

# 16 Generalised Method of Moments.

## 16.1 GMM setup

The basic principle of the method of moments is to choose the parameter estimate so that the corresponding sample moments are equal to the corresponding population moments. Moment equation models take the following form. Let  $\mathbf{g}_i(\boldsymbol{\beta})$  be a known  $\ell \times 1$  function of the  $i^{th}$  observation, and a  $k \times 1$  parameter  $\boldsymbol{\beta}$ . Denote the true parameter value in population as  $\boldsymbol{\beta}_0$ . A moment equation model is summarised by the moment equations

$$\mathbb{E}[\mathbf{g}_i(\boldsymbol{\beta}_0)] = 0.$$

**Example. Mean:**  $g_i(\mu) = y_i - \mu$ .

**Instrumental Variables:**  $\mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{w}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})$ . (Exogeneity of  $w_i$ )

**Linear Regression:**  $\mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{x}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})$ . (uncorrelatedness of  $x_i$  and  $u_i$ )

When  $\ell = k$ , the moment equations are just identified (k equations for k unknowns), and we solve the moment equations for  $\hat{\boldsymbol{\beta}}$

$$\bar{g}_n = \frac{1}{n} \sum_{i=1}^n g_i(\hat{\boldsymbol{\beta}}) = 0.$$

When  $\ell < k$  the system is under-identified we typically have no solution to the system, and cannot hope to solve for  $\hat{\boldsymbol{\beta}}$ .

When  $\ell > k$  the system is over-identified there are an infinite number of solutions to the system (more variables than unknowns), each one giving us a different MOM estimate of  $\boldsymbol{\beta}$ . 2SLS solved this problem by using the k dimensional vector  $\hat{\pi}w_i$  obtained from the first stage regression. Alternatively, GMM aims to minimise the distance between the sample and population moments. That is,

### Definition 16.1.1: GMM estimator

$$\hat{\boldsymbol{\beta}}_{GMM} = \arg \min_{\boldsymbol{\beta}} \left( \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) \right) = \arg \min_{\boldsymbol{\beta}} \bar{g}_n' W_n \bar{g}_n.$$

where  $W_n$  is a positive-definite  $\ell \times \ell$  weighting matrix.

The object we are minimising is called the GMM criterion function, further denoted as  $J_n(\boldsymbol{\beta})$ . If the criterion function is unweighted:  $W_n = I_\ell$ , then  $J_n(\boldsymbol{\beta}) = \bar{g}_n(\boldsymbol{\beta})' \bar{g}_n(\boldsymbol{\beta}) = \|\bar{g}_n(\boldsymbol{\beta})\|^2$ , the square of the Euclidean length. Since we restrict attention to positive definite weight matrices  $W$ , the criterion  $J_n$  is always non-negative.

#### Note:-

In the just identified case, the GMM estimator is the same as the MOM estimator, regardless of the choice of  $W_n$ .  $\hat{\boldsymbol{\beta}}_{MOM}$  solves  $\bar{g}_n(\boldsymbol{\beta}_{MOM}) = 0$ , hence  $J_n(\boldsymbol{\beta}_{MOM}) = 0$ . Since  $J_n(\boldsymbol{\beta}) \geq 0$ ,  $\hat{\boldsymbol{\beta}}_{MOM}$  is also the GMM estimator.

GMM is known as an extremum estimator, since it is obtained by minimising a criterion function. Another example of an extremum estimator is MLE.

**Example (Overidentified IV).** Given the moment conditions  $\mathbb{E}[z_i(y_i - x_i'\beta)] = 0$  where  $z_i$  is  $\ell \times 1$  and  $x_i$  is  $k \times 1$ , we can construct the GMM criterion function as follows:

$$J_n(\beta) = (Z'y - Z'X\beta)'W_n(Z'y - Z'X\beta)$$

The GMM estimator minimises this criterion function, the FOCs are:

$$\begin{aligned} 0 &= \frac{\partial J_n(\beta)}{\partial \beta} \\ &= \frac{\partial}{\partial \beta} (y'ZW_nZ'y - y'ZW_nZ'X\beta - \beta'X'ZW_nZ'y + \beta'X'ZW_nZ'X\beta) \\ &= -2X'ZW_nZ'y + 2X'ZW_nZ'X\beta \\ \Rightarrow \hat{\beta}_{GMM} &= (X'ZW_nZ'X)^{-1}X'ZW_nZ'y \end{aligned}$$

While the estimator depends on  $W_n$ , the dependence is only up to scale. For example, if we multiply  $W_n$  by a constant  $c > 0$ , then  $\hat{\beta}_{GMM}$  is unchanged. When  $W$  is fixed by the user we call  $\hat{\beta}_{GMM}$  the **one-step GMM estimator**. The GMM estimator is similar to the 2SLS estimator, in fact they are equal when  $W_n = (Z'Z)^{-1}$ .

**Theorem 16.1.1.** If  $W_n = (Z'Z)^{-1}$ , then  $\hat{\beta}_{GMM} = \hat{\beta}_{2SLS}$ . Furthermore if  $k = \ell$ , then  $\hat{\beta}_{GMM} = \hat{\beta}_{IV}$ .

**Example (CAPM).** Consider the case where a representative agent makes decisions about consumption and investment to maximise lifetime discounted expected utility conditional on their information set. Suppose the agent has CRRA utility

$$\mathbb{E} \left( \sum_{i=1}^{\infty} \delta_0^i U(c_{t+i} | \mathcal{I}_t) \right), \quad U(c_t) := \frac{c_t^{\gamma_0} - 1}{\gamma_0}$$

The budget constraint is

$$c_t + p_t q_t = r_t q_{t-1} + w_t$$

where  $p_t$  is the price of a financial asset,  $q_t$  is the amount of asset owned at time  $t$ ,  $r_t$  is the gross return to the asset, and  $w_t$  is the wage. The FOC (euler equation) for this problem is:

$$p_t c_t^{\gamma_0 - 1} = \mathbb{E}_t \left( \delta_0 r_{t+1} c_{t+1}^{\gamma_0 - 1} | \mathcal{I}_t \right)$$

We thus have the conditional moment condition:

$$\mathbb{E}_t \left( \delta_0 \frac{r_{t+1}}{p_t} \frac{c_{t+1}^{\gamma_0 - 1}}{c_t^{\gamma_0 - 1}} | \mathcal{I}_t \right) = 0$$

This implies an infinite number of unconditional moment conditions, we select the following (arbitrarily):

$$\mathbb{E}_t \left( z_t \left[ \delta_0 \frac{r_{t+1}}{p_t} \frac{c_{t+1}^{\gamma_0 - 1}}{c_t^{\gamma_0 - 1}} \right] \right) = 0$$

Where  $z$  is any vector of variables known by time  $t$  (EG consumption growth, returns in recent periods etc.) This is  $\dim(z_t)$  equations for 2 unknowns  $\gamma_0$  and  $\delta_0$ . We can estimate this model using GMM.

**Note:-**

A conditional moment condition is much stronger than an unconditional condition, the conditional condition implies an infinite set of unconditional moments:

$$\begin{aligned}\mathbb{E}[f(z_i)g(x_i, \beta_0)] &= \mathbb{E}[\mathbb{E}[f(z_i)g(x_i, \beta_0)|z_i]] \\ &= \mathbb{E}[f(z_i)\mathbb{E}[g(x_i, \beta_0)|z_i]] \\ &= 0\end{aligned}$$

above we arbitrarily selected  $f(z_i) = z_i$ , but any function of any subset of variables in  $z_i$  will do. See [here](#) for a brief discussion of the optimal choice, however this is beyond the scope of this course.

## 16.2 Asymptotic properties of GMM

- (GMM0)  $\frac{1}{n} \sum_{i=1}^n g_i(\beta) \xrightarrow{p} \mathbb{E}g_i(\beta)$  and  $\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta'} g_i(\beta) \xrightarrow{p} \mathbb{E} \frac{\partial}{\partial \beta'} g_i(\beta)$ , uniformly over  $\beta$  from a compact subset  $B$  of  $\mathbb{R}^k$ .
- (GMM1)  $\mathbb{E} \frac{\partial}{\partial \beta'} g_i(\beta)$  has rank  $k$  (i.e. full column rank) in a neighbourhood of  $\beta_0$ . Furthermore  $W_n \xrightarrow{p} W$ , a positive definite matrix.
- (GMM2)  $\mathbb{E}g_i(\beta_0) = 0$  and  $\beta_0$  is the only root of equation  $\mathbb{E}g_i(\beta) = 0$  for  $\beta \in B$ .
- (GMM3)  $\frac{1}{n} \sum_{i=1}^n g_i(\beta)g_i(\beta)' \xrightarrow{p} V(\beta)$  uniformly  $\beta \in B$  and  $\frac{1}{\sqrt{n}} \frac{1}{n} \sum_{i=1}^n g_i(\beta_0) \xrightarrow{d} N(0, V(\beta_0))$  with  $V(\beta_0) > 0$ .
- (GMM4)  $\mathbb{E}g_i(\beta)$  and  $g_i(\beta)$  are continuously differentiable, and  $\frac{\partial}{\partial \beta'} \mathbb{E}g_i(\beta) = \mathbb{E} \frac{\partial}{\partial \beta'} g_i(\beta)$ .

**GMM0** does not require iid-ness of the data, it just says that an LLN should hold. As discussed with serial correlation, there exist LLNs for dependent data.

**Theorem 16.2.1.** Under (GMM0)-(GMM4)

1.  $\hat{\beta}_{GMM} \xrightarrow{p} \beta_0$
2.  $\sqrt{n}(\hat{\beta}_{GMM} - \beta_0) \xrightarrow{d} N(0, (D_0' W D_0)^{-1} D_0' W V_0 W D_0 (D_0' W D_0)^{-1})$  where  $D_0 = \mathbb{E} \frac{\partial}{\partial \beta'} g_i(\beta_0)$  and  $V_0 = V(\beta_0)$  is the variance of the asymptotic distribution of .

**Proof.** 1. By GMM0 and GMM1  $W_n \rightarrow W > 0$  and  $\frac{1}{n} \sum_{i=1}^n g_i(\beta) \rightarrow \mathbb{E}g_i(\beta)$  in probability, then

$$\begin{aligned}J_n(\beta) &= \left( \frac{1}{n} \sum_{i=1}^n g_i(\beta) \right)' W_n \left( \sum_{i=1}^n g_i(\beta) \right) \\ &\xrightarrow{p} \mathbb{E}g_i(\beta)' W \mathbb{E}g_i(\beta) \geq 0\end{aligned}$$

Using GMM2, we have that  $\beta_0$  is the only root of  $\mathbb{E}g_i(\beta) = 0$ , hence  $\beta_0$  is the minimiser of  $J_n(\beta)$ , and  $\hat{\beta}_{GMM} \xrightarrow{p} \beta_0$ .

2. Given that  $\hat{\beta}_{GMM} \xrightarrow{p} \beta_0$ , and that  $\frac{\partial}{\partial \beta} g_i$  is differentiable, we can compute the taylor expansion as follows:

$$\frac{\partial}{\partial \beta} J_n(\hat{\beta}_{GMM}) \approx \frac{\partial}{\partial \beta} J_n(\beta_0) + \frac{\partial^2}{\partial \beta \partial \beta'} J_n(\hat{\beta}_{GMM})(\hat{\beta}_{GMM} - \beta_0)$$

Since the LHS = 0 by definition of  $\hat{\beta}_{GMM}$ , we have that

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta_0) \approx - \left( \frac{\partial^2}{\partial \beta \partial \beta'} J_n(\hat{\beta}_{GMM}) \right)^{-1} \left( \sqrt{n} \frac{\partial}{\partial \beta} J_n(\beta_0) \right)$$

Consider the first term in the above expression, we can write it as follows:

$$\begin{aligned} \frac{\partial^2}{\partial \beta \partial \beta'} J_n(\hat{\beta}_{GMM}) &= \frac{\partial^2}{\partial \beta \partial \beta'} \left( \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{GMM}) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{GMM}) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \beta \partial \beta'} g_i(\hat{\beta}_{GMM})' W_n g_i(\hat{\beta}_{GMM}) + \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{GMM})' W_n \frac{\partial^2}{\partial \beta \partial \beta'} g_i(\hat{\beta}_{GMM}) \\ &\quad + 2 \sum_{j,r=1}^l \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \beta \partial \beta'} g_{ij}(\beta_0) \right) W_{n,jr} \left( \frac{1}{n} \sum_{i=1}^n g_{ir}(\beta_0) \right) \\ &= 2 \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta'} g_i(\hat{\beta}_{GMM}) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta'} g_i(\hat{\beta}_{GMM}) \right) + o_p(1) \\ &\xrightarrow{d} 2 \left( \mathbb{E} \frac{\partial}{\partial \beta'} g_i(\beta_0) \right)' W \left( \mathbb{E} \frac{\partial}{\partial \beta'} g_i(\beta_0) \right) \\ &:= 2D_0' W D_0 \end{aligned}$$

$g_{ij}$  and  $g_{ir}$  are the  $j^{th}$  and  $r^{th}$  elements of  $g_i$  respectively, and  $W_{n,jr}$  is the element in the  $j$ -th row and  $r$ -th column of  $W_n$ . The cross term is  $o_p(1)$  since the final term converges to  $\mathbb{E} g_{ir}(\beta_0) = 0$  by GMM2.

Consider the final term in the above expression, we can write it as follows:

$$\begin{aligned} \sqrt{n} \frac{\partial}{\partial \beta} J_n(\beta_0) &= \sqrt{n} \frac{\partial}{\partial \beta} \left( \frac{1}{n} \sum_{i=1}^n g_i(\beta_0) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n g_i(\beta_0) \right) \\ &= 2 \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta'} g_i(\beta_0) \right)' W_n \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0) \right) \\ &\xrightarrow{d} 2 \left( \mathbb{E} \frac{\partial}{\partial \beta'} g_i(\beta_0) \right)' W (N(0, \mathbb{E}[g_i(\beta_0)g_i(\beta_0)'])) \\ &:= 2D_0' W N(0, V_0) \end{aligned}$$

Combining these two results, we have that

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{GMM} - \beta_0) &\xrightarrow{d} (2D_0' W D_0)^{-1} 2D_0' W N(0, V_0) \\ &\sim N(0, (D_0' W D_0)^{-1} D_0' W V_0 W D_0 (D_0' W D_0)^{-1}) \end{aligned}$$

□

## 16.3 Efficient GMM

The question remains as to how we pick  $W$ , and thus  $W_n$ . A simple choice is the identity matrix, i.e.  $W_n = I_\ell$ , or another fixed positive definite matrix. However these choices result in a loss of efficiency.

**Theorem 16.3.1.** The efficient choice of weighting matrix (that minimises asymptotic vari-

ance) is

$$W = V_0^{-1}$$

where  $V_0$  is the asymptotic variance of  $\sqrt{n}\bar{g}_n(\beta_0)$ .

Intuitively  $\hat{\beta}_{GMM}$  is trying to set each component of the  $\ell$  dimensional vector  $\frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{GMM})$  as close to zero as possible. However, if some components have particularly large variance (and are thus less informative), then it makes sense to place less weight on them.

**Proof.** If  $W = V_0^{-1}$ , then we can simplify the asymptotic variance as follows:

$$\begin{aligned} (D_0' W D_0)^{-1} D_0' W V_0 W D_0 (D_0' W D_0)^{-1} &= (D_0' V_0^{-1} D_0)^{-1} D_0' V_0^{-1} V_0 V_0^{-1} D_0 (D_0' V_0^{-1} D_0)^{-1} \\ &= (D_0' V_0^{-1} D_0)^{-1} D_0' V_0^{-1} D_0 (D_0' V_0^{-1} D_0)^{-1} \\ &= (D_0' V_0^{-1} D_0)^{-1} \end{aligned}$$

It remains to show that the difference between the inefficient and efficient weighting matrices  $(D_0' W D_0)^{-1} D_0' W V_0 W D_0 (D_0' W D_0)^{-1} - (D_0' V_0^{-1} D_0)^{-1}$  is positive semi-definite. For my first trick, I will pull the following matrix out of my ass:  $A = I_\ell - V_0^{-1/2} D_0 (D_0' V_0^{-1} D_0)^{-1} D_0' V_0^{-1/2}$ . Observe:

$$\begin{aligned} &(D_0' W D_0)^{-1} D_0' W V_0^{1/2} A V_0^{1/2} W D_0 (D_0' W D_0)^{-1} \\ &= (D_0' W D_0)^{-1} D_0' W V_0^{1/2} \left( I_\ell - V_0^{-1/2} D_0 (D_0' V_0^{-1} D_0)^{-1} D_0' V_0^{-1/2} \right) V_0^{1/2} W D_0 (D_0' W D_0)^{-1} \\ &= (D_0' W D_0)^{-1} D_0' W V_0^{1/2} V_0^{1/2} W D_0 (D_0' W D_0)^{-1} \\ &\quad - (D_0' W D_0)^{-1} D_0' W V_0^{1/2} V_0^{-1/2} D_0 (D_0' V_0^{-1} D_0)^{-1} D_0' V_0^{-1/2} V_0^{1/2} W D_0 (D_0' W D_0)^{-1} \\ &= (D_0' W D_0)^{-1} D_0' W V_0 W D_0 (D_0' W D_0)^{-1} \\ &\quad - \cancel{(D_0' W D_0)^{-1} D_0' W D_0 (D_0' V_0^{-1} D_0)^{-1} D_0' W D_0 (D_0' W D_0)^{-1}} \\ &= (D_0' W D_0)^{-1} D_0' W V_0 W D_0 (D_0' W D_0)^{-1} - (D_0' V_0^{-1} D_0)^{-1} \end{aligned}$$

Note that  $A$  is the residual maker matrix onto the space spanned by  $D_0 V^{-1/2}$ , is symmetric and idempotent, and thus positive semi-definite. The difference therefore has the form  $B' A B = B' A' A B = (A B)' (A B)$  and is clearly positive s.d. .  $\square$

### Estimation of $V_0$

We can estimate  $V_0$  and then estimate GMM in a two step procedure. The first step uses a consistent (but not necessarily efficient) estimate of  $\beta$  to construct the optimal weighting matrix  $W_n$ . The second step uses this weighting matrix  $W_n = \hat{V}_0^{-1}$  to estimate  $\beta$  efficiently. This is known as the two-step GMM estimator.

In the case of i.i.d.  $g_i(\beta)$ , we can estimate  $\hat{\beta}$  consistently with a sub optimal weighting matrix and compute

$$\begin{aligned} \hat{V}_0 &= \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}) g_i(\hat{\beta})' \quad \text{or} \\ \hat{V}_0 &= \frac{1}{n} \sum_{i=1}^n \left( g_i(\hat{\beta}) - \bar{g}_n(\hat{\beta}) \right) \left( g_i(\hat{\beta}) - \bar{g}_n(\hat{\beta}) \right)' \end{aligned}$$

If the data are serially correlated, a version of the Newey-West estimator can be used to estimate  $V_0$ .

$$\hat{\Gamma}_j = \frac{1}{n} \sum_{i=j+1}^n g_i(\hat{\beta}) g_{i-j}(\hat{\beta})' \quad \text{and obtain}$$

$$\hat{V}_0 = \frac{1}{n} \sum_{j=0}^G \frac{G+1-j}{G+1} (\hat{\Gamma}_j + \hat{\Gamma}_j')$$

**Example (2SLS under homoskedasticity).** Consider linear regression  $y_i = x_i' \beta_0 + \varepsilon_i$  with  $k$  endogenous variables and  $\ell > k$  instruments. The moment conditions  $g_i(\beta) = z_i(y_i - x_i' \beta)$  are distributed as:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \varepsilon_i \rightarrow N(0, \sigma^2 \mathbb{E}[z_i z_i'])$$

under IV3 (homoskedasticity). Hence  $V_0 = \sigma^2 \mathbb{E}[z_i z_i'] = \sigma^2 Z'Z$ . The efficient weighting matrix is therefore  $W_n = (Z'Z)^{-1}$ , since the weighting matrix is invariant to scale (so we can drop the  $\sigma^2$ ). From the previous example we know that the GMM estimator is given by  $\hat{\beta}_{GMM} = (X'Z W_n Z'X)^{-1} X'Z W_n Z'y$ , substituting in the efficient weighting matrix we obtain:

$$\hat{\beta}_{GMM} = (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'y = (X'P_Z X)^{-1} X'P_Z y = \hat{\beta}_{2SLS}$$

and thus, 2SLS can be viewed as efficient GMM under homoskedasticity.

However, under heteroskedasticity, 2SLS is inefficient relative to GMM which would instead be based on :

$$W_n = \left( \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 z_i z_i' \right)^{-1} \quad \text{or its centred version}$$

$$W_n = \left( \frac{1}{n} \sum_{i=1}^n \left( \hat{\varepsilon}_i z_i - \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i z_i \right) \left( \hat{\varepsilon}_i z_i - \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i z_i \right)' \right)^{-1}$$

$\hat{\varepsilon}$  can be obtained using any consistent (but inefficient) method, like 2SLS or unweighted GMM. If the model is correctly specified such that  $\mathbb{E} z_i \varepsilon_i = 0$ , then the centred and un-centred versions of  $W_n$  are asymptotically equivalent.

### Overidentification test

Previously we studied Sargan's test of over-identifying restrictions in the context of 2SLS under homo. This can be generalised to GMM (allowing for serial correlation and heteroskedasticity). In the overidentified case we have  $\ell$  equations

$$\mathbb{E}(g_i(\beta_0)) = 0$$

that have to be satisfied by  $k < \ell$  parameters. Since  $\frac{1}{n} \sum_{i=1}^n g_i(\beta_0) \xrightarrow{p} \mathbb{E} g_i(\beta_0)$  and  $\hat{\beta}_{GMM} \xrightarrow{p} \beta_0$  we can use  $\frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{GMM})$  to test the hypothesis:

$$H_0 : \mathbb{E}(g_i(\beta_0)) = 0$$

Assuming the efficient weight matrix is being used, the criterion function is

$$J(\hat{\beta}_{GMM}) = n \bar{g}_n' \hat{V}_0^{-1} \bar{g}_n = \left( \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{GMM}) \right)' \hat{V}_0^{-1} \left( \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{GMM}) \right)$$

**Theorem 16.3.2.** Under assumptions GMM0-4, as  $n \rightarrow \infty$ ,

$$J(\hat{\beta}_{GMM}) \xrightarrow{d} \chi^2(\ell - k)$$

**Proof.** This was not presented in the course, WORK IN PROGRESS.

For any positive semi-definite matrix  $A$  we can decompose it as  $A = U\Lambda U' = (U\Lambda^{1/2})(U\Lambda^{1/2})' = BB'$ . The covariance matrix  $V_0$  is positive definite, thus  $\Lambda$  is a diagonal matrix with positive entries. The inverse of  $V_0$  is thus also positive definite and can be written as  $V_0^{-1} = CC'$  (and  $V_0 = C'^{-1}C^{-1}$ ).

Subbing this into  $J$ :

$$\begin{aligned} J(\hat{\beta}_{GMM}) &= n(\bar{g}_n' \hat{V}_0^{-1} \bar{g}_n) \\ &= n(\bar{g}_n' CC' \bar{g}_n) \\ &= n(C' \bar{g}_n(\hat{\beta}))' C' \bar{g}_n(\hat{\beta}) \\ &= n(C' \bar{g}_n(\hat{\beta}))' (C' \hat{V}_0 C)^{-1} C' \bar{g}_n(\hat{\beta}) \end{aligned}$$

Where the last step follows from the fact that  $C' \hat{V}_0 C = I_\ell$ .

We can write the average of the moment conditions as

$$\bar{g}_n(\beta) = \frac{1}{n} Z' e$$

For my next trick I will pull the following matrix out of my ass:

$$D_n = I_\ell - C' \left( \frac{1}{n} Z' X \right) \left( \left( \frac{1}{n} X' Z \right) \hat{V}_0^{-1} \left( \frac{1}{n} Z' X \right) \right)^{-1} \left( \frac{1}{n} X' Z \right) \hat{V}_0^{-1} C'^{-1}$$

and observe that

$$\begin{aligned} D_n C' \bar{g}_n(\beta) &= I_\ell C' \bar{g}_n - C' \left( \frac{1}{n} Z' X \right) \left( \left( \frac{1}{n} X' Z \right) \hat{V}_0^{-1} \left( \frac{1}{n} Z' X \right) \right)^{-1} \left( \frac{1}{n} X' Z \right) \hat{V}_0^{-1} C'^{-1} C' \bar{g}_n \\ &= \dots \\ &= C' \bar{g}_n(\hat{\beta}) \end{aligned}$$

$D_n$  is a residual maker matrix onto the space spanned by  $C' Z' X$ , it converges to the following  $D_n \xrightarrow{p} I_\ell - R(R'R)^{-1}R'$  where  $R = C' \mathbb{E}(z_i x_i')$ .

Further note the asymptotic distribution of  $C' \bar{g}_n(\beta)$

$$\begin{aligned} n^{1/2} C' \bar{g}_n(\beta) &= C' \frac{1}{\sqrt{n}} \sum z_i e_i \\ &\xrightarrow{d} C' N(0, V_0) \\ &\sim N(0, C' V_0 C) \\ &\sim N(0, I_\ell) \end{aligned}$$

Thus (with some abuse of notation):

$$\begin{aligned}
J(\hat{\beta}_{GMM}) &= n(C'\bar{g}_n(\hat{\beta}))'(C'\hat{V}_0C)^{-1}C'\bar{g}_n(\hat{\beta}) \\
&= (D_n\sqrt{n}C'\bar{g}_n(\beta))'(C'\hat{V}_0C)^{-1}D_n\sqrt{n}C'\bar{g}_n(\beta) \\
&= (D_n\sqrt{n}C'\bar{g}_n(\beta))'(C'\hat{V}_0C)^{-1}D_n\sqrt{n}C'\bar{g}_n(\beta) \xrightarrow{d} (DN(0, I_\ell))'(\cancel{C'V_0C})^{-1}DN(0, I_\ell) \\
&= (DN(0, I_\ell))'DN(0, I_\ell)
\end{aligned}$$

$D$  is a residual maker matrix with rank equal to the number of non zero eigenvalues (i.e.  $\ell - k$ ), thus  $DN(0, I_\ell)$  is a  $\ell - k$  dimensional vector of independent standard normal random variables (same logic as deriving distribution of error variance in OLS). Thus the asymptotic distribution of  $J(\hat{\beta}_{GMM})$  is  $\chi^2(\ell - k)$ .  $\square$