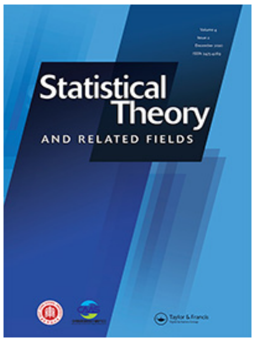

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Efficient GMM estimation with singular system of moment conditions

Zhiguo Xiao

STATISTICAL THEORY AND RELATED FIELDS
2020, VOL. 4, NO. 2, 172–178
https://doi.org/10.1080/24754269.2019.1653159

Efficient GMM estimation with singular system of moment conditions

Zhiguo Xiao

Department of Statistics, School of Management, Fudan University, Shanghai, People's Republic of China

ABSTRACT
Standard generalised method of moments (GMM) estimation was developed for nonsingular system of moment conditions. However, many important economic models are characterised by singular system of moment conditions. This paper shows that efficient GMM estimation of such models can be achieved by using the reflexive generalised inverses, in particular the Moore–Penrose generalised inverse, of the variance matrix of the sample moment conditions as the weighting matrix. We provide a consistent estimator of the optimal weighting matrix and establish its consistency. Potential issues of using generalised inverse and some remedies are also discussed.

ARTICLE HISTORY
Received 21 April 2019
Accepted 5 August 2019

KEYWORDS
Singular moment condition models; generalised method of moments; reflexive generalised inverses; imposing random noises

JEL CLASSIFICATIONS
C13; C14

1. Introduction

Over the past several decades, a great deal of statisticians' efforts has been devoted to the statistical inference of moment condition models, i.e., models where the linkage between parameter and data is specified through a set of moment restrictions (also known as estimating equations). Technically, a moment condition model specifies that the data generating process of observations $Z_1, \ldots, Z_n$ satisfies

$$\mathbf{E}\left[g(Z_i, \beta_0)\right] = 0, \qquad (1)$$

where g is a $\mathbb{R}^K$-valued known function and β_0 is the $\mathbb{R}^p$-valued parameter of interest, and $K \geq p$. The popularity of moment condition models is partially due to that a parametric likelihood form may be too strong for many real applications or scientific theories. When the dimension of parameter of interest equals the number of moment conditions, the parameter is said to be just-identified, and the classical approach of the method of moments can be applied for parameter estimation. In practice, a majority of the moment condition models investigated by applied researchers, such as the models for assets pricing and dynamic panel data, are over-identified. The generalised method of moments (GMM) of Hansen (1982) is one of the most popular techniques that are designed for the estimation of over-identified moment condition models (see, e.g., Hansen & West, 2002 and Hall, 2005).

Like many other classical statistical methods, GMM comes with the price of a set of regularity conditions which warrant its validity. Although in most applications those regularity conditions are not binding, some of them can be violated in interesting circumstances. This paper is concerned with the efficient GMM estimation when one of the regularity conditions of standard GMM, that the covariance matrix of the moment vector evaluated at the true parameter be of full rank, is violated. A typical such kind of violation appears when the system of moment conditions is singular, i.e., some components of the moment functions are linear combinations of each other.

Singular systems of moment conditions exist in a wide variety of economic studies, such as the consumer expenditure function analysis (Barten, 1969, 1977), the market share analysis (Rao, 1972; Weiss, 1968), the production function estimation (Dhrymes, 1962), the translog utility function analysis (Berndt & Christensen, 1974), the linearised dynamic stochastic general equilibrium (DSGE) modelling (Bierens, 2007; Ireland, 2004), the errors-in-variables analysis with panel data (Biørn, 2000; Biørn & Klette, 1998; Wansbeek, 2001; Xiao, Shao, & Palta, 2010a, 2010b; Xiao, Shao, Xu, & Palta, 2007), the multivariate random-effects meta-analysis models (Chen, Hong, & Riley, 2014; Riley, Abrams, Lambert, Sutton, & Thompson, 2007) and the non-Gaussian ARMA models (Alessi, Barigozzi, & Capasso, 2011; Leeper, Walker, & Yang, 2013; Mountford & Uhlig, 2009; Velasco & Lobato, 2018).

In a linear regression model with known singular disturbance covariance matrix, Theil (1971) showed that a generalised Aitken-like estimator using the Moore–Penrose generalised inverse is best linear

CONTACT Zhiguo Xiao zhiguo_xiao@fudan.edu.cn Department of Statistics, School of Management, Fudan University, Room 709 Siyuan Faculty Building, 670 Guoshun Rd., Shanghai 200433, People's Republic of China

unbiased. Following Theil (1971), Kreijger and Neudecker (1977) proposed two optimality criteria to obtain best linear unbiased estimators. Within the same context, Dhrymes and Schwarz (1987) discussed the existence issue of the estimators using generalised inverses. Haupt and Oberhofer (2006) proposed an estimator which does not use the generalised inverses and allows for additional exogenous restrictions, collinearities and generalised adding-up. Bierens and Swanson (2000) and Bierens (2007) suggested that one can obtain parameter estimate by maximising the information content of the singular system. Ireland (2004) and Lai (2008) proposed adding random noises to the singular system to implement maximum likelihood estimation.

In the GMM literature, White (1986) showed that if the estimating function g is of the form $g = (g_1', g_2')'$ such that: (i) $\Omega_1 = \mathbf{E}[g_1(Z_i, \beta_0)g_1(Z_i, \beta_0)']$ is nonsingular and (ii) components of g_2 are linear combinations of g_1, then the efficient GMM estimator is the minimiser of

$$J_n(\beta) = n\bar{g}_n(\beta)'\Omega^{-}\bar{g}_n(\beta), \qquad (2)$$

where $\bar{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g(Z_i, \beta)$ and Ω^{-} is a reflexive generalised inverse of

$$\Omega = \mathbf{E}\left[g(Z_i, \beta_0)g(Z_i, \beta_0)'\right]. \qquad (3)$$

However, in practice, the aforementioned representation of g is generally not readily obtainable (see, e.g., Schneeweiss, 2014; Velasco & Lobato, 2018; Xiao et al., 2010b).

The purpose of this article is to develop an efficient GMM estimator for a singular system of moment conditions with general form. An earlier effort appeared in Xiao (2008), which is proposed using the reflexive generalised inverses to deal with the singularity. Schneeweiss (2014) independently discussed similar ideas.

The rest of the paper is organised as follows. In Section 2, we briefly review the GMM methodology, the concepts of generalised inverses and some results of the reflexive generalised inverses. We present our main result in Section 3. Section 4 discusses further issues such as the estimation of optimal weighting matrix and the method of adding noises, and Section 5 concludes. Proofs of results are relegated to the Appendix.

2. GMM and generalised inverses

We first make a brief introduction of the standard GMM method. For book-length detailed account, see Hall (2005). For simplicity we assume that the data $Z_1, \ldots, Z_n$ are i.i.d. Assume also that $K > p$, i.e., the model is over-identified. Since the number of restrictions on parameter is greater than the dimension of parameter, in general it is impossible to obtain an estimator of the parameter by using method of moments, i.e., by setting the sample moment $\bar{g}_n$ equal to zero. The idea of GMM by Hansen (1982) is to minimise a quadratic norm of $\bar{g}_n$:

$$J_n(\beta) = n\bar{g}_n(\beta)'W_n\bar{g}_n(\beta), \qquad (4)$$

where W_n is a positive semidefinite matrix. Under a set of regularity conditions including that Ω being positive definite, and assuming W_n converges in probability to a positive semi-definite matrix W, $\hat{\beta}_{GMM}$, the minimiser of (4), is a consistent estimator for β_0 and has limiting distribution

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta_0) \xrightarrow{d} N(0, V(W)),$$

where $V(W) = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$ with $G = E[\frac{\partial g(Z_i,\beta_0)}{\partial \beta}]$. The lower bound of $V(W)$ is achieved at $W = \Omega^{-1}$, i.e.,

$$V(W) \geq V(\Omega^{-1})$$

in the sense of being nonnegative definite, for any W. In practice, a consistent estimator of Ω^{-1} can be set as

$$\hat{\Omega}^{-1} = \left[\frac{1}{n}\sum_{i=1}^{n}\left[g(Z_i, \tilde{\beta})g(Z_i, \tilde{\beta})'\right]\right]^{-1}, \qquad (5)$$

where $\tilde{\beta}$ is a consistent estimator of β_0. A typical choice of $\tilde{\beta}$ is a GMM estimator with $W = I_K$, the identity matrix of order K. Note that $\hat{\Omega}^{-1}$ converges in probability to Ω^{-1} because $\frac{1}{n}\sum_{i=1}^{n}[g(Z_i, \tilde{\beta})g(Z_i, \tilde{\beta})']$ converges in probability to Ω, and more importantly, Ω is positive definite.

Next we review the concepts of generalised inverses of a matrix and some of their properties.

Definition 2.1: Let A be a real $l \times s$ matrix. An $s \times l$ real matrix A^{-} may have one or all of the following properties:

(i) $AA^{-}A = A$;
(ii) $A^{-}AA^{-} = A^{-}$;
(iii) $(AA^{-})' = AA^{-}$;
(iv) $(A^{-}A)' = A^{-}A$.

If A^{-} satisfies (i), it is called a **generalised inverse** of A; if A^{-} satisfies (i) and (ii), it is called a **reflexive generalised inverse** (or $\mathbf{g}_2$**-inverse**) of A; if A^{-} satisfies (i)–(iv), it is called the **Moore–Penrose generalised inverse** of A. The Moore–Penrose generalised inverse of a matrix A is unique and is denoted by A^{+} hereafter.[1]

[1] For the existence and uniqueness of the Moore–Penrose generalised inverse, see, e.g., Penrose (1955) and Abadir and Magnus (2005, pp. 284–285).

We list some of the important properties of the generalised inverses by the following two propositions, proof of which can be achieved by direct verification and therefore is omitted.[2] Proposition 1 states that when the matrix of interest has natural factorisation with certain structure, some of its generalised inverses can be easily derived.

Proposition 2.1: *(i) Let* $\Omega = [A_1' \;\; A_2']'\Omega_1[A_1' \;\; A_2']$, *where* A_1 *and* Ω_1 *are nonsingular square matrices. Then* $\Omega^- = \begin{bmatrix} (A_1^{-1})'\Omega_1^{-1}A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}$ *is a reflexive generalised inverse of* Ω.

(ii) Let $\Omega = A\Omega_1 A'$, *where* A *is of full column rank and* Ω_1 *is nonsingular, then any reflexive generalised inverse* Ω^- *of* Ω *satisfies* $A'\Omega^- A = \Omega_1^{-1}$. *Moreover, we have* $A^+ = (A'A)^{-1}A'$, *and*

$$\Omega^+ = (A^+)'\Omega_1^{-1}A^+ = A(A'A)^{-1}\Omega_1^{-1}(A'A)^{-1}A'.$$

Proposition 2.2 points out that the generalised inverses (including the reflexive generalised inverses) are not unique and can be obtained by using the singular value decomposition .[3]

Proposition 2.2: *Let* Ω *be an* $m \times n$ *real valued matrix with rank* $r > 0$. *Suppose that the singular value decomposition of* Ω *is* $\Omega = S\Sigma T'$, *where* S *is* $m \times m$ *with* $S'S = I_m$, T *is* $n \times n$ *with* $T'T = I_n$, *and* $\Sigma = \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}$, *with* Σ_r *the diagonal matrix of singular values of* Ω *and* O *the matrices of zeros. Then*

(i)

$$G = T\begin{bmatrix} \Sigma_r^{-1} & X \\ Y & Z \end{bmatrix}S' \tag{6}$$

is a generalised inverse of Ω, *where* X, Y *and* Z *are arbitrary real valued matrices with appropriate dimension.*

(ii)

$$G = T\begin{bmatrix} \Sigma_r^{-1} & X \\ Y & Y\Sigma_r^{-1}X \end{bmatrix}S' \tag{7}$$

is a reflexive generalised inverse of Ω, *where* X, Y *are arbitrary real valued matrices with appropriate dimension.*

(iii)

$$G = T\begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix}S' \tag{8}$$

is the Moore–Penrose generalised inverse of Ω.

White (1986) result on GMM estimation with singular moment conditions can be stated as follows:

Theorem 2.1 (White, 1986): *Suppose there exists a matrix* Δ *such that* $G = [I_{l_1} \;\; \Delta]'G_1$ *and* $\Omega = [I_{l_1} \;\; \Delta]'\Omega_1 [I_{l_1} \;\; \Delta]$, *where* G_1 *is of full column rank and* Ω_1 *is* $l_1 \times l_1$ *positive definite with* $l_1 \geq p$, *then*

(i) For any reflexive generalised inverse Ω^- *of* Ω, $\Omega - G(G'\Omega^- G)^{-1}G'$ *is independent of the choice of* Ω^-, *and* $\Omega - G(G'\Omega^- G)^{-1}G' \geq 0$.

(ii) For any reflexive generalised inverse Ω^- *of* Ω, *and for any* W,

$$(G'WG)^{-1}G'W\Omega WG(G'WG)^{-1} \geq (G'\Omega^- G)^{-1}.$$

Hence Ω^- *is the optimal weighting matrix. In practice, one may choose the Moore–Penrose generalised inverse of* Ω, *or a special reflexive generalised inverse*

$$\Omega^- = \begin{bmatrix} \Omega_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Remark: Note that $(G'\Omega^- G)^{-1}$ – the asymptotic covariance matrix of the optimal GMM estimator – does not depend on Δ. The basic idea of Theorem 2.1 is as follows. Suppose we have two sets of instrumental variables, say Z_1 and Z_2, such that Z_1 is linearly independent and Z_2 is a linear combination of Z_1, i.e., $Z_2 = Z_1\alpha$ for some constant vector α, then one can ignore Z_2 and use Z_1 only as instruments, in doing so we achieve the same asymptotic efficiency as using $Z = (Z_1 Z_2)$. The limitation of this result is that to apply this method we have to sort all instrumental variables into two groups, such that instrumental variables in one group are linear combinations of those in the other group. This can be very tedious in practice. For example, in panel data models, there are often a very large number of instruments and it is in general impossible to sort out them.

3. Main results

We now establish some basic results about random vectors with singular covariance matrices.

Lemma 3.1: *Let* $Y = [Y_1 \ldots Y_m]'$ *be an* $m \times 1$ *random vector and* r *be the rank of the covariance matrix of* Y. *Suppose that* $r < m$. *Then*

(i) There exist a r*-dimensional subvector* $Y^{(r)} = [Y_{i_1} \ldots Y_{i_r}]'$ *of* Y *such that its covariance matrix* $\mathbf{var}(Y^{(r)})$ *is positive definite. The vector* $Y^{(r)}$ *is called an essential subvector of* Y.

(ii) Let $Y^{-(r)}$ *be the* $(m-r) \times 1$ *vector consisting of the remaining components of* Y. *Then there exist an* $(m-r) \times (m-r)$ *constant matrix* C *and an* $(m-r) \times 1$ *constant vector* d *such that*

$$Y^{-(r)} = CY^{(r)} + d, \quad \text{w.p.1.},$$

where w.p.1. means 'with probability one'. Hence there exist an $m \times r$ *constant matrix* B *of full column rank and*

[2] More results on reflexive generalised inverses can be found in Rao and Mitra (1971), Rao (2001), Bapat (2012), Fampa and Lee (2018), and Xu, Fampa, and Lee (2019).

[3] Similar results for square matrices can be found in Bapat (2012, pp. 47–48).

an $m \times 1$ *constant vector* $\widetilde{d}$ *such that*

$$Y = BY^{(r)} + \widetilde{d}, \quad \text{w.p.1.}$$

(iii) If $\mathbf{E}Y = 0$*, then* $\widetilde{d}$ *in (ii) is the zero vector. i.e.,* $Y = BY^{(r)}$*, w.p.1.*

Theorem 3.1 is the main result of this paper.

Theorem 3.1: *Consider GMM estimation for model* (1) *with* Ω *defined by* (3)*. Suppose it is known that the components of* $g(Z_i, \beta_0)$ *are linearly dependent (with probability one). Then any reflexive inverse of* Ω *is an optimal weighting matrix. Particularly, we can use the Moore–Penrose generalised inverse* Ω^+*.*

Let Ω^- be an arbitrary reflexive generalised inverse of Ω. Then the asymptotic variance matrix of the GMM estimator using Ω^- as the weighting matrix is

$$\begin{aligned} V(\Omega^-) &= (G'\Omega^- G)^{-1} G'\Omega^-\Omega\Omega^- G(G'\Omega^- G)^{-1} \\ &= (G'\Omega^- G)^{-1}, \end{aligned}$$

where $G = \mathbf{E}[\frac{\partial g(Z_i,\beta_0)}{\partial \beta}]$. A natural question is whether $V(\Omega^-)$ is a constant matrix independent of the choice of Ω^-. The answer is yes. To see this, suppose the essential subvector of $g(Z_i, \beta_0)$ is $g^{(r)}(Z_i, \beta_0) = (g_{i_1}(Z_i, \beta_0), \ldots, g_{i_r}(Z_i, \beta_0))'$, with $g(Z_i, \beta_0) = Bg^{(r)}(Z_i, \beta_0)$ (*w.p.*1.). Let $G_1 = \mathbf{E}[\frac{\partial g^{(r)}(Z_i,\beta_0)}{\partial \beta'}]$ and $\Omega_1 = \mathbf{var}(g^{(r)}(Z_i, \beta_0))$. Then we have $G = BG_1$, $B'\Omega^- B = \Omega_1^{-1}$, hence

$$\begin{aligned} V(\Omega^-) &= (G'\Omega^- G)^{-1} = (G_1' B'\Omega^- BG_1)^{-1} \\ &= (G_1'\Omega_1^{-1} G_1)^{-1}, \end{aligned}$$

which is independent of the choice of Ω^-, as the essential vector and the corresponding matrices G_1 and Ω_1 are unrelated to Ω^-. We also see that $V(\Omega^-)$ remains the same if we use another essential vector, as $(G'\Omega^- G)^{-1} = (G'\Omega^+ G)^{-1}$ is unrelated to the choice of the essential vectors. More details can be found in the proof of Theorem 3.1 in the Appendix.

Judging from the asymptotic distributions we can see that GMM estimation using moment conditions (1) and the reflexive generalised inverses as weighting matrix is asymptotically equivalent to the efficient GMM using moment conditions $\mathbf{E}[g^{(r)}(Z_i, \beta_0)] = 0$. In some situations, one can figure out the essential subvector $g^{(r)}(Z_i, \beta)$, then efficient GMM estimation can be based on $\mathbf{E}[g^{(r)}(Z_i, \beta_0)] = 0$ directly. For instance, in the errors-in-variables analysis of panel data, Xiao et al. (2010a) and Xiao et al. (2010b) found that one can obtain $g^{(r)}(Z_i, \beta)$ by using singular value decomposition. However, such simple decomposition is not available in general and it can be very inconvenient, if not impossible, to find the essential vector $g^{(r)}(Z_i, \beta)$. Theorem 2 tells us that whatever this subvector is, the GMM estimator using any of the reflexive generalised inverses, and the Moore–Penrose generalised inverse in particular, as the weighting matrix will always have the same asymptotic variance as the efficient GMM based on $\mathbf{E}[g^{(r)}(Z_i, \beta_0)] = 0$.

4. Further issues

4.1. Optimal weighting matrix estimation

Now we discuss consistent estimation of Ω^+. Let $\tilde{\beta}$ be a consistent estimator of β_0 and $\hat{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n}[g(Z_i, \tilde{\beta}) g(Z_i, \tilde{\beta})']$. Then $\hat{\Omega}_n \to \Omega$ in probability under normal regularity conditions. A natural candidate estimator of Ω^+ is $\hat{\Omega}_n^+$.

It is well known that if a sequence of nonsingular square matrices $\{A_n\}$ converges to a nonsingular square matrix A, then $A_n^{-1} \to A^{-1}$.[4] However, if A is singular, and $A_n \to A$, we may not necessarily have that $A_n^+ \to A^+$.[5] Assuming $A_n \to A$ and A is singular, then a necessary and sufficient condition for $A_n^+ \to A^+$ is:

Theorem 4.1 (Stewart, 1969): *Let* $\{A_n\}$ *be a sequence of real* $m \times n$ *matrices converging to a* $m \times n$ *matrix* A*. Then* $A_n^+ \to A^+$ *if and only if rank*$(A_n) =$ *rank*(A) *for* n *large enough.*

We now prove that $\hat{\Omega}_n^+$ is a consistent estimator for Ω^+. By Theorem 4.1, we need only to show that $rank(\hat{\Omega}_n) = rank(\Omega)$ when n is large enough. By Lemma 3.1, there exists a constant matrix B of full column rank, such that w.p.1.,

$$\begin{aligned} \hat{\Omega}_n &= \frac{1}{n}\sum_{i=1}^{n}\left[Bg^{r}(Z_i, \tilde{\beta}) g^{(r)}(Z_i, \tilde{\beta})' B'\right] \\ &= B\left(\frac{1}{n}\sum_{i=1}^{n}\left[g^{(r)}(Z_i, \tilde{\beta}) g^{(r)}(Z_i, \tilde{\beta})'\right]\right) B'. \end{aligned}$$

Since the components of $g^{(r)}(Z_i, \beta)$ are linear independent for any β,

$$rank\left(\frac{1}{n}\sum_{i=1}^{n}\left[g^{(r)}(Z_i, \hat{\beta}) g^{(r)}(Z_i, \hat{\beta})'\right]\right) = r = rank(\Omega),$$

for any n. Therefore $rank(\hat{\Omega}_n) = rank\left(\frac{1}{n}\sum_{i=1}^{n}[g^{(r)}(Z_i, \hat{\beta}) g^{(r)}(Z_i, \hat{\beta})']\right) = rank(\Omega)$, and $\hat{\Omega}_n^+$ converges to Ω^+ in probability.

[4] A sequence of real $m \times n$ matrices $\{A_n\}$ is said to converge to a $m \times n$ matrix A if $\|A_n - A\| \to 0$, where $\|\cdot\|$ is a matrix norm, such as the Euclidean norm or $\|A\| = \sup_{\|x\|=1}\{\|Ax\|\}$.

[5] For example, consider $A_n = \begin{bmatrix} 1-\frac{1}{n} & 1-\frac{1}{n^2} \\ 1-\frac{1}{n^2} & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then $A_n \to A$. Since A_n is invertible, $A_n^+ = A_n^{-1} \to \begin{bmatrix} -\infty & \infty \\ \infty & -\infty \end{bmatrix}$. Hence $A_n^+ \nrightarrow A^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

Even though using generalised inverses is theoretically sound, it can be unstable, i.e., small perturbation of a singular matrix may result in large deviation from its generalised inverses.[6] Therefore, one must be cautious when using generalised inverses. We suggest that one should first try to find the essential subvector $g^{(r)}$. In case that $g^{(r)}$ is not easily obtainable, the method introduced below can be used as an alternative to generalised inverses.

4.2. Imposing random noises

To avoid the potential bias caused by generalised inverses, we can add randomly generated noises to the system to make it nonsingular, as Bierens (2007) and Lai (2008) did in the maximum likelihood estimation of singular system of equations. Specifically, let $U_1, \ldots, U_n$ be i.i.d. $K \times 1$ random vectors generated from the multivariate normal distribution with mean zero and covariance matrix $\sigma^2 I_K$, and assume that $U_1, \ldots, U_n$ are independent from $Z_1, \ldots, Z_n$. Define $h(Z_i, Ui, \beta) = g(Z_i, \beta) + U_i$, for $i = 1, \ldots, n$. Then β_0 is the solution of the set of moment conditions

$$\mathbf{E}\left[h(Z_i, U_i, \beta)\right] = 0. \tag{9}$$

The set of moment conditions (9) is nonsingular, since $\Sigma = \mathbf{E}[h(Z_i, U_i, \beta_0)h(Z_i, U_i, \beta_0)'] = \Omega + \sigma^2 I_K > 0$. Let $\tilde{\beta}_{GMM}$ be an efficient GMM estimator of β_0 based on (9), then the asymptotic distribution of $\tilde{\beta}_{GMM}$ is $\sqrt{n}(\tilde{\beta}_{GMM} - \beta_0) \xrightarrow{d} N(0, (G'\Sigma^{-1}G)^{-1})$. Since $(G'\Sigma^{-1}G)^{-1} > (G'\Omega^{-1}G)^{-1}$, for any $\sigma > 0$, $\tilde{\beta}_{GMM}$ is asymptotically less efficient than $\hat{\beta}_{GMM}$. However, the loss of efficiency can be controlled since $(G'\Sigma^{-1}G)^{-1} \to (G'\Omega^{-1}G)^{-1}$, as $\sigma \to 0$. Similar to Lai (2008), one can also generate m independent samples of $U_1, \ldots, U_n$, obtain m GMM estimators $\tilde{\beta}^1_{GMM}, \ldots, \tilde{\beta}^m_{GMM}$ and then construct a new estimator by $\tilde{\beta}^A_{GMM} = \frac{1}{m}\sum_{j=1}^{m} \tilde{\beta}^j_{GMM}$.[7] Since $\tilde{\beta}^A_{GMM}$ combines information in $\tilde{\beta}^1_{GMM}, \ldots, \tilde{\beta}^m_{GMM}$, in theory it is asymptotically more efficient than any of $\tilde{\beta}^1_{GMM}, \ldots, \tilde{\beta}^m_{GMM}$. It is of interest to investigate the asymptotic distribution and finite sample performance of $\tilde{\beta}^A_{GMM}$ in a future study.

5. Concluding remarks

Since the moment condition models do not require researchers to specify the likelihood function of the data generating process, they have been widely used by econometricians to model economic theories. Though it is desirable that the moment conditions constructed from economic theory are linearly independent, in practice this may not always be the case. Sometimes singularity is inherent in the model or is caused by some singular transformations. In this paper, we extended the efficient GMM estimation to linearly dependent moment condition models. The result can be viewed as a natural extension of the standard GMM theory, since the generalised inverse of a matrix is a natural extension of the inverse of a matrix. Though in theory using generalised inverses yields efficient GMM estimators, in practice one must be cautious of using them, in light of the following two concerns. First, using generalised inverses ignores the intrinsic structure of the moment conditions, which sometimes contains important information. Second, the generalised inverses of a singular matrix are unstable, which could induce serious bias of the resulting GMM estimator. Therefore when there is singularity in the system, a practical strategy is to obtain an essential moment vector and apply GMM to it. In case an essential moment vector is not available, we can add random noises to the moment conditions and obtain GMM estimators based on the new set of moment conditions. We suggest using generalised inverses with discretion. The results in this paper might also shed light on other popular statistical methods (such as the empirical likelihood) for estimating equations with singularity.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

The research is supported by the National Natural Science Foundation of China (NSFC grant: 71661137005, 71473040 and 11571081).

Notes on contributor

Dr. Zhiguo Xiao received his PhD in statistics from the University of Wisconsin-Madison. His research interest includes both theoretical and applied econometrics.

References

Abadir, K. M., & Magnus, J. R. (2005). *Matrix algebra*. New York: Cambridge University Press.
Alessi, L., Barigozzi, M., & Capasso, M. (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, *79*, 16–47.
Bapat, R. B. (2012). *Linear algebra and linear models* (3rd ed.). London: Springer-Verlag.
Barten, A. (1969). Maximum likelihood estimation of a complete system of demand equations. *European Economic Review*, *1*, 7–73.
Barten, A. (1977). The system of consumer demand function approach: A review. *Econometrica*, *45*, 23–51.

[6] For example, consider a matrix $A = \begin{bmatrix} 1 & & \\ & 0 & \\ & & 0 \end{bmatrix}$. Its Moore–Penrose generalised inverse is $A^+ = A$. Adding a small number $\varepsilon = 10^{-3}$ to the first two diagonal entries of A, we obtain a new Moore–Penrose generalised inverse $\begin{bmatrix} 0.999 & & \\ & 1000 & \\ & & 0 \end{bmatrix}$.

[7] This idea is similar to bootstrap aggregating (bagging).

Berndt, E. R., & Christensen, L. R. (1974). Testing for the existence of a consistent aggregate index of labor inputs. *American Economics Review*, *64*, 391–403.
Bierens, H. J. (2007). Econometric analysis of linearized singular dynamic stochastic general equilibrium models. *Journal of Econometrics*, *136*, 595–627.
Bierens, H. J., & Swanson, N. R. (2000). The econometric consequences of the ceteris paribus condition in economic theory. *Journal of Econometrics*, *95*, 223–253.
Biørn, E. (2000). Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences. *Econometric Reviews*, *19*(4), 391–424.
Biørn, E., & Klette, T. (1998). Panel data with errors-in-variables: Essential and redundant orthogonal conditions in GMM-estimation. *Economics Letters*, *59*, 275–282.
Chen, Y., Hong, C., & Riley, R. D. (2014). An alternative pseudolikelihood method for multivariate random-effects meta-analysis. *Statistics in Medicine*, *34*(3), 361–380.
Dhrymes, P. J. (1962). On devising unbiased estimators for the parameters of a Cobb–Douglas production function. *Econometrica*, *30*, 297–304.
Dhrymes, P. J., & Schwarz, S. (1987). On the existence of generalized inverse estimators in a singular system of equations. *Journal of Forecasting*, *6*, 181–192.
Fampa, M., & Lee, J. (2018). On sparse reflexive generalized inverses. *Operations Research Letters*, *46*(6), 605–610.
Hall, A. (2005). *Generalized method of moments*. Oxford University Press: New York.
Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.
Hansen, B., & West, K. (2002). Generalized method of moments and macroeconomics. *Journal of Business & Economic Statistics*, *20*, 460–469.
Haupt, H., & Oberhofer, W. (2006). Generalized adding-up in systems of regression equations. *Economics Letters*, *92*, 263–269.
Ireland, P. N. (2004). A method for taking models to the data. *Journal of Economic Dynamics and Control*, *28*, 1205–1226.
Kreijger, R. G., & Neudecker, H. (1977). Exact linear restrictions on parameters in the general linear model with a singular covariance matrix. *Journal of the American Statistical Association*, *72*, 430–432.
Lai, H. (2008). Maximum likelihood estimation of singular systems of equations. *Economics Letters*, *99*, 51–54.
Leeper, E. M., Walker, T. B., & Yang, S.-C. S. (2013). Fiscal foresight and information flows. *Econometrica*, *81*, 1115–1145.
Mountford, A., & Uhlig, H. (2009). What are the effects of fiscal policy shocks?. *Journal of Applied Econometrics*, *24*, 960–992.
Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, *51*, 406–413.
Rao, C. R. (1972). Alternative econometric models of sales advertising relationships. *Journal of Marketing Research*, *9*, 171–181.
Rao, C. R., & Mitra, S. K. (1971). *Generalized inverse of matrices and its applications*. John Wiley & Sons, Inc.: New York.
Riley, R., Abrams, K., Lambert, P., Sutton, A., & Thompson, J. (2007). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine*, *26*(1), 78–97.
Schneeweiss, H. (2014). The linear GMM model with singular covariance matrix due to the elimination of a nuisance parameter. Technical Report 165, Dept. Statistics, Univ. Munich.
Stewart, G. W. (1969). On the continuity of the generalized inverse. *SIAM Journal of Applied Mathematics*, *17*, 33–45.
Theil, H. (1971). *Principle of econometrics*. John Wiley & Sons: New York.
Velasco, C., & Lobato, I. N. (2018). Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models. *The Annals of Statistics*, *46*(2), 555–579.
Wansbeek, T. J. (2001). GMM Estimation in panel data models with measurement error. *Journal of Econometrics*, *104*, 259–268.
Weiss, D (1968). Determinants of market share. *Journal of Marketing Research*, *5*, 290–295.
White, H. (1986). Instrumental variables analogs of generalized least squares estimators. *Advances in Statistical Analysis and Statistical Computing*, *1*, 173–227.
Xiao, Z. (2008). Generalized inverses in GMM estimation with redundant moment conditions. In *Topics in Generalized Method of Moments Estimation with Application to Panel Data with Measurement Error*, University of Wisconsin-Madison PhD thesis.
Xiao, Z., Shao, J., & Palta, M. (2010a). GMM in linear regression for longitudinal data with multiple covariates measured with error. *Journal of Applied Statistics*, *37*, 791–805.
Xiao, Z., Shao, J., & Palta, M. (2010b). Instrumental variable and GMM estimation of panel data with measurement error. *Statistica Sinica*, *20*(4), 1725–1747.
Xiao, Z., Shao, J., Xu, R., & Palta, M. (2007). Efficiency of GMM estimation in panel data models with measurement error. *Sankhya: The Indian Journal of Statistics*, *69*, 101–118.
Xu, L., Fampa, M., & Lee, J. (2019). Aspects of symmetry for sparse reflexive generalized inverses. *arXiv*:1903.05744v1 [math.OC].

Appendix: Proofs of results

Proof of Lemma 1: Let $\Gamma = \mathbf{var}(Y)$, and $\Gamma = T\begin{pmatrix}\Lambda_r & 0\\ 0 & 0\end{pmatrix}T'$ be the spectrum decomposition of Γ, with $T'T = TT' = I_m$. Let $T = [T_1\ T_2]$, then

$$T'T = \begin{bmatrix}T_1'\\ T_2'\end{bmatrix}[T_1\ T_2] = \begin{bmatrix}T_1'T_1 & T_1'T_2\\ T_2'T_1 & T_2'T_2\end{bmatrix} = I_m,$$

$$\Gamma = T\begin{bmatrix}\Lambda_r & 0\\ 0 & 0\end{bmatrix}T' = T_1\Lambda_r T_1',$$

hence $T_2'T_1 = 0$, and $\mathbf{var}(T_2'Y) = T_2'VT_2 = T_2'T_1\Lambda_r T_1'T_2 = 0$. Hence there exists an $(m-r)\times 1$ constant vector c such $T_2'Y = c$. Let $T_2' = [\mathbf{t}_1 \ldots \mathbf{t}_m]$, with $rank(T_2') = q := m - r$. Suppose $\mathbf{t}_{j_1} \ldots \mathbf{t}_{j_q}$ are linearly independent, then

$$\begin{aligned}T_2'Y &= [\mathbf{t}_1 \cdots \mathbf{t}_m]\,[Y_1 \cdot Y_m]'\\ &= \mathbf{t}_1 Y_1 + \cdots + \mathbf{t}_m Y_m\\ &= \left[\mathbf{t}_{j_1} \ldots \mathbf{t}_{j_q}\right]\left[Y_{j_1} \cdot Y_{j_q}\right]' + \hat{T}Y^{-(q)},\end{aligned}$$

hence $[\mathbf{t}_{j_1} \ldots \mathbf{t}_{j_q}][Y_{j_1} \cdot Y_{j_q}]' + \hat{T}Y^{-(q)} = c$, i.e.,

$$Y^{(q)} = \left[Y_{j_1} \cdot Y_{j_q}\right]' = -\left[\mathbf{t}_{j_1} \ldots \mathbf{t}_{j_q}\right]^{-1}\hat{T}Y^{-(q)} + \left[\mathbf{t}_{j_1} \ldots \mathbf{t}_{j_q}\right]^{-1}c.$$

Let $C = -[\mathbf{t}_{j_1} \ldots \mathbf{t}_{j_q}]^{-1}\hat{T}$, $d = [\mathbf{t}_{j_1} \ldots \mathbf{t}_{j_q}]^{-1}c$, we get $Y^{(q)} = CY^{-(q)} + d$, i.e., $Y^{-(r)} = CY^{(r)} + d$. Since $Y^{(r)}$ and $Y^{-(r)}$ are subvectors of Y, there exists a $m\times m$ nonsingular matrix A

such that

$$Y = A\begin{bmatrix} Y^{(r)} \\ Y^{-(r)} \end{bmatrix} = A\begin{bmatrix} Y^{(r)} \\ CY^{(r)} + d \end{bmatrix} = A\begin{bmatrix} I \\ C \end{bmatrix} Y^r + A\begin{bmatrix} 0 \\ d \end{bmatrix},$$

i.e., $Y = BY^{(r)} + \widetilde{d}$, with $B = A\left[\begin{smallmatrix} I \\ C \end{smallmatrix}\right], \widetilde{d} = A\left[\begin{smallmatrix} 0 \\ d \end{smallmatrix}\right]$. Hence $\mathbf{var}(Y) = B'\mathbf{var}(Y^{(r)})B$. Since $B = A\left[\begin{smallmatrix} I \\ C \end{smallmatrix}\right]$ is of full column rank, $rank(\mathbf{var}(Y)) = rank(\mathbf{var}(Y^{(r)})) = r$. This shows that $\mathbf{var}(Y^{(r)})$ is nonsingular. ■

Proof of Theorem 3.1: Let $V(W)$ denote the asymptotic variance of the GMM estimator using weighting matrix W. Then $V(W) = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$. Let Ω^- be a reflexive generalised inverse of Ω. Then we have

$$\begin{aligned} V(\Omega^-) &= (G'\Omega^- G)^{-1}G'\Omega^-\Omega\Omega^- G(G'\Omega^- G)^{-1} \\ &= (G'\Omega^- G)^{-1}. \end{aligned}$$

Hence $V(W) - V(\Omega^-) = (G'WG)^{-1}G'W[\Omega - G(G'\Omega^- G)^{-1}G']WG(G'WG)^{-1}$. To establish $V(W) - V(\Omega^-) \geq 0$, we just need to show that $\Omega - G(G'\Omega^- G)^{-1}G' \geq 0$. Let $rank(\Omega) = r$. By Lemma 3.1, there exist a subvector $g^{(r)}(Z_i, \beta_0) = (g_{i_1}(Z_i, \beta_0), \ldots, g_{i_r}(Z_i, \beta_0))'$ and a matrix B of full column rank such that $g(Z_i, \beta_0) = Bg^r(Z_i, \beta_0)$ a.s., with $\Omega_1 = \mathbf{var}(g^{(r)}(Z_i, \beta_0))$ positive definite. Then $\Omega = \mathbf{var}(g(Z_i, \beta_0)) = B\Omega_1 B'$, and $G = \mathbf{E}[\frac{\partial g(Z_i,\beta_0)}{\partial \beta}] = BG_1$, with $G_1 = \mathbf{E}[\frac{\partial g^{(r)}(Z_i,\beta_0)}{\partial \beta'}]$. Hence

$$\begin{aligned} \Omega - G(G'\Omega^- G)^{-1}G' &= B\Omega_1 B' - BG_1(G'\Omega^- G)^{-1}G_1'B' \\ &= B\left[\Omega_1 - G_1(G'\Omega^- G)^{-1}G_1'\right]B'. \end{aligned}$$

So we just need to show that $\Omega_1 - G_1(G'\Omega^- G)^{-1}G_1' \geq 0$. By Proposition 2.1, $B'\Omega^- B = \Omega_1^{-1}$, hence

$$\begin{aligned} &\Omega_1 - G_1(G'\Omega^- G)^{-1}G_1' \\ &= \Omega_1 - G_1(G_1'B'\Omega^- BG_1)^{-1}G_1' \\ &= \Omega_1 - G_1(G_1'\Omega_1^{-1}G_1)^{-1}G_1' \\ &= \Omega_1^{\frac{1}{2}}\left[I - \Omega_1^{-\frac{1}{2}}G_1(G_1'\Omega_1^{-1}G_1)^{-1}G_1'\Omega_1^{-\frac{1}{2}}\right]\Omega_1^{\frac{1}{2}} \\ &\geq 0, \end{aligned}$$

since $I - \Omega_1^{-\frac{1}{2}}G_1(G_1'\Omega_1^{-1}G_1)^{-1}G_1'\Omega_1^{-\frac{1}{2}}$ is idempotent and symmetric. ■