)$ under A2(viii) that m is finite imply,

$$P'_0 a_n \hat{\Omega} P_0 = P'_0 a_n \Omega_n P_0 + a_n O_p(n^{-1/2}) \quad (39)$$

where $a_n O_p(n^{-1/2}) = o_p(1)$ by assumption on a_n . Similar to the reasoning above (38),(39) imply $a_n P'_0 \Omega_n P_0 = P'_0 D P_0 := H$ where $P'_0 D P_0$ is full rank and p.d by NSD. This establishes that $\Omega_n = P_+ \Lambda_+ P'_+ + P_0 \Lambda_0 / a_n P'_0$ for some full rank p.d matrix Λ_0 .

Appendix C: Proof of Main Theorems

PROOF OF THEOREM 1

We first show (13), namely that $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$, which will prove useful when deriving the limit distribution in (14). To do this we first show $\hat{Q}(\hat{\beta})$ is bounded.

$$\hat{Q}_{opt}(\hat{\beta}) = O_p(1) \quad (40)$$

By definition of $\hat{\beta}$,

$$\hat{Q}_{opt}(\hat{\beta}) \leq \hat{Q}_{opt}(\beta_0) \quad (41)$$

where $\hat{Q}_{opt}(\beta_0) = \hat{Q}_{opt}^+(\beta_0) + \hat{Q}_{opt}^0(\beta_0)$. We show both $\hat{Q}_{opt}^+(\beta_0) = O_p(1)$, $\hat{Q}_{opt}^0(\beta_0)$ are $O_p(1)$ establishing $\hat{Q}_{opt}(\hat{\beta}) = O_p(1)$.

By repeated applications of CS,

$$\hat{Q}_{opt}^+(\beta_0) \leq \|n^{1/2}\hat{g}(\beta_0)\| \|\hat{P}_+\|\hat{\Lambda}_+(\tilde{\beta})\|^{-1} \quad (42)$$

where $\|\hat{P}_+\| = \bar{m} = O(1)$ by A2(viii) and $\|n^{1/2}\hat{g}(\beta_0)\| = O_p(1)$ by A2(i),(ii),(viii) and $\hat{\Lambda}_+(\tilde{\beta}) = O_p(1)$ since by A1, and given $\tilde{\beta} = \beta_0 + o_p(1)$ then $\hat{\Lambda}_+(\tilde{\beta}) = \Lambda_+ + o_p(1)$ by Theorem 4 where $\|\Lambda_+\| = O(1)$ by A2(vii).

Now to show $\hat{Q}_{opt}^0(\beta_0) = O_p(1)$ multiplying and dividing by n^δ to express $\hat{Q}_{opt}^0(\beta_0)$ as,

$$\hat{Q}_{opt}^0(\beta_0) = n^{1+\delta}\hat{g}(\beta_0)'\hat{P}_0(\tilde{\beta})(n^\delta\hat{\Lambda}_0(\tilde{\beta}))^{-1}\hat{P}_0(\tilde{\beta})'\hat{g}(\beta_0) \quad (43)$$

where by repeated application of CS,

$$\hat{Q}_{opt}^0(\beta_0) \leq \|n^{(1+\delta)/2}\hat{P}_0(\tilde{\beta})'\hat{g}(\beta_0)\|^2 \|n^\delta\hat{\Lambda}_0(\tilde{\beta})\|^{-1} \quad (44)$$

where $(n^\delta\hat{\Lambda}_0(\tilde{\beta}))^{-1} = \Lambda_0^{-1} + o_p(1)$ by Theorem 4 since $\tilde{\beta} = \beta_0 + O_p(n^{-1/2})$ and where $\|\Lambda_0\| = O(1)$ by A2(vii) so that,

$$\|\hat{\Lambda}_0(\tilde{\beta})\|^{-1} = O_p(1) \quad (45)$$

by CMT and since $m = O(1)$ by A1(viii). By T,

$$\|n^{(1+\delta)/2}\hat{P}_0(\tilde{\beta})'\hat{g}(\beta_0)\| \leq \|n^{(1+\delta)/2}(\hat{P}_0(\tilde{\beta}) - P_0)'\hat{g}(\beta_0)\| + \|n^{(1+\delta)/2}P_0'\hat{g}(\beta_0)\| \quad (46)$$

where $n^{1/2}\|\hat{g}(\beta_0)\| = O_p(1)$ as shown above and $\|n^{1/2}(\hat{P}_0(\tilde{\beta}) - P_0)\| = O_p(1)$ by Theorem 4 then since $\delta < 1$ by repeated application of CS it is straightforward to establish (47).

$$\|n^{(1+\delta)/2}(\hat{P}_0(\tilde{\beta}) - P_0)'\hat{g}(\beta_0)\| = o_p(1) \quad (47)$$

Finally as $Var(n^{(1+\delta)/2}P_0'n^{1/2}\hat{g}(\beta_0)) = \Lambda_0$ where Λ_0 is bounded by A2(vii)

and g_i is i.i.d by A1(i) then by the LLCT,

$$n^{(1+\delta)/2} P_0' \hat{g}(\beta_0) \xrightarrow{d} N(0, \Lambda_0) \quad (48)$$

so that,

$$\|n^{(1+\delta)/2} P_0' \hat{g}(\beta_0)\| = O_p(1) \quad (49)$$

where (47), (49) together with (46) establishes,

$$\|n^{(1+\delta)/2} \hat{P}_0(\tilde{\beta})' \hat{g}(\beta_0)\| = O_p(1) \quad (50)$$

which together with (45) implies $\hat{Q}_{opt}^0(\hat{\beta}) = O_p(1)$.

We now use $\hat{Q}_{opt}(\hat{\beta}) = O_p(1)$ to show (13).

$$\hat{Q}_{opt}(\hat{\beta}) \geq \hat{P}(\hat{\beta})' \hat{g}(\hat{\beta})^2 / \sup_j \hat{\lambda}_j(\tilde{\beta}) \quad (51)$$

Since $\hat{Q}_{opt}(\hat{\beta}) \geq \inf_j \hat{\lambda}_j(\tilde{\beta})^{-1} (\hat{P}(\hat{\beta})' \hat{g}(\hat{\beta}))^2$.

Note that $\hat{\lambda}_j(\tilde{\beta}) = \lambda_j + (\hat{\lambda}_j(\tilde{\beta}) - \lambda_j)$ for all $j = \{1, \dots, m\}$.

Since $\|\hat{\Lambda}(\tilde{\beta}) - \Lambda\| \leq \|\hat{\Omega}(\tilde{\beta}) - \Omega\|$ by Lemma 4.2 of Bosq (2000) where $\hat{\lambda}_j(\tilde{\beta}) - \lambda_j \leq \|\hat{\Lambda}(\tilde{\beta}) - \Lambda\|$ for all $j = \{1, \dots, m\}$ so that,

$$\sup_j \hat{\lambda}_j(\tilde{\beta}) \leq \sup_j \lambda_j + \|\hat{\Omega}(\tilde{\beta}) - \Omega\| \quad (52)$$

where $\|\hat{\Omega}(\tilde{\beta}) - \Omega\| = o_p(1)$ by (85) since $\tilde{\beta} = \beta_0 + O_p(n^{-1/2})$ and $\delta < 1$ by assumption where $\sup_j \lambda_j < K < \infty$ for some $K > 0$ by A2(vii) hence along with (52) implies,

$$\sup_j \hat{\lambda}_j \leq K + o_p(1) \quad (53)$$

noting also that $\sup_j \hat{\lambda}_j > 0$ w.p.a.1 since $\sup_j \lambda_j > 0$ by WSV then (53) along

with (51) implies,

$$\hat{Q}_{opt}(\hat{\beta}) \geq (K + o_p(1)) \|n^{1/2} \hat{g}(\hat{\beta})\|^2 \|\hat{P}(\tilde{\beta})\|^2 \quad (54)$$

by repeated application of CS. By definition $\|\hat{P}(\tilde{\beta})\|^2 = m = O(1)$ by A2(viii). The finally using the identification from by A1(ii) for some $C > 0$,

$$\|n^{1/2} \hat{g}(\hat{\beta})\|^2 \geq (C \|n^{1/2}(\hat{\beta} - \beta_0)\| + \hat{M})^2 \quad (55)$$

since $\hat{M} = O_p(1)$ and $\|n^{1/2}(\hat{\beta} - \beta_0)\| \geq 0$ then since $\hat{Q}_{opt}(\hat{\beta}) \leq O_p(1)$ by (40) imply $\|n^{1/2}(\hat{\beta} - \beta_0)\| = O_p(1)$ which establishes (13).

We now establish the limit distribution of GMM as stated in (14). By definition $\hat{\beta}$ solves,

$$\hat{G}(\hat{\beta})' \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\hat{\beta}) = 0 \quad (56)$$

performing a Mean Value Expansion around β_0 ,

$$n^{1/2}(\hat{\beta} - \beta_0) = -(\hat{G}(\hat{\beta})' \hat{\Omega}(\tilde{\beta})^{-1} \hat{G}(\tilde{\beta}))^{-1} \hat{G}(\hat{\beta})' \hat{\Omega}(\tilde{\beta})^{-1} n^{1/2} \hat{g}(\beta_0) \quad (57)$$

where $\tilde{\beta} = \alpha \hat{\beta} + (1 - \alpha) \beta_0$ for some potentially data dependent $\alpha \in [0, 1]$.

Define $\hat{A}_1 = -(\hat{G}(\hat{\beta})' \hat{\Omega}(\tilde{\beta})^{-1} \hat{G}(\tilde{\beta}))$ and $\hat{A}_2 = \hat{G}(\hat{\beta})' \hat{\Omega}(\tilde{\beta})^{-1} n^{1/2} \hat{g}(\beta_0)$ hence $n^{1/2}(\hat{\beta} - \beta_0) = \hat{A}_1^{-1} \hat{A}_2$. We can rewrite $\hat{A}_1^{-1} \hat{A}_2 = B_n^{-1} (B_n'^{-1} \hat{A}_1 B_n^{-1})^{-1} B_n'^{-1} \hat{A}_2$ where $B_n n^{1/2}(\hat{\beta} - \beta_0) = (B_n'^{-1} \hat{A}_1 B_n^{-1})^{-1} B_n'^{-1} \hat{A}_2$.

$$B_n'^{-1} \hat{A}_1 B_n^{-1} = I_n^{-1} B' \hat{G}(\hat{\beta})' \hat{\Omega}(\tilde{\beta})^{-1} \hat{G}(\tilde{\beta}) B I_n^{-1} \quad (58)$$

$$= \hat{A}_3 + \hat{A}_4 \quad (59)$$

where $\hat{A}_3 := I_n^{-1} B' \hat{G}(\hat{\beta})' \hat{P}_+(\tilde{\beta}) \hat{\Lambda}_+(\tilde{\beta})^{-1} \hat{P}_+(\tilde{\beta})' B I_n^{-1}$,

$\hat{A}_4 := I_n^{-1} B' \hat{G}(\hat{\beta})' \hat{P}_0(\tilde{\beta}) \hat{\Lambda}_0(\tilde{\beta})^{-1} \hat{P}_0(\tilde{\beta})' \hat{G}(\tilde{\beta}) B I_n^{-1}$. Firstly we show,

$$\hat{A}_3 = \begin{pmatrix} 0_{\bar{p} \times \bar{m}} \\ G_{BP_+}^{p-\bar{p}} \end{pmatrix} \Lambda_+^{-1} \begin{pmatrix} 0_{\bar{p} \times \bar{m}} \\ G_{BP_+}^{p-\bar{p}} \end{pmatrix}' + o_p(1) \quad (60)$$

by A2(iii) $\hat{G}(\hat{\beta}) = G + O_p(\|\hat{\beta} - \beta_0\|)$ where $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/2})$ by (13) and,

$$\hat{G}(\hat{\beta})' \hat{P}_+(\tilde{\beta}) = G' P_+ + O_p(n^{-1/2}) \quad (61)$$

since $\hat{P}_0(\tilde{\beta}) = P_0 + O_p(n^{-1/2})$ by Theorem 4. Since G and P_+ are bounded matrices this establishes (61).

Since B is bounded by assumption then by (61) and CMT imply,

$$B \hat{G}(\hat{\beta})' \hat{P}_+(\tilde{\beta}) = B' G' P_+ + O_p(n^{-1/2}) \quad (62)$$

where we define $B' G' P_+ := G_{BP_+}$. Noting that $I_n^{-1} = \begin{pmatrix} n^{-\delta/2} I_{\bar{p}} & 0_{\bar{p} \times \bar{p}} \\ 0_{(p-\bar{p}) \times (p-\bar{p})} & I_{p-\bar{p}} \end{pmatrix}$

where $\|I_n^{-1}\| = O(1)$ along with (62) imply,

$$I_n^{-1} B' \hat{G}(\hat{\beta})' \hat{P}_+(\tilde{\beta}) = I_n^{-1} G_{BP_+} + O_p(n^{-1/2}) \quad (63)$$

where $I_n^{-1} G_{BP_+} \rightarrow \begin{pmatrix} 0_{\bar{p} \times \bar{m}} \\ G_{BP_+}^{p-\bar{p}} \end{pmatrix}$ where $G_{BP_+} = \begin{pmatrix} G_{BP_+}^{\bar{p}} \\ G_{BP_+}^{p-\bar{p}} \end{pmatrix}$ as $\delta > 0$, together with (63) implies (60).

By definition $\hat{A}_4 := I_n^{-1} B' \hat{G}(\hat{\beta})' \hat{P}_0(\tilde{\beta}) \hat{\Lambda}_0(\tilde{\beta})^{-1} \hat{P}_0(\tilde{\beta})' \hat{G}(\tilde{\beta}) B I_n^{-1}$,

$$\hat{A}_4 = n^{\delta/2} I_n^{-1} B' \hat{G}(\hat{\beta})' \hat{P}_0(\tilde{\beta}) (n^\delta \hat{\Lambda}_0(\tilde{\beta}))^{-1} \hat{P}_0(\tilde{\beta})' \hat{G}(\tilde{\beta}) B n^{\delta/2} I_n^{-1} \quad (64)$$

$$= n^{\delta/2} I_n^{-1} B' G P_0 (n^\delta \hat{\Lambda}_0(\tilde{\beta}))^{-1} P_0' G B n^{\delta/2} I_n^{-1} + o_p(1) \quad (65)$$

$$= \begin{pmatrix} G_{BP_0}^{\bar{p}} \\ 0_{(p-\bar{p}) \times \bar{m}} \end{pmatrix} \Lambda_0^{-1} \begin{pmatrix} G_{BP_0}^{\bar{p}} \\ 0_{(p-\bar{p}) \times \bar{m}} \end{pmatrix}' + o_p(1) \quad (66)$$

where (64) follows by multiplying and deriving by n^δ and (65) since $B'\hat{G}(\hat{\beta})'\hat{P}_0(\tilde{\beta}) = B'G'P_0 + O_p(n^{-1/2})$ as shown similarly above for $B'\hat{G}(\hat{\beta})'\hat{P}_0(\tilde{\beta})$ under A1,A2 and Theorem 4 and since $I_n^{-1}n^{\delta/2} = O(n^{\delta/2})$ then $n^{\delta/2}I_n^{-1}B'\hat{G}(\hat{\beta})'\hat{P}_0(\tilde{\beta}) = n^{\delta/2}I_n^{-1}B'G'P_0 + O_p(n^{\delta/2})O_p(n^{-1/2}) = n^{\delta/2}I_n^{-1}G'P_0 + o_p(1)$ since $\delta < 1$ then $(n^\delta\hat{\Lambda}_0(\tilde{\beta}))^{-1} = O_p(1)$ which holds under A1,A2 by Theorem 4 shows (65).

Note by definition of B , $B'G'P_0 = \begin{pmatrix} G_{BP_0}^{\bar{p}} \\ 0_{(p-\bar{p}) \times \bar{m}} \end{pmatrix}$.

As $n^{\delta/2}I_n^{-1} = \begin{pmatrix} I_{\bar{p}} & 0_{\bar{p} \times \bar{p}} \\ 0_{(p-\bar{p}) \times (p-\bar{p})} & n^{\delta/2}I_{p-\bar{p}} \end{pmatrix}$ then $n^{\delta/2}I_n^{-1}B'G'P_0 = \begin{pmatrix} G_{BP_0}^{\bar{p}} \\ 0_{(p-\bar{p}) \times \bar{m}} \end{pmatrix}$

where $(n^\delta\hat{\Lambda}_0(\tilde{\beta}))^{-1} \xrightarrow{p} \Lambda_0^{-1}$ by Theorem 4 which together along with the CMT establish (66).

Together (60),(66) and the CMT imply,

$$B_n'^{-1}\hat{A}_1B_n^{-1} = G'_{BP}\Lambda^{-1}G_{BP} + o_p(1) \quad (67)$$

where $G_{BP} := \begin{pmatrix} G_{BP_+}^{\bar{p}} & 0_{\bar{p} \times \bar{m}} \\ 0_{(p-\bar{p}) \times (m-\bar{m})} & G_{BP_0}^{p-\bar{p}} \end{pmatrix}'$ and $\Lambda = \text{diag}(\Lambda_+, \Lambda_0)$.

We now show that,

$$B_n'^{-1}\hat{A}_2 \xrightarrow{d} N(0, G'_{BP}\Lambda^{-1}G_{BP}) \quad (68)$$

rewriting $B_n'^{-1}\hat{A}_2 = \hat{A}_5\hat{A}_6$ where $\hat{A}_5 = B_n'^{-1}\hat{G}(\hat{\beta})'\hat{\Omega}(\tilde{\beta})^{-1/2}$ and $\hat{A}_6 = n^{1/2}\hat{\Omega}(\tilde{\beta})^{-1/2}\hat{g}(\beta_0)$.

First note that $\hat{A}_5\hat{A}_5' = B_n'^{-1}\hat{G}(\hat{\beta})'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G}(\hat{\beta})B_n^{-1}$ where as shown similarly for $B_n'^{-1}\hat{A}_1B_n^{-1} = B_n'^{-1}\hat{G}(\hat{\beta})'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G}(\tilde{\beta})B_n^{-1} \xrightarrow{p} G'_{BP}\Lambda^{-1}G_{BP}$ by (67). Noting that $\hat{\beta} = \tilde{\beta} + O_p(n^{-1/2})$ we can likewise show $\hat{A}_5\hat{A}_5' \xrightarrow{p} G'_{BP}\Lambda^{-1}G_{BP}$.

Defining $A_5 = G'_{BP}\Lambda^{-1/2}$,

$$\hat{A}_5\hat{A}_5' \xrightarrow{p} A_5A_5' \quad (69)$$

then $\|x' \hat{A}_5\|^2 = x' \hat{A}_5 \hat{A}_5' x \xrightarrow{p} \|x' A_5\|^2$ for all $x \in \mathbb{R}^p$. For example take the scalar case then $\hat{A}_5^2 \xrightarrow{p} A_5^2$ where $\hat{A}_5 = |\hat{A}_5^2|^{1/2} \xrightarrow{p} |A_5^2|^{1/2} = A_5$ by CMT since $f(\cdot) = |\cdot|^{1/2}$ is a continuous function of A_5 . Hence by the definition of convergence in probability for matrices.

$$\hat{A}_5 \xrightarrow{p} A_5 \quad (70)$$

We now show $\hat{A}_6 \xrightarrow{d} N(0, I_m)$. For brevity we omit the dependence of the eigenvalues/vectors of $\hat{\Omega}(\tilde{\beta})$ on, $\tilde{\beta}$ such that $\hat{P}_0 = \hat{P}_0(\tilde{\beta})/$

$$\hat{A}_6 = \hat{P}_+ \hat{\Lambda}_+^{-1/2} \hat{P}'_+ n^{1/2} \hat{g}(\beta_0) + \hat{P}_0 \hat{\Lambda}_0^{-1/2} \hat{P}'_0 n^{1/2} \hat{g}(\beta_0) \quad (71)$$

$$= \hat{P}_+ \hat{\Lambda}_+^{-1/2} \hat{P}'_+ n^{1/2} \hat{g}(\beta_0) + \hat{P}_0 (n^\delta \hat{\Lambda}_0)^{-1/2} \hat{P}'_0 n^{1/2} n^{\delta/2} \hat{g}(\beta_0) \quad (72)$$

$$= P_+ \Lambda_+^{-1/2} P'_+ n^{1/2} \hat{g}(\beta_0) + P_0 \Lambda_0^{-1/2} P'_0 n^{1/2} n^{\delta/2} \hat{g}(\beta_0) + o_p(1) \quad (73)$$

Where the LHS of (73) follows since $\hat{P}_+ = P_+ + o_p(1)$, $\hat{\Lambda}_+ = \Lambda_+ + o_p(1)$ by Theorem 4 under A1, A2 and $n^{1/2} \hat{g}(\beta_0) = O_p(1)$ by A1 (i),(ii). The RHS of (73) holds as $\hat{P}'_0 n^{1/2} n^{\delta/2} \hat{g}(\beta_0) = n^{1/2} n^{\delta/2} P'_0 \hat{g}(\beta_0) + o_p(1)$ as shown in (50) where $\hat{P}_0 = P_0 + o_p(1)$ $n^\delta \hat{\Lambda}_0 = \Lambda_0 + o_p(1)$ hold by Theorem 4.

Finally since $P'_0 n^{1/2} n^{\delta/2} \hat{g}(\beta_0) \xrightarrow{d} N(0, I_{\bar{m}})$ as shown in (48) and $P'_+ n^{1/2} \hat{g}(\beta_0) \xrightarrow{d} N(0, I_{m-\bar{m}})$ by the LLLT under A1(i),(ii) where $P'_+ \Omega P_+ = \Lambda_+$ and both terms in (73) are orthogonal then,

$$\hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\tilde{\beta}) \xrightarrow{d} N(0, I_m) \quad (74)$$

hence (70), (74) imply (68) by the application of Slutsky's Theorem.

PROOF OF THEOREM 2

Define $\Phi := G'_{BP} \Lambda^{-1} G_{BP}$ and write $\Phi^{-1} = C'C$ by the Cholesky Decomposition for some $C \in \mathbb{R}^{p \times p}$.

Theorem 2 holds if,

$$\hat{V}(\hat{\beta}^*)^{-1/2} B_n^{-1} \xrightarrow{p} C \quad (75)$$

since $\hat{V}(\hat{\beta}^*)^{-1/2} n^{1/2}(\hat{\beta} - \beta_0) = \hat{V}(\hat{\beta}^*)^{-1/2} B_n^{-1} n^{1/2} B_n(\hat{\beta} - \beta_0)$ where $B_n n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Phi^{-1})$ by Theorem 1.

Define $\hat{C} = \hat{V}(\hat{\beta}^*)^{-1/2} B_n^{-1}$. Then

$$\hat{C}' \hat{C} \xrightarrow{p} C' C \quad (76)$$

holds by (67) which implies $\hat{C} \xrightarrow{p} C$ by similar arguments to (70).

Hence $\hat{V}(\hat{\beta}^*)^{-1/2} n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} CN(0, \Phi^{-1})$ where $C\Phi^{-1}C' = I_{p \times p}$ by definition of C where $CC' = I_{p \times p}$, establishing the result.

PROOF OF THEOREM 3

The proof of Theorem 3 follows the standard method of proof for the distribution of the J-Statistic in the strongly identified (e.g Hansen (1982)).

$$\hat{Q}_{opt}(\hat{\beta}) = n\hat{g}(\hat{\beta})' \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\hat{\beta}) \quad (77)$$

$$= (n^{1/2} \hat{\Omega}(\tilde{\beta})^{-1/2} \hat{g}(\hat{\beta}))' (n^{1/2} \hat{\Omega}(\tilde{\beta})^{-1/2} \hat{g}(\hat{\beta})) \quad (78)$$

By a mean value expansion of $\hat{g}(\hat{\beta})$,

$$\hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\hat{\beta}) = \hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\beta_0) + \hat{\Omega}(\tilde{\beta})^{-1/2} \hat{G}(\bar{\beta}) n^{1/2} (\hat{\beta} - \beta_0) \quad (79)$$

where $\bar{\beta}$ lies between $\hat{\beta}$ and β_0 .

Plugging in the expansion of $\hat{\beta} - \beta_0$ from (57) in to (79),

$$\hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\hat{\beta}) = (I_m - \hat{F}(\hat{F}'\hat{F})^{-1} \hat{F}') n^{1/2} \hat{\Omega}(\tilde{\beta})^{-1/2} \hat{g}(\beta_0) \quad (80)$$

Where $\hat{F} = \hat{\Omega}(\tilde{\beta})^{-1/2} \hat{G}(\tilde{\beta})$.

Then $\hat{F} B_n^{-1} \xrightarrow{p} \Lambda^{-1/2} G_{BP}$ similarly to \hat{A}_5 in (70). Since $\hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\beta_0) = O_p(1)$ as shown in (74) defining $F = \Lambda^{-1/2} G_{BP}$

$$\hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\hat{\beta}) = (I - F(F'F)^{-1}F') \hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\beta_0) + o_p(1) \quad (81)$$

and using the result from (74) $\hat{\Omega}(\tilde{\beta})^{-1/2} n^{1/2} \hat{g}(\hat{\beta}) \xrightarrow{d} N(0, I_m)$ then since $I - F(F'F)^{-1}F'$ is idempotent of rank $m - p$ the result follows.

PROOF OF THEOREM 4

The proof of Theorem 4 borrows heavily from the proof of Theorem 2 of Grant (2014). A2 combines Assumptions 1 and 2 of Grant (2014) used to derive the asymptotic eigen-system expansions with the additional assumption of WSV. The proof for non-zero eigenvalues (20),(21) follows under A2 directly from Theorem 3 in Grant (2014). Grant (2014) considers the case $\Lambda_{0n} = 0$, this paper considers the case $\Lambda_{0n} = \Lambda_0/b_n$ where Λ_0 is full rank. When $b_n \rightarrow \infty$ then $\Lambda_{0n} \rightarrow 0$.

To first prove (22) expand $\hat{\Omega}(\beta_n)$ around Ω_n , then under A2 by Theorem 3 of Grant (2014),

$$\hat{P}_0(\beta_n) = P_0 - \Omega_+^* (\hat{\Omega}(\beta_n) - \Omega_n) P_0 + O_p(\|\hat{\Omega}(\beta_n) - \Omega_n\| \wedge \|\Delta_n\|)^2 \quad (82)$$

where by T,

$$\|\hat{\Omega}(\beta_n) - \Omega_n\| \leq \|\hat{\Omega}(\beta_n) - \hat{\Omega}\| + \|\hat{\Omega} - \Omega_n\| \quad (83)$$

by A2(v) $\|\hat{\Omega}(\beta_n) - \hat{\Omega}\| = O_p(\|\Delta_n\|) = O_p(n^{-1/2})$ and $\|\hat{\Omega} - \Omega_n\| = O_p(n^{-1/2})$

under A1(i)(ii). Since $\|P_0\| = \bar{m} < m = O(1)$ and $\|\Omega_+^*\| \leq \|P_+\|^2 \|\Lambda_+\|^{-1} = O(1)$. Together along with (82) along with repeated applications of CS imply $\hat{P}_0(\beta_n) - P_0 = O_p(n^{-1/2})$.

To prove (23) expand $\hat{\Omega}(\beta_n)$ around Ω then by Theorem 3 of Grant (2013) under A2 (since $P_0'\Omega = 0$),

$$\hat{\Lambda}_0(\beta_n) = P_0'\hat{\Omega}(\beta_n)P_0 - P_0'\hat{\Omega}(\beta_n)\Omega_+^*\hat{\Omega}(\beta_n)P_0 + O_p((\|\hat{\Omega}(\beta_n) - \Omega\| \wedge \|\Delta_n\|)^3) \quad (84)$$

where by T,

$$\|\hat{\Omega}(\beta_n) - \Omega\| \leq \|\hat{\Omega}(\beta_n) - \hat{\Omega}\| + \|\hat{\Omega} - \Omega_n\| + \|\Omega_n - \Omega\| = O_p(n^{-(1/2 \vee \delta)}) \quad (85)$$

which holds since by A2(v) $\|\hat{\Omega}(\beta_n) - \hat{\Omega}\| = O_p(\|\Delta_n\|) = O_p(n^{-1/2})$ and $\|\hat{\Omega} - \Omega_n\| = O_p(n^{-1/2})$. Since $0 < \delta < 1$ by WSV where $\Omega = \lim_{n \rightarrow \infty} \Omega_n = P_+\Lambda_+P_+'$ then $\|\Omega_n - \Omega\| = \|P_0\Lambda_0/b_nP_0'\| = O_p(b_n^{-1}) = O_p(n^{-\delta})$, together imply $\|\hat{\Omega}(\beta_n) - \Omega\| = O_p(n^{-(1/2 \vee \delta)})$. Hence $\|\hat{\Omega}(\beta_n) - \Omega\|^3 = O_p((n^{-3(1/2 \vee \delta)})$. Then $n^\delta O_p((n^{-3(1/2 \vee \delta)}) = o_p(1)$ since $0 < \delta < 1$ establishing (86).

$$n^\delta \hat{\Lambda}_0(\beta_n) = n^\delta P_0'\hat{\Omega}(\beta_n)P_0 - n^\delta P_0'\hat{\Omega}(\beta_n)\Omega_+^*\hat{\Omega}(\beta_n)P_0 + o_p(1) \quad (86)$$

We now establish that,

$$n^\delta P_0'\hat{\Omega}(\beta_n)\Omega_+^*\hat{\Omega}(\beta_n)P_0 = o_p(1) \quad (87)$$

Let $\bar{\beta}_n$ lie between β_0 and β_n and define $\bar{G}_i = G_i(\bar{\beta}_n)$ then using a Taylor expansion of $\hat{\Omega}(\beta_n)$ around β_0 ,

$$\hat{\Omega}(\beta_n) = \hat{\Omega} + \frac{1}{n} \sum_{i=1}^n g_i \Delta_n' \bar{G}_i' + \frac{1}{n} \sum_{i=1}^n \bar{G}_i \Delta_n g_i' + \frac{1}{n} \sum_{i=1}^n \bar{G}_i \Delta_n \Delta_n' \bar{G}_i' \quad (88)$$

$$n^{\delta/2}P_0'\hat{\Omega}(\beta_n) = n^{\delta/2}P_0'(\hat{\Omega} + \frac{1}{n}\sum_{i=1}^n g_i\Delta_n'\bar{G}_i' + \frac{1}{n}\sum_{i=1}^n \bar{G}_i\Delta_n g_i' + \frac{1}{n}\sum_{i=1}^n \bar{G}_i\Delta_n\Delta_n'\bar{G}_i') \quad (89)$$

where $n^{\delta/2}P_0'\hat{\Omega} = n^{\delta/2}P_0'(\hat{\Omega} - \Omega_n) + n^{\delta/2}P_0'\Omega_n$ where $n^{\delta/2}\|\hat{\Omega} - \Omega_n\| = O_p(n^{-(1+\delta)/2})$ under A1(i),(ii), and $n^{\delta/2}\|P_0'\Omega_n\| = n^{\delta/2}\|\Lambda_0/b_n\|P_0\| = O_p(n^{-\delta/2})$ hence $P_0'\hat{\Omega} = O_p(n^{-(1-\delta)/2}) + O_p(n^{-\delta/2}) = O_p(n^{-(1-\delta)/2})$ since $0 < \delta < 1$.

By repeated application of CS,

$$n^{\delta/2}\|P_0'\frac{1}{n}\sum_{i=1}^n g_i\Delta_n'\bar{G}_i'\| \leq n^{\delta/2}\|P_0\|\|\Delta_n\|\frac{1}{n}\sum_{i=1}^n \|c_i\|\frac{1}{n}\sum_{i=1}^n \|\bar{G}_i\| \quad (90)$$

where $c_i := n^{\delta/2}P_0'g_i$ and $\|c_i\|^2 = \text{tr}(n^{\delta}P_0g_i g_i'P_0)$ and

$E[\|c_i\|^2] = \text{tr}(n^{\delta}P_0'E[g_i g_i']P_0) = \text{tr}(\Lambda_0) = O(1)$ by A2(viii). Hence $E[\|c_i\|] = O(1)$ then by the i.i.d assumption and $M\frac{1}{n}\sum_{i=1}^n \sum_{i=1}^n \|c_i\| = O_p(1)$. By T $\frac{1}{n}\sum_{i=1}^n \|\bar{G}_i\| \leq \frac{1}{n}\sum_{i=1}^n \|\bar{G}_i - G_i\| + \frac{1}{n}\sum_{i=1}^n \|G_i\|$ where $\frac{1}{n}\sum_{i=1}^n \|\bar{G}_i - G_i\| = O_p(\|\Delta_n\|)$ by A2(iii) and $\frac{1}{n}\sum_{i=1}^n \|G_i\| = O_p(1)$ by A2(i),(ii). Then finally since $\|\Delta_n\| = O_p(n^{-1/2})$ and $\|P_0\| = O(1)$ by A2(viii) together along with (90) imply,

$$n^{\delta/2}P_0'\frac{1}{n}\sum_{i=1}^n g_i\Delta_n'\bar{G}_i = O_p(n^{-(1-\delta)/2}) = o_p(1) \quad (91)$$

since $\delta < 1$.

Similarly by the analysis above we can show,

$$\|n^{\delta/2}P_0'\frac{1}{n}\sum_{i=1}^n \bar{G}_i\Delta_n g_i'\| = O_p(n^{-(1-\delta)/2}) \quad (92)$$

and also, $n^{\delta/2}\|P_0'\frac{1}{n}\sum_{i=1}^n \bar{G}_i\Delta_n\Delta_n'\bar{G}_i'\| = O_p(\|\Delta_n\|^2 n^{\delta/2}) = O_p(n^{-1+\delta/2})$. So that,

$$\|n^{\delta/2}P_0'\hat{\Omega}(\beta_n)\| = O_p(n^{-(1-\delta)/2}) \quad (93)$$

and

$$\|n^\delta P'_0 \hat{\Omega}(\beta_n) \Omega_+^* \hat{\Omega}(\beta_n) P_0\| \leq \|n^{\delta/2} P'_0 \hat{\Omega}(\beta_n)\|^2 \|\Omega_+^*\| = O_p(n^{-1+\delta}) = o_p(1) \quad (94)$$

Hence (94) plugged in to (86) implies (95).

$$\hat{\Lambda}_0(\beta_n) = P'_0 \hat{\Omega}(\beta_n) P_0 + o_p(1) \quad (95)$$

We now establish that,

$$n^\delta P'_0 \hat{\Omega}(\beta_n) P_0 = \Lambda_0 + o_p(1) \quad (96)$$

which along with (95) establishes (23).

Note that,

$$n^\delta P'_0 \hat{\Omega}(\beta_n) P_0 = n^\delta P'_0 \hat{\Omega} P_0 + n^\delta P'_0 (\hat{\Omega}(\beta_n) - \hat{\Omega}) P_0 \quad (97)$$

where by (89),

$$n^\delta P'_0 (\hat{\Omega}(\beta_n) - \hat{\Omega}) P_0 = n^\delta P'_0 \frac{1}{n} \sum_{i=1}^n g_i \Delta'_n \bar{G}'_i + \frac{1}{n} \sum_{i=1}^n \bar{G}_i \Delta_n g'_i + \frac{1}{n} \sum_{i=1}^n \bar{G}_i \Delta_n \Delta'_n \bar{G}'_i P_0 \quad (98)$$

and repeated application of CS,

$$n^\delta \|P'_0 \frac{1}{n} \sum_{i=1}^n g_i \Delta'_n \bar{G}'_i\| \leq n^{\delta/2} \|P_0\| \frac{1}{n} \sum_{i=1}^n \|c_i\| \frac{1}{n} \sum_{i=1}^n \|\bar{G}_i\| \quad (99)$$

where $\frac{1}{n} \sum_{i=1}^n \|c_i\| = O_p(1)$, $\frac{1}{n} \sum_{i=1}^n \|\bar{G}_i\| = O_p(1)$, $\|\Delta_n\| = O_p(n^{-1/2})$ and $n^\delta \|P'_0 \frac{1}{n} \sum_{i=1}^n g_i \Delta'_n \bar{G}'_i\| = O_p(n^{-(1-\delta)/2})$ and $n^\delta P'_0 \frac{1}{n} \sum_{i=1}^n \bar{G}_i \Delta_n \Delta'_n \bar{G}'_i P_0 = O_p(\|\Delta_n\|^2 n^\delta) = O_p(n^{-(1-\delta)})$ follow by the arguments above in the proof of (93). Hence,

$$n^\delta P'_0 \hat{\Omega}(\beta_n) P_0 = n^\delta P'_0 \hat{\Omega} P_0 + o_p(1) \quad (100)$$

so that finally by definition of c_i $n^\delta P'_0 \hat{\Omega} P_0 = \frac{1}{n} \sum_{i=1}^n c_i c'_i$ where c_i is i.i.d by A2(i), and $E[c_i c'_i] = n^\delta P'_0 \Omega_n P_0 = \Lambda_0$ and $\|\Lambda_0\| = O(1)$ by A2(vii), hence by KWLLN

$$n^\delta P'_0 \hat{\Omega} P_0 = \Lambda_0 + o_p(1) \tag{101}$$

together with (100) establishes,

$$n^\delta \hat{\Lambda}_0(\beta_n) = \Lambda_0 + o_p(1) \tag{102}$$

establishing the result.

Bibliography

- Aguilar, M., 2009. A Latent Factor Model Of Multivariate Conditional Heteroscedasticity. *Journal of Financial Econometrics* 7, 481-503.
- 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.
- Bosq, D., 2000. *Linear Processes in Function Spaces*. Springer, New York.
- Caner, M., 2008. Nearly-singular design in GMM and generalized empirical likelihood estimators. *Journal of Econometrics* 144, 511-523.
- Caner, M. and Yildiz, N., 2012. CUE with many weak instruments and nearly singular design. *Journal of Econometrics* 170, 422-441.
- 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.
- Grant, N.L., 2014. Efficient Inference from Efficient GMM from Moments with Singular Variance. Unpublished working paper.
- Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-54.
- Knight, K., and Fu, W., 2000. Asymptotics for lasso-type estimators. *The Annals of Statistics* 28, 1356-1378.

- Knight, K., 2008. Shrinkage Estimation for Nearly Singular Designs. *Econometric Theory*, 24, 323-337.
- 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.
- 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.
- West, K., Wong, K., and Anatolyev, S., 2009. Instrumental Variables Estimation of Heteroskedastic Linear Models Using All Lags of Instruments. *Econometric Reviews* 28, 441-467.