The University of Manchester

MANCHESTER 1824

Economics
Discussion Paper Series
EDP-1315

Identification Robust Inference with Singular Variance

Nicky Grant

September 2013

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

Identification Robust Inference with Singular Variance

Nicky Grant[†]

[†]Faculty of Economics
University of Manchester

September 3, 2013

Abstract

This paper studies identification robust inference when moments have singular variance at the true parameter θ_0. Existing robust methods assume non-singular moment variance at θ_0 up to a particular known matrix of parameters, Andrews and Cheng (2012). This is shown to restrict the class of identification failure for which current results on robust methods hold. General conditions under which the GAR statistic has a χ^2_m limit distribution are derived utilizing second order asymptotic eigensystem expansions of the sample variance matrix around θ_0. This method prevents the necessity of restrictive assumptions on the form of singular variance. A crucial condition for this result requires that the null space of the moment variance lies within that of the outer product of the expected first order derivative at θ_0. When this condition is violated the GAR statistic is $O_p(n)$, which is termed the 'moment-singularity bias'. Empirically relevant examples of this problem are provided and the bias verified in a simulation.

Keywords: Generalized Anderson Rubin Statistic, Identification Failure, Singular Variance, Non-linear models, Matrix Perturbation Theory.

1 Introduction

Identification robust methods of inference have gained increasing prominence in the econometrics literature in the last decade. Broadly its objective has been to provide asymptotically valid methods of inference on some unknown parameter θ_0 robust to failures of either global or first-order identification. A substantive part of this literature derives confidence sets containing θ_0 with asymptotically correct probability inverting a pre-specified test statistic over a parameter space.

A large part of this literature has focussed on Linear Instrumental Variable (IV) settings with its roots in the work of Anderson and Rubin (1949). A now sizeable literature has developed providing alternative procedures aiming to make as few possible assumptions to justify asymptotically valid inference on θ_0, including but not limited to Kleibergen (2002,2005), Moreira (2003), Chernozhukov & Hansen (2008), Kleibergen & Mavroeidis (2009), Magnusson (2010), Guggenberger et al (2012).

General non-linear moment functions have received relatively little attention in this literature, a notable exception[1] being the GAR statistic of Newey and Windmeijer (2009). Also known as the Continuous Updating Estimator (CUE) statistic, Guggenberger, Ramalho and Smith (2005) and confidence regions based on the GAR statistic defined as 'S-Sets' in Stock and Wright (2000) .

Let w_i $(i = 1, .., n)$ be an independent and identically distributed (i.i.d) data set with a known $m \times 1$ moment function $g(w, \theta)$ satisfying the moment condition $\mathbb{E}[g(w_i, \theta)] = 0$ at the true parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^p$. Define the sample moment function and corresponding variance matrix respectively $\hat{g}(\theta) := \frac{1}{n}\sum_{i=1}^{n} g_i(w_i, \theta)$, $\hat{\Omega}(\theta) := \frac{1}{n}\sum_{i=1}^{n} g_i(w_i, \theta) g_i(w_i, \theta)'$. The GAR statistic is defined

$$\hat{T}_{GAR}(\theta) := n\hat{g}(\theta)'\hat{\Omega}(\theta)^{-1}\hat{g}(\theta)$$

Under a set of assumptions including the asymptotic moment variance $\Omega := \mathbb{E}[g(w_i, \theta_0)g(w_i, \theta_0)']$ is non-singular $\hat{T}_{GAR}(\theta_0)$ converges in distribution to χ^2_m (e.g Stock and Wright (2000)). The majority of the literature on identification robust inference makes no explicit assumption of first order identification. Namely that $G := \mathbb{E}[G_i(\theta_0)]$ is full column rank where $G_i(\theta) := \partial g(w_i, \theta)/\partial\theta'$.

The impetus for this paper stems from the fact that Ω must be singular when G is not full rank for a broad class of non-linear moment functions. This result has mainly gone unmentioned in the identification robust literature. In light of this issue current results in the identification robust literature justify valid inference for a restricted class of identification failure, limited largely to linear models.

[1] The K-Statistic of Kleibergen (2005) also permits general non-linear moment functions, however the proof of asymptotic validity does not adequately account for singular variance in the transformed moment function considered. This issue is beyond the scope of this paper, however the author intends to work on this in future research

An exception is (Cheng (2008), Andrews and Cheng (2012)) who note the link between identification failure and singular variance for a particular form of identification failure in semi-linear regression models. Cheng (2008) derives the limit distribution of the Non-Linear Least Squares (NLS) estimator for such models. Using this result the distribution of the t, Wald and Quasi-Likelihood Ratio (QLR) statistic are evaluated and methods of identification robust inference are proposed based on such statistics. These results are extended in Andrews & Cheng (2012) to general extremum estimation.

Both papers overcome the issue of singularity of Ω arising from identification failure for asymptotic analysis by an assumption that the form of the singularity is known up to a matrix of model parameters. The class of identification failure (and hence singular variance) that satisfy this assumption is shown to be restrictive, being difficult to motivate outside of the particular examples of identification failure studied in both papers.

This paper differs from Andrews and Cheng (2012) in two ways. (i) Conditions which the GAR statistic is asymptotically χ^2_m are provided for general forms of identification failure requiring no assumptions on the form of moment singularity. (ii) To achieve (i) the GAR statistic is expanded around θ_0 via second order asymptotic expansions of the eigensystem of $\hat{\Omega}(\theta)^{-1}$. This method is of interest in its own right and would prove useful extending results for other identification robust statistics and estimators to allow for general forms of identification failure.

Second order asymptotic expansions of the eigenvectors of $\hat{\Omega}(\theta)$ around θ_0 are derived borrowing results from Matrix Perturbation Theory with its roots in Kato (1982). This field has not readily made it in to the mainstream econometric literature- exceptions being Ratsimalahelo (2002) who consider tests of matrix rank, Moon and Weidner (2010) derive expansions of the Quasi Maximum Likelihood profile function for panel data models and Hassani et al (2011) use such expansions for Singular Spectral Analysis.

Utilizing this result general second order eigenvalue expansions of $\hat{\Omega}(\theta)$ around θ_0 are established. Specific expansions under an i.i.d assumption (along with requisite regularity conditions) are then derived. These eigensystem expansions will prove useful when extending the results of this paper to non-i.i.d settings and are new in the identification literature.

In order for the result (i) to hold further conditions on $G_i(\theta)$ and $\hat{\Omega}(\theta)$ at θ_0 are required when considering general forms of identification failure. A key condition requires those $\delta \in \mathbb{R}^m$ such that $\delta'\Omega = 0$ imply $\delta' G = 0$ (i.e the null space of Ω is a subset of that of GG'). For example this rules out singular variance when the strong identification conditions hold in just-identified models. In this case the GAR statistic is shown to be bounded in probability of order n. This issue currently unknown in the literature is termed the 'moment-singularity bias'.

Simulation evidence demonstrates this bias in a Linear IV Simultaneous Equation setup. The small sample approximation of the GAR statistic by a χ^2_m distribution is shown to be poor when the null space of Ω almost does not lie within that of GG' (i.e when $\delta'\Omega \approx 0$ and $\delta' G \neq 0$). In this case the GAR statistic is shown to be oversized even for large sample sizes.

Numerous examples of singular variance for commonly used moment functions are provided including financial econometric models and Non-Linear IV Simultaneous Equations. Many cases where the assumption on the form of the singularity in Andrews and Cheng (2012) is violated are provided.

Section 2 explores the relationship between G and Ω for conditional moment restrictions. Section 3 sets out the asymptotic approach, deriving second order asymptotic expansions of the eigensystem of $\hat{\Omega}(\theta)$ and specific expansions in the case w_i is i.i.d. Section 4 provides conditions under which the GAR statistic is asymptotically locally χ^2_m and explains the 'moment-singularity bias'. An extensive simulation study is also provided demonstrating the main results of this paper. Section 5 presents conclusions and directions for further research. An Appendix collects proofs of the main theorems.

2 Identification and Singular Variance

The link between identification failure and singular variance is not a new idea in the identification robust literature. Andrews and Cheng (2012) provide asymptotic results under the assumption that there exists $B(\theta)$,

$$B(\theta) = \text{diag}(I_{m^* \times m^*}, \iota(\theta) I_{\bar{m} \times \bar{m}}) \tag{1}$$

Where $m^* = \text{Rank}(\Omega)$ and $\bar{m} = m - m^*$, $\iota(\theta) = \|\theta\|$ such that,

$$B(\theta_n)^{-1} \hat{\Omega}(\theta_n) B(\theta_n)^{-1} \xrightarrow{p} \bar{\Omega} \tag{2}$$

For all $\theta_n = \theta_0 + \Delta_n$ where $\|\Delta_n\| > 0$ and $\|\Delta_n\| = o_p(n^{-1/2})$ and $\text{Rank}(\bar{\Omega}) = m$.

They derive asymptotic properties of (functions) of general extremum estimators working with the transformed moment function $B(\theta_n)^{-1}\sqrt{n}\hat{g}(\theta_n)$ where asymptotic singularity of $\sqrt{n}\hat{g}(\theta_n)$ is eradicated. Once the moment function is transformed the limit variance is non-singular and standard asymptotic analysis is feasible. The existence of such a matrix $B(\theta)$ satisfying (2) is restrictive, being difficult to motivate generally outside of piecewise linear models with particular forms of identification failure, see Section 2.2.

Section 2.1 studies the relationship between G and Ω more generally from moment conditions derived from a system of conditional moment restrictions. Conditions under which $\text{Null}(\Omega) \subseteq \text{Null}(GG')$ are derived for general non-linear models with arbitrary forms of identification failure. As demonstrated in Section 4.2 this condition turns out to be crucial for $\hat{T}_{GAR}(\theta_n)$ to be bounded in probability with a χ^2_m limit distribution. Empirically relevant examples are given where this condition does not hold in Section 2.2.2

2.1 Conditional Moments

Consider a $J \times 1$ residual function $\rho(\theta) := \rho(x, \theta)$ where $\rho(\cdot, \cdot) : \mathfrak{X} \times \Theta \longmapsto \mathbb{R}^J$, $x \in \mathfrak{X} \subseteq \mathbb{R}^l$ with a $h \times 1$ instrument z satisfying,

$$\mathbb{E}[\rho(\theta)|z] = 0 \qquad \text{at} \quad \theta = \theta_0 \tag{3}$$

Broadly speaking there are two types of moment function derived from (3) depending upon whether $\mathbb{E}[\partial\rho(\theta)/\partial\theta'|z]$ must be estimated beforehand.

Namely whether (i) $\mathbb{E}[\partial\rho(\theta)/\partial\theta'|z] = \partial\rho(\theta)/\partial\theta'$ $a.s(z)$ for example Non-Linear Least Squares (NLS) and Unconditional Maximum Likelihood where $x = z$ or (ii) $\mathbb{E}[\partial\rho(\theta)/\partial\theta'|z] \neq \partial\rho(\theta)/\partial\theta'$ for z with measure greater than zero, for example non-linear instrumental variables where generally $z \neq x$.

2.1.1 Case (i): $\mathbb{E}[\partial\rho(\theta)/\partial\theta'|z] = \partial\rho(\theta)/\partial\theta'$

Define $D(\theta, z) := \mathbb{E}[\partial\rho(\theta)/\partial\theta'|z]'$ and $\Omega_\rho(\theta, z) = \mathbb{E}[\rho(\theta)\rho(\theta)'|z]$. In the i.i.d setting the optimal instrument is $D(\theta_0, z)\Omega_\rho(\theta_0, z)^{-1}$, Newey (1993).

Take the case $J = 1$ forming the moment $g(\theta) = D(\theta, z)\rho(\theta)$,

$$\Omega = \mathbb{E}[\rho(\theta_0)^2 D(\theta_0, z)D(\theta_0, z)']$$

$$G = \mathbb{E}[D(\theta_0, z)D(\theta_0, z)']$$

Hence for any $\delta \in \mathbb{R}^p$ such that $\delta'\Omega\delta = 0$ implies

$$\mathbb{E}[\rho(\theta_0)^2(\delta' D(\theta_0, z))^2] = 0$$

$\delta' D(\theta_0, z) = 0$ a.s(z). Therefore $\delta' G\delta = \mathbb{E}[(\delta' D(\theta_0, z))^2] = 0$. The reverse is also simple to establish, so that $\text{Null}(\Omega) \equiv \text{Null}(G) \equiv \text{Null}(GG')$. First order under-identification and singular variance are equivalent for single equation NLS. This result may break down for $J \geq 2$ if $\Omega_\rho(\theta_0, z)$ is singular $a.s(z)$ existing cases where the null space of GG' and Ω are not equivalent[2].

PROPOSITION 1: *For* $g(\theta) = D(\theta, z)\rho(\theta)$
$\text{Null}(\Omega) \subseteq \text{Null}(GG')$ iff $\nexists\ \delta \in \mathbb{R}^p$ such that $D(\theta_0, z)'\delta \in \text{Null}(\Omega_\rho(\theta_0), z)/0\ a.s(z)$
PROOF For $\delta \neq 0$, $\delta'\Omega\delta = \mathbb{E}[\delta' D(\theta_0, z)\Omega_\rho(\theta_0, z)D(\theta_0, z)'\delta] = 0$ iff $\exists \delta \in \mathbb{R}^p$ such that $D(\theta_0, z)'\delta$ lies in the null space of $\Omega_\rho(\theta_0, z)$ $a.s(z)$ since $\delta' G = 0$ iff $\delta' D(\theta_0, z) = 0$ a.s(z).

Q.E.D

2.1.2 Case (ii) $\mathbb{E}[\partial\rho(\theta)/\partial\theta'|z] \neq \partial\rho(\theta)/\partial\theta'$

Commonly when $D(\theta_0, z)$ is not known a priori the fact that (3) implies the following moment condition for any $m \times 1$ $Z := (\phi_1(z), .., \phi_m(z))'$ where $\{\phi_j(.) : j = \{1, .., m\}\}$ are arbitrary functions of z (e.g polynomials in z up to order m),

$$\mathbb{E}[\rho(\theta) \otimes Z] = 0 \quad \text{at} \quad \theta = \theta_0$$

For example the Consumption Capital Asset Pricing Model moment conditions in Stock and Wright (2000). In this case

$$G = \mathbb{E}[D(\theta_0, z)' \otimes Z]$$

$$\Omega = \mathbb{E}[\Omega_\rho(\theta_0, z) \otimes ZZ']$$

Where G is an $mJ \times p$ matrix and Ω is $mJ \times mJ$. In this case in general the null space of Ω and GG' are not necessarily linked. Given that Z includes no linearly redundant combinations of instruments then Ω may be less than full rank only when $\Omega_\rho(\theta_0, z)$ is not full rank $a.s(z)$. Define $\delta := (\delta_1', .., \delta_J')'$ where $\delta_j \in \mathbb{R}^m$ for $j = \{1, .., m\}$.

PROPOSITION 2: $\text{Null}(\Omega) \subseteq \text{Null}(GG')$ iff $\nexists\ \delta \in \mathbb{R}^{mJ}$ where$(\delta_1' Z, .., \delta_J' Z)' \in \text{Null}(\Omega_\rho(\theta_0, z))$ $a.s(z)$ such that $\delta \propto \nu$ for some $\nu \in \mathbb{R}^{mJ}$ s.t $\nu' G \neq 0$

PROOF: For $\delta \neq 0$ then $\delta'\Omega\delta = \mathbb{E}[(\delta_1' Z, .., \delta_J' Z)\Omega_\rho(\theta_0, z)(\delta_1' Z, .., \delta_J' Z)']$ hence $\delta'\Omega = 0$ iff $(\delta_1' Z, .., \delta_J' Z)' \in \text{Null}(\Omega_\rho(\theta_0, z))$ $a.s(z)$. The null space of Ω will not lie in that of GG' iff

[2]Note a similar result can also be shown based utilizing an estimate of the optimal instrument based on an a consistent estimator of a generalized inverse $\Omega_\rho(\theta_0, z)^-$ noting that the $\text{Rank}(\Omega_\rho(\theta_0, z)) = \text{Rank}(\Omega_\rho(\theta_0, z)^-)$.

$\exists\delta \in \mathbb{R}^{mJ}$ such that $(\delta_1' Z, .., \delta_J' Z) \in \text{Null}(\Omega_\rho(\theta_0, z))$ where $\delta \propto \nu$ for some $\nu \in \mathbb{R}^{mJ}$ where $\nu' G \neq 0$.

$Q.E.D$

REMARKS:

(i) When $\Omega_\rho(\theta_0, z)$ is homoscedastic (i.e $\Omega_\rho(\theta_0, z) = \Omega_\rho\ a.s(z)$ for some p.s.d symmetric $m \times m$ matrix Ω_ρ) then it is straightforward to show that $\text{Rank}(\Omega) = m(J - r)$ where $r = \text{Rank}(\Omega_\rho)$.

(ii)If for any function $a(.)$ of z $\exists\ \pi \in \mathbb{R}^m$ such that

$$\mathbb{E}[(\pi' Z - a(z))^2] \to 0 \tag{4}$$

For $m \to \infty$ then $\text{Rank}(\Omega) \leq mJ - r^*$ (as $m \to \infty$) where $r^* = J - \text{Rank}(\Omega_\rho(\theta_0, z))$ $a.s(z)$. Since by (4) there will exist at least r^* linearly independent vectors $\delta \in \mathbb{R}^{mJ}$ s.t $(\delta_1' Z, .., \delta_J' Z)'$ can be expressed as some linear combination of elements of the null space of $\Omega(\theta_0, z)$ $a.s(z)$ for m large.

Especially a concern is (ii) as even if $\rho(\theta_0)$ has no perfectly correlated (linear combination of) elements ($\mathbb{E}[\Omega_\rho(z, \theta_0)]$ is full rank), Ω will be singular for m large when there exists perfect conditional correlation in elements of $\rho(\theta_0)$ (i.e $r^* > 0$). This would violate the condition for GAR to be asymptotically χ^2_m. An example of this case is provided Example 3 in Section 2.2 with a corresponding simulation provided in Section 4.2.

2.2 Examples of Singular Variance

This sections provides examples of moment functions with singular variance both with and without identification- specifically when the condition that $\text{Null}(\Omega) \subseteq \text{Null}(GG')$ holds or does not.

2.2.1 Singular Variance : Null$(\Omega) \subseteq$ Null(GG')

A class of identification failure satisfying $\text{Null}(\Omega) \subseteq \text{Null}(GG')$ is the stochastic semi-linear parametric equations (for $J = 1$) considered in Cheng (2008)[3].

$$y = \alpha' x + \pi f(z, \gamma) + \epsilon$$

[3] Cheng (2008) allow for a vector of non-linear functions though for simplicity this special case is highlight to demonstrate the infeasibility of the assumption on the form of the singular variance made in both Cheng (2008) and Andrews and Cheng (2012).

Where $\theta = (\alpha, \gamma, \pi)$, $\alpha \in \mathbb{R}^q$, $\pi \in \mathbb{R}$, $\gamma \in \mathbb{R}^l$ and $f(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^l \to \mathbb{R}$ is a continuously differentiable function.

Let $w = (y, x, z)$ where y is a scalar random variable, x is $q \times 1$ and z is $d \times 1$ where $\mathbb{E}[\epsilon|x, z] = 0$ at $\theta = \theta_0$ for some parameter vector $\theta_0 = (\alpha_0, \gamma_0, \pi_0)$.

Define $f(\gamma) := f(z, \gamma)$, $\epsilon(\theta) := y - \alpha' x - \pi f(\gamma)$,

$$\frac{\partial \epsilon(\theta)}{\partial \theta} = (x, f(\gamma), \pi \partial f(\gamma)/\partial \gamma)'$$

Then the moment function utilized in NLS is

$$g(\theta) = \epsilon(\theta)(x, f(\gamma), \pi \partial f(\gamma)/\partial \gamma)'$$

Under the i.i.d assumption the variance of the moments at any $\theta \in \Theta$ is

$$\Omega(\theta) = \mathbb{E}\epsilon(\theta)^2 \begin{bmatrix} xx' & f(\gamma)x & \pi x \partial f(\gamma)/\partial \gamma' \\ f(\gamma)x' & \pi f(\gamma)^2 & \pi f(\gamma) \partial f(\gamma)/\partial \gamma' \\ \pi \partial f(\gamma)/\partial \gamma x' & \pi f(\gamma) \partial f(\gamma)/\partial \gamma & \pi^2 \partial f(\gamma)/\partial \gamma \partial f(\gamma)/\partial \gamma' \end{bmatrix}$$

Ω would be singular in the following three cases (and potentially others),

(i) $\theta_0 = (\alpha, \gamma, 0)$ for any $(\alpha, \gamma) \in \mathbb{R}^{q+l}$.

(ii) $f(\gamma_0) = \delta' x$ for some $\delta \in \mathbb{R}^q$.

(iii) $\delta_1' \partial f(\gamma_0)/\partial \gamma = \delta_2' x$ for some $\delta_1 \in \mathbb{R}^l$ and $\delta_2 \in \mathbb{R}^q$. where $\|\delta_1\| > 0$.

Case (i) falls under the assumption of Andrews and Cheng (2012). Namely for the matrix $B(\theta) = \begin{pmatrix} I_{2\times 2} & 0_2 \\ 0_2 & \pi \end{pmatrix}$ then $B(\theta)^{-1}\Omega(\theta)B(\theta)^{-1}$ is no longer a function of π. In this case singularity cased by $\pi_0 = 0$ is removed. However there exist no matrix of the form $B(\theta)$ that will remove the singularity for cases (ii) and (iii) and more generally for arbitrary forms of singularity that depend upon the Data Generating Process.

Example 1: Heckman Selection Consider a Heckman Selection Regression where $f(z, \gamma) = \phi(z'\gamma)/\Phi(-z'\gamma)$ is the Inverse Mills Ratio and z corresponds to variables which govern sample selection. If $z'\gamma_0 = c$ for some constant c and x includes a constant then singularity arises from (ii). Even if this condition does not hold, as noted by Puhani (2000) and others the Inverse Mills Ratio is approximately linear for a wide range of γ. In this case if x and z contain coinciding variables then NLS would be weakly identified with almost singular variance.

Example 2 provides a case of a general non-linear moment function where $\text{Null}(\Omega) \subseteq \text{Null}(GG')$. Also note that in this case there exists no matrix $B(\theta)$ satisfying (1.2).

Example 2: Interest Rate Dynamics

$$r - r_{-1} = a(b - r_{-1}) + \epsilon\sigma r^{\gamma}$$

Where r_{-1} is the first lag of the interest rate r. Define $\theta = (a, b, \sigma, \gamma)$. Under the assumption that ϵ is stationary at $\theta = \theta_0$ where $\theta_0 = (a_0, b_0, \sigma_0, \gamma_0)$ then using the test-function approach of Hansen and Scheinkman (1995) the following moment function is derived in Jagannathan and Wang (2002),

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

satisfying $\mathbb{E}[g(\theta)] = 0$ at $\theta = \theta_0$.

When $\sigma_0 = a_0$, $\gamma_0 = 1/2(a_0 + 1)$ or $\gamma_0 = 1/2(\sigma_0 + 1)$ redundant moments exist at the true parameter. For example if all three conditions held simultaneously the rank of Ω and G would both be 1.

2.2.2 Singular Variance: Null(Ω) $\not\subseteq$ Null(GG')

Common causes of singular variance arise from a lack of identification. It is however plausible that singular variance occurs where $\text{Null}(\Omega) \not\subseteq \text{Null}(GG')$, for example in just-identified settings when G is full rank (first-order identified) though Ω is singular.

Example 3: IV Simultaneous Equations

Consider an example of a conditional moment restriction where $J = 2$,

$$\rho_1(\theta_0) = \epsilon\sqrt{h_1(z)}$$

$$\rho_2(\theta_0) = \epsilon\sqrt{h_2(z)}$$

Where $\mathbb{E}[\epsilon^2|z] = 1$ and $h_1(z)$ and $h_2(z)$ are the conditional heteroscedasticity for equations 1 and 2 respectively. Let Z be an $m \times 1$ vector function of z used as instruments.

Let $\delta = (\delta_1', \delta_2')'$ where $\delta_1, \delta_2 \in \mathbb{R}^m$ then

$$\delta'\Omega\delta = \mathbb{E}[h_1(z)(\delta_1' Z)^2] + \mathbb{E}[h_2(z)(\delta_2' Z)^2] + 2\mathbb{E}[\sqrt{h_1(z)h_2(z)}\delta_1' Z \delta_2' Z]$$

For example if $\delta_1' Z = 1/\sqrt{h_1(z)}$, $\delta_2' Z = -1/\sqrt{h_2(z)}$ then Ω is singular. In the case where $h_1(z) = h_2(z)$ then any $\delta_1, \delta_2 \in \mathbb{R}^m$ where $\delta_1' Z = -\delta_2' Z$ would yield $\delta'\Omega\delta = 0$. This is an example of Proposition 2 and in general $\delta'\Omega = 0$ does not imply $\delta' G = 0$. Take for example

$$\rho_1(\theta) = y_1 - \theta_1 x_1$$

$$\rho_2(\theta) = y_2 - \theta_2 x_2$$

Where $\theta = (\theta_1, \theta_2)$, $x = (y_1, y_2, x_1, x_2)$ with instrument vector $Z = (1, z)$. Assuming $\mathbb{E}[x_1|z] = \bar{\pi}(1+z)$, $\mathbb{E}[x_2|z] = -\bar{\pi}(1+z^2)$ and $z \sim N(0,1)$ it is straightforward to establish,

$$G = \begin{pmatrix} \bar{\pi}(1,1)' & 0_2 \\ 0_2' & \bar{\pi}(-2,0)' \end{pmatrix}$$

If $h_1(z) = h_2(z)$ then $\delta_1 = (c, 0)$, $\delta_2 = (-c, 0)$ for $c \neq 0$ imply $\delta'\Omega = 0$ however $\delta' G = (c\bar{\pi}, 2c\bar{\pi}) \neq 0$ when $\bar{\pi} \neq 0$. Note that if instruments were irrelevant ($\bar{\pi} = 0$) then $\delta' G = 0$ for all directions $\delta \in \mathbb{R}^4$.

Though the example here is somewhat pathological (requiring $\rho_1(\theta_0), \rho_2(\theta_0)$ be perfectly correlated) the problem extends also to the case where no equations are perfectly correlated, i.e $h_1(z) \neq h_2(z)$).

For example if $h_1(z) = \exp(-\zeta_1 z)$ and $h_2(z) = \exp(-\zeta_2 z)$ (where $\zeta_1 \neq \zeta_2$) if Z includes polynomial orders of z up to m then δ_1 and δ_2 such that $\delta_1' z = 1 + 1/2\zeta_1 z + ... + (1/2\zeta_1 z)^m/m!$ and $\delta_2' z = -(1 + 1/2\zeta_2 z + ... + (1/2\zeta_2 z)^m/m!)$ will well approximate $1/\sqrt{h_1(z)}$ and $-1/\sqrt{h_2(z)}$ respectively for m large. When using many instruments (and/or with J large) it is entirely plausible there exist directions in which $\delta'\Omega = 0$ that do not imply $\delta' G$=0.

3 Matrix Perturbation Theory

Section 1.4 derives conditions under which $\hat{T}_{GAR}(\theta_n)$ converges in distribution to a χ^2_m limit for any local sequence $\theta_n = \theta_0 + \Delta_n$ where $\|\Delta_n\| = o_p(n^{-1/2})$ without an assumption the form of the singularity is known. To do so the GAR statistic at θ_n is expanded around the point of singularity θ_0, requiring second order expansions of the eigensystem of $\hat{\Omega}(\theta_n)$ around θ_0. This section is concerned with deriving these expansions.

Firstly definitions for the eigensystem of the functional matrix $\Omega(\theta)$ and $\hat{\Omega}(\theta)$ are outlined. By construction both matrices are p.s.d and symmetric hence the following decompositions can be made for all $\theta \in \Theta$. Let the $m \times m$ matrix $P(\theta)$ be the matrix of population eigenvalues where $\Omega(\theta) = P(\theta)\Lambda(\theta)P(\theta)'$ Such that $P(\theta)'P(\theta) = I_m$ and $\Lambda(\theta)$ contains the eigenvalues of Ω across the diagonal and zeros on the off-diagonal. Define the rank of $\Omega(\theta)$ as $m - \bar{m}(\theta)$ where $0 \leq m(\theta) \leq m$. Express $P(\theta) = (P_+(\theta), P_0(\theta))$ and $\Lambda(\theta) = \begin{pmatrix} \Lambda_+(\theta) & 0 \\ 0 & \Lambda_0(\theta) \end{pmatrix}$ where $\Lambda_+(\theta)$ is an $(m - \bar{m}(\theta)) \times (m - \bar{m}(\theta))$ diagonal matrix with the non-zero eigenvalues of $\Omega(\theta)$ on the diagonal with corresponding eigenvector matrix $P_+(\theta)$. $\Lambda_0(\theta) = 0_{\bar{m}(\theta)\times\bar{m}(\theta)}$ with corresponding eigenvector matrix $P_0(\theta)$. Performing an eigenvalue decomposition re-write $\Omega(\theta)$ as

$$\Omega(\theta) = P_+(\theta)\Lambda_+(\theta)P_+(\theta)' + P_0(\theta)\Lambda_0(\theta)P_0(\theta)'$$

Performing a similar decomposition for $\hat{\Omega}(\theta)$

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

Where $\hat{P}_+(\theta)$ is an $(m - \bar{m}(\theta)) \times (m - \bar{m}(\theta))$ matrix of sample eigenvector estimates of $P_+(\theta)$ with corresponding sample eigenvalue $\hat{\Lambda}_+(\theta)$. $\hat{P}_0(\theta)$ and $\hat{\Lambda}_0(\theta)$ are similarly the sample estimates of $P_0(\theta)$ and $\Lambda_0(\theta)$ respectively letting $\hat{P}(\theta) := (\hat{P}_+(\theta), \hat{P}_0(\theta))$.
Define $\Omega = \Omega(\theta_0)$ and $\hat{\Omega} = \hat{\Omega}(\theta_0)$ and $\bar{m}(\theta_0) := \bar{m}$ for notational simplicity throughout and let the eigenvalues/vector matrices of both Ω and $\hat{\Omega}$ be defined without θ_0, for example $P := P(\theta_0)$, $\hat{P} := \hat{P}(\theta_0)$ and so on.

3.1 Asymptotic Eigensystem Expansions

Borrowing results from the Matrix Perturbation literature second order expansions of the eigenvectors of $\hat{\Omega}(\theta_n)$ are derived, Hassani et al. (2011). Using this result second order expansions of the eigenvalues around θ_0 are established. These results for the sample moment variance matrix are new in the literature and of interest in their own right.

Assumption 1 (A1): *General Eigensystem Expansions*
(i) $c \leq [\Lambda_+]_{jj} \leq K$ for some $0 < c \leq K < \infty$ $\forall$ $j = \{1, .., \bar{m}\}$, *(ii)* $\|\hat{\Omega}(\theta) - \hat{\Omega}(\theta^*)\| \leq \hat{M}\|\theta - \theta^*\|$ $\forall \theta, \theta^* \in \Theta$ *for some* $\hat{M} = O_p(1)$, *(iii)* $m < \infty$

A1(i) is a relatively trivial condition which assumes the non-zero eigenvalues are well separated from zero and bounded. A2(ii) requires an asymptotic Lipschitz condition on

the sample variance matrix. A3(iii) is an assumption of a finite number of moments which is made for simplicity, all results could readily be extended to allow $m \to \infty$ with appropriate rate restrictions relative to n.
Define $\Omega_+ = P_+\Lambda_+P'_+$ and $\Omega^*_+ = P_+\Lambda_+^{-1}P'_+$

Theorem 1 (T1): *General Eigensystem expansions of*
Under A1

$$\hat{P}_+(\theta_n) = P_+ + O_p(\|\hat{\Omega} - \Omega\| \wedge \|\Delta_n\|) \tag{5}$$

$$\hat{\Lambda}_+(\theta_n) = \Lambda_+ + O_p(\|\hat{\Omega} - \Omega\| \wedge \|\Delta_n\|) \tag{6}$$

$$\hat{P}_0(\theta_n) = P_0 - \Omega^*_+\hat{\Omega}(\theta_n)P_0 + O_p((\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|)^2) \tag{7}$$

$$\hat{\Lambda}_0(\theta_n) = P'_0\hat{\Omega}(\theta_n)P_0 - P'_0\hat{\Omega}(\theta_n)\Omega^*_+\hat{\Omega}(\theta_n)P_0 + O_p((\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|)^3) \tag{8}$$

Second order expansions for the eigenvectors/values corresponding to non-zero eigenvalues are also provided in Lemma A2. As shown in Section 1.4 second order terms in $\hat{\Lambda}_+(\theta_n)$,$\hat{P}_+(\theta_n)$ do not enter first order asymptotics for $\hat{T}_{GAR}(\theta_n)$ these results are omitted here for brevity. Theorem 2 provides expansions of the eigensystem of $\hat{\Omega}(\theta_n)$ around θ_0 under an i.i.d assumption on w_i with corresponding regularity conditions.

Assumption 2 (A2) : *i.i.d Eigensystem Expansions*
(i) $w_i(i = 1, .., n)$ *is an i.i.d sequence, (ii)* $\mathbb{E}[\|g_i\|^2] < \infty$,*(iii)* $\frac{1}{n}\sum_{i=1}^n \|G_i(\theta) - G_i(\theta^*)\| \leq \hat{M}\|\theta - \theta^*\|\ \forall\theta, \theta^* \in \Theta$ *where* $\hat{M} = O_p(1)$, *(iv)* $\mathbb{E}[\|G_i\|^2] < \infty$

A2(i) is made largely for simplicity, all results could be extended to allow for dependence and heteroscedasticity under further regularity conditions. A2(iii) requires that for n large enough the average of any elements of $G_i(\theta)$ is sufficiently continuous. This is a weaker condition than $G_i(\theta)$ is continuous, though a sufficient condition for A2(ii) is that $G_i(\cdot)$ satisfies the Lipschitz condition. A2 (ii), (iv) are both required such that the remainder terms in the eigensystem expansions are bounded.
For any arbitrary sequence Δ_n where $\|\Delta_n\| > 0$, $\|\Delta_n\| = o_p(n^{-1/2})$ define $\bar{\Delta}_n = \|\Delta_n\|^{-1}\Delta_n$ where $\bar{\Delta}_n \xrightarrow{p} \Delta$ where $\|\Delta\| > 0$ and is bounded. Define $g_i := g_i(\theta_0)$, $G_i := G_i(\theta_0)$ and the following[4] $\Gamma = P'_0\mathbb{E}[G_i\Delta\Delta'G'_i]P_0$,$\Psi = P'_0\mathbb{E}[G_i\Delta g'_i]$, $\Phi := \Gamma - \Psi\Omega^*_+\Psi'$.

[4]For simplicity w omit dependence of Γ, Ψ on the arbitrary limit Δ.

Theorem 2 (T2): *i.i.d Eigensystem Expansions*
Under A1, A2

$$\hat{P}_+(\theta_n) \xrightarrow{p} P_+ \tag{9}$$

$$\hat{\Lambda}_+(\theta_n) \xrightarrow{p} \Lambda_+ \tag{10}$$

$$\|\Delta_n\|^{-1}(\hat{P}_0(\theta_n) - P_0) \xrightarrow{p} \Omega_+^* \Psi' \tag{11}$$

$$\|\Delta_n\|^{-2}\hat{\Lambda}_0(\theta_n) \xrightarrow{p} \Phi \tag{12}$$

4 Generalized Anderson Rubin Statistic with Singular Variance

This section derives conditions under which the GAR statistic has a χ^2_m limit distribution making no assumption on the form of singularity.
The GAR statistic $\hat{T}_{GAR}(\theta)$ does not exist at $\theta = \theta_0$ when Ω is singular. However if $\text{Rank}(\Omega(\theta)) = m$ for all $\theta \in \mathbb{B}(\theta_0, \epsilon)/\theta_0$ for some $\epsilon > 0$ then $T_{GAR}(\theta_n)$ will exist w.p.1. The formal condition under which $\hat{T}_{GAR}(\theta_n)$ exists asymptotically is given in Assumption 3(ii). As such the limit of the GAR statistic is derived at a point arbitrarily close to θ_0[5].

Assumption 3 (A3) : *Limit Distribution of GAR Statistic*
(i) $\text{Null}(\Omega) \subseteq \text{Null}(GG')$, *(ii)* Φ *is s.p.d.*

A3(i) is a crucial condition needed for the GAR statistic to have the standard χ^2_m limit distribution. Note that this assumption always holds for NLS where $J = 1$ by the results in Section 1.2.1. When A3(i) is violated the GAR statistic in general is $O_p(n)$ as shown in Theorem 4. This is termed the 'moment-singularity bias' in Section 1.4.2.
A3(ii) is required for $\hat{T}_{GAR}(\theta_n)$ to exist w.p.1 when the function does not exist at θ_0 due to a singularity in Ω. $\Phi = P_0'(\mathbb{E}[G_i\Delta\Delta'G_i'] - \mathbb{E}[G_i\Delta g_i']\Omega_+^*\mathbb{E}[g_i\Delta'G_i'])P_0$ is p.s.d and in general will be p.d unless $P_0'G_i = 0$ which would arise in the case where $P_0'g_i(\theta) = 0$ for all $\theta \in \mathbb{B}(\theta_0, \epsilon)$ where $\epsilon > 0$. In this case $P_0'G_i(\theta) = 0$ which implies $P_0'G_i = 0$. Singularity

[5]Note that this is not an assumption that the true parameter is a sequence converging to θ_0 at some rate, merely that we are evaluating the distribution of $T_{GAR}(\theta)$ at points arbitrarily close to θ_0. Using these results the true parameter could be modeled as some sequence converging to a limit θ_0 which is commonly used to model certain forms of weak-identification in the literature, for example Stock and Wright (2000), Andrews and Cheng (2012).

from identification failure invariably occurs at discrete points in the parameter space. It is unusual to find a case of identification failure where some linear combination $x \in \mathbb{R}^m$ such that $x'g(\theta) = 0$ for all $\theta \in \mathbb{B}(\theta_0, \epsilon)$ where $\epsilon > 0$. Example 1,2 and 3 all have singular variance occurring at a point θ_0 where the variance is non-singular at some perturbation away from θ_0.

Theorem 3 (T3): *Under A1, A2,A3*

$$\hat{T}_{GAR}(\theta_n) \xrightarrow{d} \chi^2_m \tag{13}$$

Remarks (i) Note that in the standard case where Ω is assumed to be non-singular, A2 (iii), (iv) and A3 are not made. In this case all that is required to establish (13) is $\sqrt{n}\hat{g}(\theta_n) \xrightarrow{d} N(0, \Omega)$ which holds under A2(i),(ii) and that $\hat{\Omega}(\theta)$ is (asymptotically) continuous around θ_0 which follows from A1(ii). It is then straightforward to show that $\hat{T}_{GAR}(\theta_n) \xrightarrow{d} \chi^2_m$.

(ii) When Ω is singular second order terms in the eigensystem expansions of $\hat{\Omega}(\theta_n)^{-1}$ enter first order asymptotics. As such second order terms in $\sqrt{n}\hat{g}(\theta_n)$ impact first order asymptotics, requiring further regularity conditions on the first order derivative. These conditions are currently unknown in the literature.

Theorem 3 is confirmed in a simulation based on the Heckman Selection example in Section 1.2.2. In this case the crucial assumption that $\text{Null}(\Omega) \subseteq \text{Null}(GG')$ holds as this is NLS with $J = 1$.

4.1 Simulation : Heckman Selection

Consider the setup in Example 1 where

$$y = \theta_1 + \theta_2 x + \theta_3 \frac{\phi(\theta_4 + \theta_5 x)}{\Phi(-(\theta_4 + \theta_5 x))} + \epsilon$$

Where (x, e) are i.i.d and $x \sim N(0, 1)$ and $\epsilon|x \sim N(0, 1)$.

Setting $\theta_0 = (1, 1, 0.2, 0.1, \kappa)$ for $\kappa = \{0.05, 0.5, 1\}$, $N = \{100, 500, 1000, 5000, 50000\}$

For κ close to zero NLS is poorly identified as the Inverse Mills Ratio is approximately linear for arguments less than 2, Puhani (2000). Rejection probabilities for the event that the GAR function at $\theta_0 + 1/n$ is less than the 90% quantile of a χ^2_5 based on $R = 10000$ simulations are calculated. Evidence in Table 1.1 shows that for n large the GAR has correct coverage based on a χ^2_5 approximation. Hence the GAR statistic provides correct coverage asymptotically for general non-linear forms of identification failure under A1-A3 as shown in Theorem 3.

Table 1: GAR Rejection Probabilities: Heckman Selection

	$\kappa = 0.05$	$\kappa = 0.5$	$\kappa = 1$
$n = 100$	0.101	0.079	0.076
$n = 500$	0.087	0.085	0.09
$n = 1000$	0.0892	0.088	0.095
$n = 5000$	0.093	0.094	0.096
$n = 50000$	0.103	0.102	0.109

4.2 Moment-Singularity Bias when Null(Ω) $\not\subseteq$ Null(GG')

A3(i) is critical in the proof of Theorem 3. When this condition is violated- with examples given in Section 1.2.2 in general $\hat{T}_{GAR}(\theta_n)$ is unbounded in probability.

THEOREM 4 (T4) : *Under A1, A2, A3(ii) when A3(i) is violated*

$$\hat{T}_{GAR}(\theta_n)/n \xrightarrow{p} \Delta' G' P_0 \Phi^{-1} P_0' G \Delta \tag{14}$$

Where $\Delta' G' P_0 \Phi^{-1} P_0' G \Delta > 0$ since Φ is full rank by A3(ii). Hence the GAR statistic is $O_p(n)$ when A3(ii) is violated. When A3(i) is almost violated the GAR statistic is shown in the simulation below to be potentially very oversized even for large sample sizes.
Theorem 4 is particularly striking as it implies there exist cases of correctly specified moments which strongly identify θ_0 where identification robust inference based on the GAR statistic would (asymptotically) yield the empty set. This would usually regarded as a sign of moment misspecification.

4.2.1 Simulation : Linear IV Simultaneous Equations

Consider Example 3 where

$$y_1 = x_1 + \epsilon_1$$

$$y_2 = 0.5x_2 + \epsilon_2$$

$$x_1 = \bar{\pi}(1 + z) + \eta_1$$

$$x_2 = -\bar{\pi}(1+z^2) + \eta_2$$

$$\eta_1 = \upsilon_1 \exp(-\zeta_1 z), \quad \eta_2 = \upsilon_2 \exp(-\zeta_2 z)$$

$$\upsilon_1 = \sqrt{\frac{1+\rho}{2}}\zeta_1 + \sqrt{\frac{1-\rho}{2}}\zeta_2, \quad \upsilon_2 = \sqrt{\frac{1+\rho}{2}}\zeta_1 - \sqrt{\frac{1-\rho}{2}}\zeta_2$$

$$(\zeta_1, \zeta_2, \epsilon_1, \epsilon_2)'|z \overset{i.i.d}{\sim} N(0_4, \Xi) \quad \Xi = \begin{pmatrix} 1 & 0 & 0.3 & 0 \\ 0 & 1 & 0.5 & 0 \\ 0.3 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 1 \end{pmatrix}$$

For each $\bar{\pi} = \{0, 0.1, 0.5\}$ (uncorrelated, weak, strong) instruments the following simulation is performed. For instrument sets $I_1 = \{1, z\}$, $I_2 = \{1, z, z^2\}$, $I_3 = \{1, z, z^2, z^3\}$ which respectively yield $m = \{4, 6, 8\}$ moments rejection probabilities are formulated for the GAR statistic based on a the 0.9 quantile of the relevant χ^2_m based on 5000 repetitions where $\theta_n = (1, 1/2) + 1/n$ for $z \overset{i.i.d}{\sim} N(0,1)$ $n = \{100, 500, 1000, 5000, 50000\}$, $\rho = \{0.9995, 0.999995, 1\}$ $(\zeta_1, \zeta_2) = \{(0,0), (0, 0.5), (0, 1)\}$.

When $\bar{\pi} = 0$ the condition $\text{Null}(\Omega) \subseteq \text{Null}(GG')$ is automatically satisfied, in which case the GAR statistic should have a rejection probability around 0.1 for large sample sizes and is verified in Table 1.2. For brevity only the case $\zeta_1 = \zeta_2 = 0$ is reported, similar results were found for both other cases.

When $\bar{\pi} \neq 0$ then when Ω is singular in directions G does not vanish the GAR statistic is in general oversized

(i) When $\rho = 1$ and $\zeta_1 = \zeta_2 = 0$ then Ω is singular as shown in Example 3 $\delta'\Omega = 0$ implies $\delta' G = 0$ if and only if $\bar{\pi} = 0$. The stronger the instruments (the larger is $\bar{\pi}$) the more oversized the rejection probability for any m.

(ii) When $\rho = 1$ and $\zeta_1 \neq \zeta_2$ then Ω approaches a singular matrix as m increases. Fixing $\zeta_1 = 0$ and let ζ_2 equal 0.5 and 1. The larger is ζ_2 the less well that any m polynomials of z can approximate $\exp(\zeta_2/2z)$ (i.e $h_2(z)^{-1/2}$ from notation in Example 3). The GAR rejection probability is decreasing in ζ_2 for any given m, $\bar{\pi}$ and increasing in both m and $\bar{\pi}$.

(iii) When $\rho < 1$ then Ω is full rank, however the closer ρ is to 1 in general the larger the GAR statistic as $\bar{\pi}$ increases. Even for large sample sizes the rejection probabilities can be very close to 1.

Table 1.3 shows the rejection probabilities for the weak instrument case. As expected when $\rho = 1$ and $\zeta_1 = \zeta_2 = 0$ the rejection probabilities converge to 1 as n increases (since GAR is unbounded in this case for any m). For $\rho = 0.999995$ and 0.9995 the rejection

Table 2: GAR Rejection Probabilities $\bar{\pi} = 0$

		$\rho = 0.9995$			$\rho = 0.99995$			$\rho = 1$		
		$m = 4$	$m = 6$	$m = 8$	$m = 4$	$m = 6$	$m = 8$	$m = 4$	$m = 6$	$m = 8$
$\zeta_1 = \zeta_2 = 0$	$n = 100$	0.099	0.080	0.074	0.090	0.092	0.077	0.099	0.904	0.074
	$n = 500$	0.099	0.099	0.097	0.095	0.093	0.087	0.101	0.094	0.084
	$n = 1000$	0.010	0.102	0.0891	0.097	0.103	0.096	0.098	0.094	0.09
	$n = 5000$	0.098	0.093	0.103	0.093	0.106	0.104	0.101	0.097	0.099
	$n = 50000$	0.010	0.102	0.096	0.098	0.101	0.098	0.102	0.091	0.102

probabilities for any n,m are smaller then when $\rho = 1$ however still oversized in small samples.

As ζ_2 increases then in general the rejection probabilities decrease for any ρ as for any given m the instrument set less well approximate the null space of $\Omega(z, \theta_0)$. As m increases the rejection probabilities increase.

This pattern is again observed in Table IV for strong instruments. In this case the rejection probabilities for any given n, m, ρ, ζ_2 is relatively more oversized in general than when $\bar{\pi} = 0.1$. This corresponds to the fact the condition $\text{Null}(\Omega) \subseteq \text{Null}(GG')$ is potentially more strongly violated in this case.

Table 3: GAR Rejection Probabilities $\bar{\pi} = 0.1$

		$\rho = 0.9995$			$\rho = 0.999995$			$\rho = 1$		
		$m = 4$	$m = 6$	$m = 8$	$m = 4$	$m = 6$	$m = 8$	$m = 4$	$m = 6$	$m = 8$
$\zeta_1 = \zeta_2 = 0$	$n = 100$	0.135	0.123	0.198	0.428	0.38	0.8	0.492	0.421	0.867
	$n = 500$	0.11	0.114	0.132	0.727	0.724	0.998	0.995	0.996	1
	$n = 1000$	0.106	0.1	0.12	0.628	0.6412	0.992	1	1	1
	$n = 5000$	0.091	0.103	0.095	0.251	0.253	0.599	1	1	1
	$n = 50000$	0.092	0.1	0.108	0.117	0.11	0.15	1	1	1
$\zeta_1 = 0 \quad \zeta_2 = 0.5$	$n = 100$	0.117	0.118	0.36	0.204	0.292	0.8	0.218	0.329	0.85
	$n = 500$	0.102	0.107	0.267	0.124	0.542	1	0.119	0.954	1
	$n = 1000$	0.106	0.103	0.194	0.105	0.461	1	0.104	1	1
	$n = 5000$	0.105	0.098	0.109	0.106	0.196	0.986	0.094	1	1
	$n = 50000$	0.103	0.105	0.103	0.099	0.107	0.278	0.095	0.676	1
$\zeta_1 = 0 \quad \zeta_2 = 1$	$n = 100$	0.080	0.107	0.521	0.089	0.234	0.739	0.076	0.263	0.764
	$n = 500$	0.094	0.099	0.623	0.086	0.247	1	0.094	0.314	1
	$n = 1000$	0.087	0.099	0.42	0.099	0.162	1	0.095	0.199	1
	$n = 5000$	0.098	0.088	0.150	0.093	0.102	0.972	0.102	0.096	1
	$n = 50000$	0.101	0.096	0.095	0.099	0.095	0.230	0.104	0.098	1

Table 4: GAR Rejection Probabilities $\bar{\pi} = 0.5$

		$\rho = 0.9995$			$\rho = 0.999995$			$\rho = 1$		
		$m = 4$	$m = 6$	$m = 8$	$m = 4$	$m = 6$	$m = 8$	$m = 4$	$m = 6$	$m = 8$
$\zeta_1 = \zeta_2 = 0$	$n = 100$	0.927	0.893	0.999	1	1	1	1	1	1
	$n = 500$	0.495	0.477	0.939	1	1	1	1	1	1
	$n = 1000$	0.317	0.286	0.706	1	1	1	1	1	1
	$n = 5000$	0.145	0.144	0.222	1	1	1	1	1	1
	$n = 50000$	0.104	0.102	0.110	0.530	0.544	0.964	1	1	1
$\zeta_1 = 0 \quad \zeta_2 = 0.5$	$n = 100$	0.761	0.895	1	0.988	1	1	0.992	1	1
	$n = 500$	0.292	0.480	1	0.690	1	1	0.713	1	1
	$n = 1000$	0.193	0.283	1	0.404	1	1	0.425	1	1
	$n = 5000$	0.106	0.125	0.642	0.148	1	1	0.148	1	1
	$n = 50000$	0.100	0.106	0.150	0.102	0.340	1	0.107	1	1
$\zeta_1 = 0 \quad \zeta_2 = 1$	$n = 100$	0.171	0.707	1	0.200	1	1	0.194	0.996	1
	$n = 500$	0.101	0.277	1	0.097	0.996	1	0.089	0.955	1
	$n = 1000$	0.096	0.182	1	0.091	0.936	1	0.098	0.349	1
	$n = 5000$	0.088	0.109	0.978	0.094	0.306	1	0.092	0.102	1
	$n = 50000$	0.095	0.105	0.220	0.102	0.108	1	0.091	0.102	1

5 Conclusion

This paper studies identification robust inference based on the GAR statistic with general forms of identification failure. As demonstrated the non-singular variance assumption is inextricably linked to the assumption of first order identification. This issue has largely been overlooked in the identification literature. A notable exception is Andrews and Cheng (2012) who deal with the singular variance from identification failure under an assumption the form of singular variance is known up to model parameters.

In order to study properties of the GAR statistic with singular variance second order expansions of the eigensystem of the moment variance matrix around the true parameter were derived. This asymptotic approach is new in the identification literature and will prove useful for extending results for other identification robust statistics.

Without making any identification assumptions (and hence allowing for general forms of singular variance) the GAR statistic is asymptotically χ^2_m under a further set of conditions. Crucially one condition requires the null space of the moment variance matrix lie within that of the outer product of the expected first order derivative matrix. When this assumption is violated the GAR statistic is unbounded. In this case confidence sets based on inverting the GAR statistic would asymptotically yield the empty set. This result is unknown in the literature and is termed the 'moment-singular bias'

Examples of how this condition could be violated are provided. Roughly speaking this problem can occur when moments are not weakly identified and are highly correlated at the true parameter. This paper models moments as exactly singular, an interesting extension would model moments as weakly-singular. Namely model the smallest eigenvalues as shrinking to zero at some rate, analogous to the weak-instrument methodology for modeling weak identification. Simulation evidence shows that when the condition on the null space of Ω and GG' is almost not satisfied that the GAR statistic in general is oversized.

The majority of the literature on properties of estimators and identification robust inference make the assumption moments have non-singular variance or singular variance of known form. This paper is the first step in providing a platform to extend results in other settings without making a non-singular variance assumption, or assumptions on the form of singularity Andrews & Cheng (2012). Examples include dropping the non-singular variance assumption for identification robust inference from the GEL objective function made in Guggenberger, Ramalho & Smith (2008).

6 Appendix

Definitions

For any random variables x $\mathbb{E}[x]$ refers to the mathematical expectation taken with respect to (w.r.t) the density of x. Denote $\xrightarrow{p}$,$\xrightarrow{d}$, as convergence in probability and convergence in distribution respectively. For any deterministic sequence a_n and constant b then $a_n \to b$ denotes b as the deterministic limit of a_n. $\overset{d}{\sim}$ is shorthand for ' is distributed as" and 'w.p.a.1' denotes 'with probability approaching 1' and 'w.p.1' denotes 'with probability 1'. $o_p(a)$ refers 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 A refer to any arbitrary matrix then $\mathcal{R}(\cdot)$ be such that $\mathcal{R}(A)$ denote the rank of A .$\|A\|$ and $tr(A)$ are the Euclidean Norm and Trace of a matrix A respectively. CMT refers to the Continuous Mapping Theorem. For any $a > 0$ $I_{a\times a}$ refers to the $a \times a$ Identity Matrix and 0_a an $a \times 1$ vector of zeroes. $a \wedge b := \max\{a, b\}$

Appendix A1: Auxiliary Lemmas

Lemma A1: w.p.1

$$\hat{\Lambda}_0 = 0$$

Proof of Lemma A1:

$P_0'\Omega P_0 = 0$ by definition of P_0.

$$\mathbb{E}[P_0' g_i g_i' P_0] = 0$$

Since $P_0' g_i g_i' P_0$ is p.d then $P_0' g_i = 0$ $a.s(z)$ Hence $P_0'\hat{\Omega}P_0 = \frac{1}{n}\sum_{i=1}^{n} P_0' g_i g_i' P_0 = 0$ So the rank of $\hat{\Omega}(\theta_0) \le m - \bar{m}$ w.p.1 and hence $\hat{\Lambda}_0(\theta_0) = 0$.

Q.E.D

Lemma A2: *Let $\hat{A}$ and A be two square matrices of dimension r where* Rank$(A) = \bar{r}$ and $\|\hat{A} - A\| = O_p(\epsilon_n)$ *for some bounded non-negative sequence ϵ_n. Eigen-decompose $A = RDR'$* where $RR' = I_{r\times r}$ *and $RDR' = R_+D_+R_+' + R_0D_0R_0'$ where $D_0 = 0_{\bar{r}\times\bar{r}}$ and D_+ is a full rank diagonal $(r-\bar{r})\times(r-\bar{r})$ matrix with the eigenvalues of A on the diagonal where $0 \le \|D_+\| \le K$ for $K < \infty$. Similarly express $\hat{A} = \hat{R}\hat{D}\hat{R}' = \hat{R}_+\hat{D}_+\hat{R}_+' + \hat{R}_0\hat{D}_0\hat{R}_0'$. Define $B = \hat{A} - A$*

$$\hat{R}_+ = R_+ - R_0R_0'B'R_+D_+^{-1} + O_p(\epsilon_n^2)$$

$$\hat{R}_0 = R_0 - R_+D_+^{-1}R_+'BR_0 + O_p(\epsilon_n^2)$$

By CS$\|R_0R_0'B'R_+D_+^{-1}\| \le \|R_0\|^2 O_p(\|D_+\|)O_p(\|B\|) = O_p(\epsilon_n)$ since $\|D_+\| = O(1)$ then

$$\hat{R}_+ = R_+ + O_p(\epsilon_n)$$

Proof of Lemma A2: This result follows from equations (8),(9) in Hassani et al. (2011) .

Q.E.D

Lemma A3: *Under A1,A2*

$$\|\Delta_n\|^{-2}P_0'\hat{\Omega}(\theta_n)P_0 \xrightarrow{p} \Gamma$$

Proof of Lemma A3:

$$\hat{\Omega}(\theta_n) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta_n)g_i(\theta_n)'$$

Taylor expand $g_i(\theta_n)$ around θ_0

$$g_i(\theta_n) = g_i + G_i(\bar{\theta}_n)\Delta_n \tag{15}$$

Where $\bar{\theta}_n$ is a vector between θ_0 and θ_n Define $\bar{G}_i := G_i(\bar{\theta}_n)$

$$\hat{\Omega}(\theta_n) = \hat{\Omega} + \frac{1}{n}\sum_{i=1}^{n}\bar{G}_i\Delta_n\Delta_n'\bar{G}_i' + \frac{1}{n}\sum_{i=1}^{n} g_i\Delta_n'\bar{G}_i' + \frac{1}{n}\sum_{i=1}^{n}\bar{G}_i\Delta_n g_i' \tag{16}$$

By Lemma A1(i) $\Pr\{P_0'g_i(\theta_0) = 0\} = 1$ so that w.p.1

$$P_0'\hat{\Omega}(\theta_n)P_0 = \frac{1}{n}\sum_{i=1}^{n} P_0'\bar{G}_i\Delta_n\Delta_n'\bar{G}_i'P_0 \tag{17}$$

$$= \frac{1}{n}\sum_{i=1}^{n} P_0'((\bar{G}_i - G_i)\Delta_n\Delta_n'\bar{G}_i' + G_i\Delta_n\Delta_n'(\bar{G}_i - G_i)')P_0 + \frac{1}{n}\sum_{i=1}^{n} P_0'G_i\Delta_n\Delta_n'G_i'P_0$$

By repeated application of CS

$$\|\frac{1}{n}\sum_{i=1}^{n} P_0'(\bar{G}_i - G_i)\Delta_n\Delta_n'G_i'P_0\| \le \|\Delta_n\|^2\|\|P_0\|^2\frac{1}{n}\sum_{i=1}^{n}\|G_i'(\bar{G}_i - G_i)\| \tag{18}$$

$$\le \|\Delta_n\|^2\|\|P_0\|^2\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|G_i\|\tfrac{1}{n}\sum_{i=1}^{n}\|\bar{G}_i - G_i\|$$

By A2(iii) $\frac{1}{n}\sum_{i=1}^{n}\|\bar{G}_i - G_i\| = O_p(\|\Delta_n\|)$ and $\frac{1}{n}\sum_{i=1}^{n}\|G_i\| = O_p(1)$ by A2 (i),(iv). Since $\|P_0\| = \bar{m} < \infty$ by A1(iii)

$$\|\frac{1}{n}\sum_{i=1}^{n} P_0'((\bar{G}_i - G_i)\Delta_n\Delta_n'G_i'P_0\| = O_p(\|\Delta_n\|^3) \tag{19}$$

Similarly it can be shown that $\|\frac{1}{n}\sum_{i=1}^{n} P_0'((\bar{G}_i - G_i)\Delta_n\Delta_n'\bar{G}_i'P_0\| = O_p(\|\Delta_n\|^3)$ Define $\hat{\Gamma}_n = P_0'\frac{1}{n}\sum_{i=1}^{n} G_i\bar{\Delta}_n\bar{\Delta}_n'G_i'P_0$, $\Gamma_n = P_0'\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[G_i\bar{\Delta}_n\bar{\Delta}_n'G_i']P_0$, Then by (18) and (19) substituted in to (17) implies

$$\|\Delta_n\|^{-2}P_0'\hat{\Omega}(\theta_n)P_0 = \hat{\Gamma}_n + O_p(\|\Delta_n\|) \tag{20}$$

Finally to show $\hat{\Gamma}_n \xrightarrow{p} \Gamma$ establishing the result
As $\mathbb{E}[\hat{\Gamma}_n] = \Gamma_n$ and by application of CS

$$\|\Gamma_n\| \leq \|\bar{\Delta}_n\|^2\mathbb{E}[\|G_i\|^2] = O(1) \tag{21}$$

Where $\bar{\Delta}_n = O(1)$ and by A2(iv) $\mathbb{E}[\|G_i\|^2] = O(1)$
Under A2(i) $w_i(i = 1, .., n)$ is i.i.d and $\mathbb{E}[\hat{\Gamma}_n] = \Gamma_n \to \Gamma$ by CMT (since $\bar{\Delta}_n\bar{\Delta}_n' \to \Delta\Delta'$ and Γ_n is a continuous function of the bounded sequence $\bar{\Delta}_n$). An application of the Khinctine Weak Law of Large of Numbers (KWLLN) element by element to $\hat{\Gamma}_n$ then $\hat{\Gamma}_n \xrightarrow{p} \Gamma$ and by (20) noting that $\|\Delta_n\| = o_p(n^{-1/2})$ establishes the result.

Q.E.D

Lemma A4: *Under A1, A2*

$$\|\Delta_n\|^{-1}P_0'\hat{\Omega}(\theta_n) \xrightarrow{p} \Psi$$

Proof of Lemma 4:

By Lemma A1(i) and (16)

$$P_0'\hat{\Omega}(\theta_n) = P_0'\sum_{i=1}^{n}\bar{G}_i\Delta_n g_i' + P_0'\frac{1}{n}\sum_{i=1}^{n}\bar{G}_i\Delta_n\Delta_n'\bar{G}_i' \tag{22}$$

Where $\|P_0'\frac{1}{n}\sum_{i=1}^{n}\bar{G}_i\Delta_n\Delta_n'\bar{G}_i'\| = O_p(\|\Delta_n\|^2)$ as shown in the proof of Lemma A3 as $\|\Gamma\| = O(1)$ by (1.21)

$$P_0'\hat{\Omega}(\theta_n) = P_0'\sum_{i=1}^{n}\bar{G}_i\Delta_n g_i' + O_p(\|\Delta_n\|^2) \tag{23}$$

By CS,

$$\|P_0'\frac{1}{n}\sum_{i=1}^{n}(\bar{G}_i - G_i)\Delta_n g_i'\| \leq \|P_0\|\frac{1}{n}\sum_{i=1}^{n}\|\bar{G}_i - G_i\|\|\Delta_n\|\|\frac{1}{n}\sum_{i=1}^{n}\|g_i\| \tag{24}$$

Where $\frac{1}{n}\sum_{i=1}^{n}\|g_i\| = O_p(1)$ by KWLLN under A2(i) and A2(ii) that $\mathbb{E}[\|g_i\|^2] = O(1)$ and $\frac{1}{n}\sum_{i=1}^{n}\|\bar{G}_i - G_i\| = O_p(\|\Delta_n\|)$ by A2(iii) so that $\|P_0'\frac{1}{n}\sum_{i=1}^{n}(\bar{G}_i - G_i)\Delta_n g_i'\| = O_p(\|\Delta_n\|^2)$.
Define $\hat{\Psi}_n := P_0'\frac{1}{n}\sum_{i=1}^{n}G_i\bar{\Delta}_n g_i'$, $\Psi_n = P_0'\mathbb{E}[G_i\bar{\Delta}_n g_i']$ then by (24)

$$\|\Delta_n\|^{-1}P_0'\frac{1}{n}\sum_{i=1}^{n}G_i\Delta_n g_i' = \hat{\Psi}_n + O_p(\|\Delta_n\|) \tag{25}$$

Since $\mathbb{E}[\hat{\Psi}_n] = \Psi_n$ where Ψ_n is bounded for all n since by CS

$$\|\Psi_n\| \leq \|\bar{\Delta}_n\|\mathbb{E}[\|G_i\|]\mathbb{E}[\|g_i\|] \tag{26}$$

Where $\mathbb{E}[\|G_i\|] = O(1)$ $\mathbb{E}[\|g_i\|] = O(1)$ by A2 (ii),(iv). By the KWLLN $\hat{\Psi}_n \xrightarrow{p} \Psi_n$ where $\|\bar{\Delta}_n\| = O(1)$ where $\Psi_n \to \Psi$ by CMT establishing the result.

Q.E.D

Appendix A2: Main Theorems

Proof of Theorem 1:

Define the following from Lemma A2
$\hat{A} = \hat{\Omega}(\theta_n)$, $A = \Omega$ where $B = \hat{\Omega}(\theta_n) - \Omega$ and $\|\hat{\Omega}(\theta_n) - \Omega\| \leq \|\hat{\Omega}(\theta_n) - \hat{\Omega}\| + \|\hat{\Omega} - \Omega\|$ by T , $\|\hat{\Omega}(\theta_n) - \hat{\Omega}\| = O_p(\|\Delta_n\|)$ by A1(ii)so that $\epsilon_n := \|\hat{\Omega} - \Omega\| \wedge \|\Delta_n\|$ where $R_+ = P_+$, $R_0 = P_0$, $\hat{R}_+ = \hat{P}_+(\theta_n)$, $\hat{R}_0 = \hat{P}_0(\theta_n)$ and $D_+ = \Lambda_+$ then Since $\|\Lambda_+^{-1}\|\|P_0\|\|P_+\|\|\hat{\Omega}(\theta_n) - \hat{\Omega}\| = O(1)O_p(\|\Delta_n\|)$ since $m = O(1)$ by A1 (iii) hence $\|P_0\| = \bar{m} = O(1)$ where $0 \leq \bar{m} \leq m$ and $\|P_+\| = m - \bar{m} = O(1)$ where $\|\Lambda_+^{-1}\| = O(1)$ by A1(i).
Then by Lemma A2

$$\hat{P}_+(\theta_n) = P_+ + O_p(\|\hat{\Omega} - \Omega\| \wedge \|\Delta_n\|) \tag{27}$$

Establishing (1.5).

$$\|\hat{\Lambda}(\theta_n) - \Lambda\| \leq \|\hat{\Omega}(\theta_n) - \Omega\| \tag{28}$$

By Theorem 4.2 of Bosq (2000). Where it has been shown that $\|\hat{\Omega}(\theta_n) - \Omega\| = O_p(\|\hat{\Omega} - \hat{\Omega}\| \wedge \|\Delta_n)$ establishing (6).
Now to show (7) and (8) again using Lemma A2,

$$\hat{P}_0(\theta_n) = P_0 - \Omega_+^*\hat{\Omega}(\theta_n)P_0 + O_p((\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|)^2) \tag{29}$$

Establishing (7).

$$\hat{\Lambda}_0(\theta_n) = \hat{P}_0(\theta_n)'\hat{\Omega}(\theta_n)\hat{P}_0(\theta_n) \tag{30}$$

$$= (\hat{P}_0(\theta_n) - P_0)'\hat{\Omega}(\theta_n)(\hat{P}_0(\theta_n) - P_0) + P_0'\hat{\Omega}(\theta_n)(\hat{P}_0(\theta_n) - P_0)$$
$$+(\hat{P}_0(\theta_n) - P_0)'\hat{\Omega}(\theta_n)P_0 + P_0'\hat{\Omega}(\theta_n)P_0$$

Where by (7) $\hat{P}_0(\theta_n) - P_0 = -\Omega_+^* \hat{\Omega}(\theta_n) P_0 + O_p(\|\Delta_n\|^2)$
Noting that $\Omega = \Omega_+$ and by CS $\|\Omega_+^* \hat{\Omega}(\theta_n) P_0\| \leq \|\Omega_+^*\| \|P_0\| \|\hat{\Omega}(\theta_n) - \hat{\Omega}(\theta_0)\| = O_p(\|\Delta_n\|)$ since $\|\Omega_+^*\| = O(1)$ by A1(i) and $P_0 \hat{\Omega}(\theta_n) = P_0'(\hat{\Omega}(\theta_n) - \hat{\Omega}(\theta_0))$ by Lemma A1(i) so that,

$$(\hat{P}_0(\theta_n) - P_0)' \hat{\Omega}(\theta_n)(\hat{P}_0(\theta_n) - P_0) \tag{31}$$

$$= P_0' \hat{\Omega}(\theta_n) \Omega_+^* \hat{\Omega}(\theta_n) P_0 + O_p(\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|)^3)$$

$$P_0' \hat{\Omega}(\theta_n)(\hat{P}_0(\theta_n) - P_0) \tag{32}$$

$$= -P_0' \hat{\Omega}(\theta_n) \Omega_+^* (\hat{\Omega}(\theta_n) P_0 + O_p((\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|)^3)$$

Hence plugging (31),(32) in to(30)

$$\hat{\Lambda}_0(\theta_n) = P_0' \hat{\Omega}(\theta_n) P_0 - P_0' \hat{\Omega}(\theta_n) \Omega_+^* \hat{\Omega}(\theta_n) P_0 + O_p((\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|)^3) \tag{33}$$

Which establishes (8).

Q.E.D

PROOF OF THEOREM 2:
By (1.7)

$$\hat{P}_+ = P_+ + O_p(\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|) \tag{34}$$

$$\hat{\Lambda}_+ = \Lambda_+ + O_p(\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|) \tag{35}$$

Where $\|\hat{\Omega} - \Omega\| = O_p(n^{-1/2})$ by A2(i),(ii) and $\|\Delta_n\| = o_p(n^{-1/2})$ establishing (9),(10).
By T1

$$\|\Delta_n\|^{-1}(\hat{P}_0(\theta_n) - P_0) = -\|\Delta_n\|^{-1} \Omega_+^* \hat{\Omega}(\theta_n) P_0 + o_p(n^{-1/2}) \tag{36}$$

Since $\|\Delta_n\|^{-1} O_p((\|\Delta_n\| \wedge \|\hat{\Omega} - \Omega\|)^2) = o_p(n^{-1/2})$ since $\|\Delta_n\| = o_p(n^{-1/2})$ By the CMT and Lemma A3 $\|\Delta_n\|^{-1} \Omega_+^* \hat{\Omega}(\theta_n) P_0 \xrightarrow{p} \Omega_+^* \Psi'$ establishing (1.11).

By (8)

$$\|\Delta\|^{-2} \hat{\Lambda}_0(\theta_n) = \|\Delta_n\|^{-2} P_0' \hat{\Omega}(\theta_0) P_0 \tag{37}$$

$$-\|\Delta_n\|^{-2} P_0' \hat{\Omega}(\theta_n) \Omega_+^* \hat{\Omega}(\theta_n) P_0 + o_p(n^{-1/2})$$

Since $\|\Delta_n\|^{-2}O_p((\|\Delta_n\| \wedge \|\hat{\Omega}-\Omega\|)^3) = o_p(n^{-1/2})$ By Lemma A2 $\|\Delta_n\|^{-2}P_0'\hat{\Omega}(\theta_0)P_0 \xrightarrow{p} \Gamma$ and by Lemma A3 and CMT
$\|\Delta_n\|^{-2}P_0'\hat{\Omega}(\theta_n)\Omega_+^*\hat{\Omega}(\theta_n)P_0 \xrightarrow{p} \Psi\Omega_+^*\Psi'$ establishing (1.12).

Q.E.D

Proof of Theorem 3:

$$\hat{T}_{GAR}(\theta_n) = n\hat{P}_+(\theta_n)'\hat{g}(\theta_n)'\hat{\Lambda}_+(\theta_n)^{-1}\hat{P}_+'(\theta_n)\hat{g}(\theta_n) \tag{38}$$

$$+n\hat{P}_0(\theta_n)'\hat{g}(\theta_n)'\hat{\Lambda}_0(\theta_n)^{-1}n\hat{P}_0(\theta_n)'\hat{g}(\theta_n)$$

Using the expansion of $\hat{g}(\theta_n)$ around θ_0 summed across i in (15)

$$\sqrt{n}\hat{g}(\theta_n) = \sqrt{n}\hat{g}(\theta_0) + \sqrt{n}\hat{G}(\bar{\theta}_n)\Delta_n \tag{39}$$

By repeated application of CS,

$$\|\sqrt{n}(\hat{G}(\bar{\theta}_n) - \hat{G}(\theta_0))\Delta_n\| \leq \sqrt{n}\|\Delta_n\|\frac{1}{n}\sum_{i=1}^{n}\|\bar{G}_i - G_i\| = O_p(n^{1/2}\|\Delta_n\|^2) \tag{40}$$

By A2 (ii) where $\|\Delta_n\|^2 n^{1/2} = o_p(n^{-1/2})$ hence

$$\sqrt{n}\hat{g}(\theta_n) = \sqrt{n}\hat{g}(\theta_0) + \sqrt{n}\hat{G}(\theta_0)\Delta_n + o_p(n^{-1/2}) \tag{41}$$

Firstly establish that

$$n(\hat{P}_+(\theta_n)'\hat{g}(\theta_n))'\hat{\Lambda}(\theta_n)^{-1}\hat{P}_+(\theta_n)'\hat{g}(\theta_n) = n(P_+'\hat{g}(\theta_0))'\Lambda_+^{-1}P_+'\hat{g}(\theta_0) + o_p(1) \tag{42}$$

By (1.9) $\hat{P}_+(\theta_n) = P_+ + o_p(1)$ and (41)

$$\hat{P}_+(\theta_n)'\sqrt{n}\hat{g}(\theta_n) = P_+'(\sqrt{n}\hat{g}(\theta_0) + \hat{G}(\theta_0)\sqrt{n}\Delta_n) + o_p(1) \tag{43}$$

$$= P_+'\sqrt{n}\hat{g}(\theta_0) + o_p(1) \tag{44}$$

Since $\|P_+'\hat{G}(\theta_0)\sqrt{n}\Delta_n\| \leq n^{1/2}\|P_+\|\|\hat{G}(\theta_0)\|\||\Delta_n\| = n^{1/2}O(1)O_p(1)o_p(n^{-1/2}) = o_p(1)$.

$\hat{\Lambda}_+(\theta_n) = \Lambda_+ + o_p(1)$ by (10) and under A1 (i) then Λ_+^{-1} exists so that by CMT

$$\hat{\Lambda}_+(\theta_n)^{-1} = \Lambda_+^{-1} + o_p(1) \tag{45}$$

Together with (44) implies (42) so that $n(\hat{P}_+(\theta_n)'\hat{g}(\theta_n))'\hat{\Lambda}(\theta_n)^{-1}\hat{P}_+(\theta_n)'\hat{g}(\theta_n) \xrightarrow{d} \chi^2_{m-\bar{m}}$. Since $\sqrt{n}P_0'\hat{g}(\theta_0) \xrightarrow{p} N(0,\Lambda_+)$ by A2(i),(ii) and the Lindberg-Levy Central Limit Theorem.

Under A1, A2 , A3 it can be shown that

$$\|\Delta_n\|^{-1}\sqrt{n}\hat{P}_0(\theta_n)'\hat{g}(\theta_n) = P_0'\sqrt{n}(\hat{G}(\theta_0) - G)\bar{\Delta}_n - \Psi\Omega_+^*\sqrt{n}\hat{g}(\theta_0) + o_p(1) \tag{46}$$

By (1.11) $\|\Delta_n\|^{-1}(\hat{P}(\theta_n) - P_0) = -\Omega_+^*\Psi' + o_p(1)$

$$\|\Delta_n\|^{-1}\sqrt{n}\hat{P}_0(\theta_n)'\hat{g}(\theta_n) = (-\Omega_+^*\Psi' + o_p(1))'\sqrt{n}\hat{g}(\theta_n) + \|\Delta_n\|^{-1}P_0'\sqrt{n}\hat{g}(\theta_n) \tag{47}$$

Where by (39) $\sqrt{n}\hat{g}(\theta_n) = \sqrt{n}\hat{g}(\theta_0)+o_p(1)$ hence $(-\Omega_+^*\Psi'+o_p(1))'\sqrt{n}\hat{g}(\theta_n) = -\Psi\Omega_+^*\sqrt{n}\hat{g}(\theta_0)+o_p(1)$ To established the first part on the right hand side of (1.46) note that

$$\|\Delta_n\|^{-1}P_0'\sqrt{n}\hat{g}(\theta_n) = P_0'(\hat{G}(\theta_n) - G)\bar{\Delta}_n + o_p(1) \tag{48}$$

Since by Lemma A1 (i) $P_0'\sqrt{n}\hat{g}(\theta_0) = 0$ w.p.1. and by A3(i) $P_0'G = 0$. By (12)

$$\|\Delta_n\|^{-2}\hat{\Lambda}_0(\theta_n) = \Phi + o_p(1) \tag{49}$$

Where Φ is p.d by A3(ii). By CMT and (49)

$$(\|\Delta_n\|^{-2}\hat{\Lambda}_0(\theta_n))^{-1} = \Phi^{-1} + o_p(1) \tag{50}$$

Together (146),(50) establish that w.p.a.1

$$n(\hat{P}_0(\theta_n)'\hat{g}(\theta_n))'\hat{\Lambda}_0(\theta_n)^{-1}\hat{P}_0(\theta_n)'\hat{g}(\theta_n)) \tag{51}$$

$$= (P_0'(\sqrt{n}(\hat{G}(\theta_0)-G)\bar{\Delta}_n - \Psi\Omega_+^*\sqrt{n}\hat{g}(\theta_0)))'\Phi^{-1}(P_0'(\sqrt{n}(\hat{G}(\theta_0)-G)\bar{\Delta}_n - \Psi\Omega_+^*\sqrt{n}\hat{g}(\theta_0)))$$

Now it can be established that

$$P_0'(\sqrt{n}(\hat{G}(\theta_0) - G)\bar{\Delta}_n - \Psi'\Omega_+^*\sqrt{n}\hat{g}(\theta_0)) \xrightarrow{d} N(0, \Phi) \tag{52}$$

Define $b_i = P_0'((G_i - G) - \Psi\Omega_+^* g_i)$

$$P_0'(\sqrt{n}(\hat{G}(\theta_0) - G)\bar{\Delta}_n - \Psi\Omega_+^*\sqrt{n}\hat{g}(\theta_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} b_i \tag{53}$$

Where $\mathbb{E}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} b_i] = 0$

$$\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} b_i b_i'] = P_0'\mathbb{E}[G_i\bar{\Delta}_n\bar{\Delta}_n{}'G_i']P_0 - \Psi_n\Omega_+^*\Omega\Omega_+^*\Psi_n' \tag{54}$$

By A1 (i) that w_i is i.i.d and by definition $\Psi_n = P_0'\mathbb{E}[G_i\bar{\Delta}_n g_i'] \to \Psi$ since $\bar{\Delta}_n \to \Delta$ where $\|\Delta\| < \infty$ (and likewise $\Gamma_n := P_0'\mathbb{E}[G_i\bar{\Delta}_n\bar{\Delta}_n' G_i']P_0 \to \Gamma$ by CMT) as $\mathbb{E}[\|G_i\|^2] < \infty$, $\mathbb{E}[\|g_i\|^2] < \infty$ by A2(ii),(iv).

$$\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} b_i b_i'] \to \Phi \tag{55}$$

As w_i is i.i.d then so is b_i and Φ then by the Multivariate Lindberg-Levy Central Limit theorem

$$P_0'(\sqrt{n}(\hat{G}(\theta_0) - G)\bar{\Delta}_n - \Psi\Omega_+^*\sqrt{n}\hat{g}(\theta_0)) \stackrel{d}{\rightarrow} N(0, \Phi) \tag{56}$$

Hence (51) converges in distribution to χ_m^2 since both terms on right hand side of (38) are orthogonal asymptotically and the sum of the two is asymptotically χ_m^2.

Q.E.D

Proof of Theorem 4:

Divide equation (51) by n (noting that $P_0'G \neq 0$ since A3(i) is violated) it is straightforward to establish that

$$(\hat{P}_0(\theta_n)'\hat{g}(\theta_n))'\hat{\Lambda}_0(\theta_n)^{-1}\hat{P}_0(\theta_n)'\hat{g}(\theta_n)) \stackrel{p}{\rightarrow} \Delta' G' P_0 \Phi^{-1} P_0' G \Delta \tag{57}$$

By A2(i),(iv) then $P_0\hat{G}(\theta_0) \stackrel{p}{\rightarrow} P_0'G$. Since the first term on the right hand side of (38) converges to zero in probability when divided by n, then it is straightforward to establish that $\hat{T}_{GAR}(\theta_n)/n \stackrel{p}{\rightarrow} \Delta' G' P_0 \Phi^{-1} P_0' G \Delta$.

Q.E.D

References

[1] ANDREWS, K. (19987). *Asymptotic Results for Generalized Wald Tests*, Econometric Theory, Cambridge University Press, vol. 3(03), pages 348-358, June.

[2] BATHIA, N. , YAO, Q. and ZIEGELMANN, F. (2010). *Identifying the finite dimensionality of curve time series*, Ann. Statist. Volume 38, Number 6 (2010), 3352-3386.

[3] BOSQ, D. (2000). *Linear Processes in Function Spaces*, New York: Springer- Verlag.

[4] GRANT, N. (2012). *GMM with Weakly Singular Variance* (work in progress).

[5] HANSEN, H.L (1982). *Large Sample Properties of Generalized Method of Moments Estimators*, Econometrica, 50, issue 4, p. 1029-54

[6] JAGANNATHAN, R. and WANG, G. (2002). *Generalized Method of Moments: Applications in Finance*, Journal of Business & Economic Statistics, American Statistical Association, vol. 20(4), pages 470-81, October.

[7] KATO, T. (1982). *A short introduction to the perturbation theory of linear operators.*, Springer-Verlag 1982.

[8] KLEIBERGEN, F. (2005). *Testing Parameters in GMM Without Assuming that They Are Identified*, Econometrica, Econometric Society, vol. 73(4), pages 1103-1123, 07.

[9] KLEIBERGEN, F. (2011). *Improved accuracy of weak instrument robust GMM statistics through bootstrap and Edgeworth approximations*, working paper.

[10] NEWEY, W. and MCFADDEN,D. (1994). *Large Sample Estimation and Hypothesis Testing*,Handbook of Econometrics, Vol.4, 2111-2245.

[11] NEWEY, W.N. and WINDMEIJER, F. (2009). *Generalized Method of Moments With Many Weak Moment Conditions*, Econometrica, Econometric Society, vol. 77(3), pages 687-719, 05.

[12] PENARANDA, F. and SENTANA, E. (2010). *Spanning tests in return and stochastic discount factor mean-variance frontiers: A unifying approach*, Economics Working Papers 1101, Department of Economics and Business, Universitat Pompeu Fabra, revised Sep 2010.

[13] Renault, E. and Donovon, P. (2009). *GMM Overidentification Test with First Order Underidentification*, ? working paper (2009).

[14] Sargan, J. D. (1983). *Identification and Lack of Identification*, Econometrica, Econometric Society, vol. 51(6), pages 1605-33, November

[15] Staiger, D. and Stock, J.H. (1997). *Instrumental Variables Regression with Weak Instruments*, Econometrica, Econometric Society, vol. 65(3):557-586.

[16] Stock, J.H. and Wright, J. (2000). *GMM with Weak Identification*,Econometrica, Econometric Society, vol. 68(5), pages 1055-1096, September.