

Chapter 13

Generalized Method of Moments

13.1 Introduction

One of the most popular estimation methods in applied econometrics is the Generalized Method of Moments (GMM). GMM generalizes classical method of moments by allowing for more equations than unknown parameters (so are overidentified) and by allowing general nonlinear functions of the observations and parameters. Together this allows for a fairly rich and flexible estimation framework. GMM includes as special cases OLS, IV, multivariate regression, and 2SLS. It includes both linear and nonlinear models. In this chapter we focus primarily on linear models.

The GMM label and methods were introduced to econometrics in a seminal paper by Lars Hansen (1982). The ideas and methods build on the work of Amemiya (1974, 1977), Gallant (1977), and Gallant and Jorgenson (1979). The ideas are closely related to the contemporaneous work of Halbert White (1980, 1982) and White and Domowitz (1984). The methods are also related to what are called **estimating equations** in the statistics literature. For a review of the latter see Godambe (1991).

13.2 Moment Equation Models

All of the models that have been introduced so far can be written as **moment equation models** where the population parameters solve a system of moment equations. Moment equation models are broader than the models so far considered and understanding their common structure opens up straightforward techniques to handle new econometric models.

Moment equation models take the following form. Let $g_i(\beta)$ be a known $\ell \times 1$ function of the i^{th} observation and a $k \times 1$ parameter β . A moment equation model is summarized by the moment equations

$$\mathbb{E}[g_i(\beta)] = 0 \tag{13.1}$$

and a parameter space $\beta \in B$. For example, in the instrumental variables model $g_i(\beta) = Z_i(Y_i - X_i'\beta)$.

In general, we say that a parameter β is **identified** if there is a unique mapping from the data distribution to β . In the context of the model (13.1) this means that there is a unique β satisfying (13.1). Since (13.1) is a system of ℓ equations with k unknowns, then it is necessary that $\ell \geq k$ for there to be a unique solution. If $\ell = k$ we say that the model is **just identified**, meaning that there is just enough information to identify the parameters. If $\ell > k$ we say that the model is **overidentified**, meaning that there is excess information. If $\ell < k$ we say that the model is **underidentified**, meaning that there is insufficient information to identify the parameters. In general, we assume that $\ell \geq k$ so the model is either just identified or overidentified.

13.3 Method of Moments Estimators

In this section we consider the just-identified case $\ell = k$.

Define the sample analog of (13.5)

$$\bar{g}_n(\beta) = \frac{1}{n} \sum_{i=1}^n g_i(\beta). \quad (13.2)$$

The **method of moments estimator (MME)** $\hat{\beta}_{\text{mm}}$ is the parameter value which sets $\bar{g}_n(\beta) = 0$. Thus

$$\bar{g}_n(\hat{\beta}_{\text{mm}}) = \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}_{\text{mm}}) = 0. \quad (13.3)$$

The equations (13.3) are known as the **estimating equations** as they are the equations which determine the estimator $\hat{\beta}_{\text{mm}}$.

In some contexts (such as those discussed in the examples below) there is an explicit solution for $\hat{\beta}_{\text{mm}}$. In other cases the solution must be found numerically.

We now show how most of the estimators discussed so far in the textbook can be written as method of moments estimators.

Mean: Set $g_i(\mu) = Y_i - \mu$. The MME is $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i$.

Mean and Variance: Set

$$g_i(\mu, \sigma^2) = \begin{pmatrix} Y_i - \mu \\ (Y_i - \mu)^2 - \sigma^2 \end{pmatrix}.$$

The MME are $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\mu})^2$.

OLS: Set $g_i(\beta) = X_i(Y_i - X_i'\beta)$. The MME is $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$.

OLS and Variance: Set

$$g_i(\beta, \sigma^2) = \begin{pmatrix} X_i(Y_i - X_i'\beta) \\ (Y_i - X_i'\beta)^2 - \sigma^2 \end{pmatrix}.$$

The MME is $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i'\hat{\beta})^2$.

Multivariate Least Squares, vector form: Set $g_i(\beta) = \bar{X}_i'(Y_i - \bar{X}_i\beta)$. The MME is $\hat{\beta} = \left(\sum_{i=1}^n \bar{X}_i'\bar{X}_i\right)^{-1} \left(\sum_{i=1}^n \bar{X}_i'Y_i\right)$ which is (11.4).

Multivariate Least Squares, matrix form: Set $g_i(\mathbf{B}) = \text{vec}(X_i(Y_i' - X_i'\mathbf{B}))$. The MME is $\hat{\mathbf{B}} = \left(\sum_{i=1}^n X_iX_i'\right)^{-1} \left(\sum_{i=1}^n X_iY_i'\right)$ which is (11.6).

Seemingly Unrelated Regression: Set

$$g_i(\beta, \Sigma) = \begin{pmatrix} \bar{X}_i\Sigma^{-1}(Y_i - \bar{X}_i'\beta) \\ \text{vec}\left(\Sigma - (Y_i - \bar{X}_i'\beta)(Y_i - \bar{X}_i'\beta)'\right) \end{pmatrix}.$$

The MME is $\hat{\beta} = \left(\sum_{i=1}^n \bar{X}_i\hat{\Sigma}^{-1}\bar{X}_i'\right)^{-1} \left(\sum_{i=1}^n \bar{X}_i\hat{\Sigma}^{-1}Y_i\right)$ and $\hat{\Sigma} = n^{-1} \sum_{i=1}^n (Y_i - \bar{X}_i'\hat{\beta})(Y_i - \bar{X}_i'\hat{\beta})'$.

IV: Set $g_i(\beta) = Z_i(Y_i - X_i'\beta)$. The MME is $\hat{\beta} = (\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{Y})$.

Generated Regressors: Set

$$g_i(\beta, \mathbf{A}) = \begin{pmatrix} \mathbf{A}' \mathbf{Z}_i (Y_i - \mathbf{Z}_i' \mathbf{A} \beta) \\ \text{vec}(\mathbf{Z}_i (\mathbf{X}_i' - \mathbf{Z}_i' \mathbf{A})) \end{pmatrix}.$$

The MME is $\hat{\mathbf{A}} = (\sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i')^{-1} (\sum_{i=1}^n \mathbf{Z}_i \mathbf{X}_i')$ and $\hat{\beta} = (\hat{\mathbf{A}}' \mathbf{Z}' \mathbf{Z} \hat{\mathbf{A}})^{-1} (\hat{\mathbf{A}}' \mathbf{Z}' \mathbf{Y})$.

A common feature of these examples is that the estimator can be written as the solution to a set of estimating equations (13.3). This provides a common framework which enables a convenient development of a unified distribution theory.

13.4 Overidentified Moment Equations

In the instrumental variables model $g_i(\beta) = \mathbf{Z}_i (Y_i - \mathbf{X}_i' \beta)$. Thus (13.2) is

$$\bar{g}_n(\beta) = \frac{1}{n} \sum_{i=1}^n g_i(\beta) = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i (Y_i - \mathbf{X}_i' \beta) = \frac{1}{n} (\mathbf{Z}' \mathbf{Y} - \mathbf{Z}' \mathbf{X} \beta). \quad (13.4)$$

We have defined the method of moments estimator for β as the parameter value which sets $\bar{g}_n(\beta) = 0$. However, when the model is overidentified (if $\ell > k$) this is generally impossible as there are more equations than free parameters. Equivalently, there is no choice of β which sets (13.4) to zero. Thus the method of moments estimator is not defined for the overidentified case.

While we cannot find an estimator which sets $\bar{g}_n(\beta)$ equal to zero we can try to find an estimator which makes $\bar{g}_n(\beta)$ as close to zero as possible.

One way to think about this is to define the vector $\mu = \mathbf{Z}' \mathbf{Y}$, the matrix $\mathbf{G} = \mathbf{Z}' \mathbf{X}$ and the “error” $\eta = \mu - \mathbf{G} \beta$. Then we can write (13.4) as $\mu = \mathbf{G} \beta + \eta$. This looks like a regression equation with the $\ell \times 1$ dependent variable μ , the $\ell \times k$ regressor matrix \mathbf{G} , and the $\ell \times 1$ error vector η . The goal is to make the error vector η as small as possible. Recalling our knowledge about least squares we deduct that a simple method is to regress μ on \mathbf{G} , obtaining $\hat{\beta} = (\mathbf{G}' \mathbf{G})^{-1} (\mathbf{G}' \mu)$. This minimizes the sum-of-squares $\eta' \eta$. This is certainly one way to make η “small”.

More generally we know that when errors are non-homogeneous it can be more efficient to estimate by weighted least squares. Thus for some weight matrix \mathbf{W} consider the estimator

$$\hat{\beta} = (\mathbf{G}' \mathbf{W} \mathbf{G})^{-1} (\mathbf{G}' \mathbf{W} \mu) = (\mathbf{X}' \mathbf{Z} \mathbf{W} \mathbf{Z}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Z} \mathbf{W} \mathbf{Z}' \mathbf{Y}).$$

This minimizes the weighted sum of squares $\eta' \mathbf{W} \eta$. This solution is known as the **generalized method of moments (GMM)**.

The estimator is typically defined as follows. Given a set of moment equations (13.2) and an $\ell \times \ell$ weight matrix $\mathbf{W} > 0$ the GMM criterion function is defined as

$$J(\beta) = n \bar{g}_n(\beta)' \mathbf{W} \bar{g}_n(\beta).$$

The factor “ n ” is not important for the definition of the estimator but is convenient for the distribution theory. The criterion $J(\beta)$ is the weighted sum of squared moment equation errors. When $\mathbf{W} = \mathbf{I}_\ell$ then $J(\beta) = n \bar{g}_n(\beta)' \bar{g}_n(\beta) = n \|\bar{g}_n(\beta)\|^2$, the square of the Euclidean length. Since we restrict attention to positive definite weight matrices \mathbf{W} the criterion $J(\beta)$ is non-negative.

Definition 13.1 The **Generalized Method of Moments (GMM)** estimator is

$$\hat{\beta}_{\text{gmm}} = \underset{\beta}{\text{argmin}} J(\beta).$$

Recall that in the just-identified case $k = \ell$ the method of moments estimator $\hat{\beta}_{\text{mm}}$ solves $\bar{g}_n(\hat{\beta}_{\text{mm}}) = 0$. Hence in this case $J(\hat{\beta}_{\text{mm}}) = 0$ which means that $\hat{\beta}_{\text{mm}}$ minimizes $J(\beta)$ and equals $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{mm}}$. This means that GMM includes MME as a special case. This implies that all of our results for GMM apply to any method of moments estimator.

In the over-identified case the GMM estimator depends on the choice of weight matrix \mathbf{W} and so this is an important focus of the theory. In the just-identified case the GMM estimator simplifies to the MME which does not depend on \mathbf{W} .

The method and theory of the generalized method of moments was developed in an influential paper by Lars Hansen (1982). This paper introduced the method, its asymptotic distribution, the form of the efficient weight matrix, and tests for overidentification.

13.5 Linear Moment Models

One of the great advantages of the moment equation framework is that it allows both linear and non-linear models. However, when the moment equations are linear in the parameters then we have explicit solutions for the estimates and a straightforward asymptotic distribution theory. Hence we start by confining attention to linear moment equations and return to nonlinear moment equations later. In the examples listed earlier the estimators which have linear moment equations include the sample mean, OLS, multivariate least squares, IV, and 2SLS. The estimates which have nonlinear moment equations include the sample variance, SUR, and generated regressors.

In particular, we focus on the overidentified IV model with moment equations

$$g_i(\beta) = Z_i(Y_i - X_i'\beta) \quad (13.5)$$

where Z_i is $\ell \times 1$ and X_i is $k \times 1$.

13.6 GMM Estimator

Given (13.5) the sample moment equations are (13.4). The GMM criterion can be written as

$$J(\beta) = n(\mathbf{Z}'\mathbf{Y} - \mathbf{Z}'\mathbf{X}\beta)' \mathbf{W} (\mathbf{Z}'\mathbf{Y} - \mathbf{Z}'\mathbf{X}\beta).$$

The GMM estimator minimizes $J(\beta)$. The first order conditions are

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta} J(\hat{\beta}) \\ &= 2 \frac{\partial}{\partial \beta} \bar{g}_n(\hat{\beta})' \mathbf{W} \bar{g}_n(\hat{\beta}) \\ &= -2 \left(\frac{1}{n} \mathbf{X}' \mathbf{Z} \right) \mathbf{W} \left(\frac{1}{n} \mathbf{Z}' (\mathbf{Y} - \mathbf{X}\hat{\beta}) \right). \end{aligned}$$

The solution is given as follows.

Theorem 13.1 For the overidentified IV model

$$\hat{\beta}_{\text{gmm}} = (\mathbf{X}' \mathbf{Z} \mathbf{W} \mathbf{Z}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Z} \mathbf{W} \mathbf{Z}' \mathbf{Y}). \quad (13.6)$$

While the estimator depends on \mathbf{W} the dependence is only up to scale. This is because if \mathbf{W} is replaced by $c\mathbf{W}$ for some $c > 0$, $\hat{\beta}_{\text{gmm}}$ does not change. When \mathbf{W} is fixed by the user we call $\hat{\beta}_{\text{gmm}}$ a **one-step GMM** estimator. The formula (13.6) applies for the over-identified ($\ell > k$) and the just-identified ($\ell = k$) case. When the model is just-identified then $\mathbf{X}'\mathbf{Z}$ is $k \times k$ so expression (13.6) simplifies to

$$\hat{\beta}_{\text{gmm}} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{W}^{-1} (\mathbf{X}'\mathbf{Z})^{-1} (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{Y}) = (\mathbf{Z}'\mathbf{X})^{-1} (\mathbf{Z}'\mathbf{Y}) = \hat{\beta}_{\text{iv}}$$

the IV estimator.

The GMM estimator (13.6) resembles the 2SLS estimator (12.29). In fact they are equal when $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$. This means that the 2SLS estimator is a one-step GMM estimator for the linear model.

Theorem 13.2 If $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ then $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{2sls}}$. Furthermore, if $k = \ell$ then $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{iv}}$.

13.7 Distribution of GMM Estimator

Let $\mathbf{Q} = \mathbb{E}[\mathbf{Z}\mathbf{X}']$ and $\Omega = \mathbb{E}[\mathbf{Z}\mathbf{Z}'e^2]$. Then

$$\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\mathbf{W}\left(\frac{1}{n}\mathbf{Z}'\mathbf{X}\right) \xrightarrow{p} \mathbf{Q}'\mathbf{W}\mathbf{Q}$$

and

$$\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\mathbf{W}\left(\frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}\right) \xrightarrow{d} \mathbf{Q}'\mathbf{W}\mathbf{N}(0, \Omega).$$

We conclude:

Theorem 13.3 Asymptotic Distribution of GMM Estimator. Under Assumption 12.2, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) \xrightarrow{d} \mathbf{N}(0, \mathbf{V}_\beta)$ where

$$\mathbf{V}_\beta = (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} (\mathbf{Q}'\mathbf{W}\Omega\mathbf{W}\mathbf{Q}) (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1}. \quad (13.7)$$

The GMM estimator is asymptotically normal with a “sandwich form” asymptotic variance.

Our derivation treated the weight matrix \mathbf{W} as if it is non-random but Theorem 13.3 applies to the random weight matrix case so long as $\widehat{\mathbf{W}}$ converges in probability to a positive definite limit \mathbf{W} . This may require scaling the weight matrix, for example replacing $\widehat{\mathbf{W}} = (\mathbf{Z}'\mathbf{Z})^{-1}$ with $\widehat{\mathbf{W}} = (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$. Since rescaling the weight matrix does not affect the estimator this is ignored in implementation.

13.8 Efficient GMM

The asymptotic distribution of the GMM estimator $\hat{\beta}_{\text{gmm}}$ depends on the weight matrix \mathbf{W} through the asymptotic variance \mathbf{V}_β . The asymptotically optimal weight matrix \mathbf{W}_0 is that which minimizes \mathbf{V}_β . This turns out to be $\mathbf{W}_0 = \Omega^{-1}$. The proof is left to Exercise 13.4.

When the GMM estimator $\hat{\beta}$ is constructed with $\mathbf{W} = \mathbf{W}_0 = \Omega^{-1}$ (or a weight matrix which is a consistent estimator of \mathbf{W}_0) we call it the **Efficient GMM** estimator:

$$\hat{\beta}_{\text{gmm}} = (\mathbf{X}'\mathbf{Z}\Omega^{-1}\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}\Omega^{-1}\mathbf{Z}'\mathbf{Y}).$$

Its asymptotic distribution takes a simpler form than in Theorem 13.3. By substituting $\mathbf{W} = \mathbf{W}_0 = \Omega^{-1}$ into (13.7) we find

$$\mathbf{V}_\beta = (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1}(\mathbf{Q}'\Omega^{-1}\Omega\Omega^{-1}\mathbf{Q})(\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1} = (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1}.$$

This is the asymptotic variance of the efficient GMM estimator.

Theorem 13.4 Asymptotic Distribution of GMM with Efficient Weight Matrix. Under Assumption 12.2 and $\mathbf{W} = \Omega^{-1}$, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) \xrightarrow{d} \mathbf{N}(0, \mathbf{V}_\beta)$ where $\mathbf{V}_\beta = (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1}$.

Theorem 13.5 Efficient GMM. Under Assumption 12.2, for any $\mathbf{W} > 0$,

$$(\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1}(\mathbf{Q}'\mathbf{W}\Omega\mathbf{W}\mathbf{Q})(\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} - (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1} \geq 0.$$

The inequality “ \geq ” can be replaced with “ $>$ ” if $\mathbf{W} \neq \Omega^{-1}$. Thus if $\hat{\beta}_{\text{gmm}}$ is the efficient GMM estimator and $\tilde{\beta}_{\text{gmm}}$ is another GMM estimator, then

$$\text{avar}[\hat{\beta}_{\text{gmm}}] \leq \text{avar}[\tilde{\beta}_{\text{gmm}}].$$

For a proof see Exercise 13.4.

This means that the smallest possible GMM covariance matrix (in the positive definite sense) is achieved by the efficient GMM weight matrix.

$\mathbf{W}_0 = \Omega^{-1}$ is not known in practice but it can be estimated consistently as we discuss in Section 13.10. For any $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}_0$ the asymptotic distribution in Theorem 13.4 is unaffected. Consequently we call any $\hat{\beta}_{\text{gmm}}$ constructed with an estimate of the efficient weight matrix an efficient GMM estimator.

By “efficient” we mean that this estimator has the smallest asymptotic variance in the class of GMM estimators with this set of moment conditions. This is a weak concept of optimality as we are only considering alternative weight matrices $\hat{\mathbf{W}}$. However, it turns out that the GMM estimator is semiparametrically efficient as shown by Gary Chamberlain (1987). If it is known that $\mathbb{E}[g_i(\beta)] = 0$ and this is all that is known this is a semi-parametric problem as the distribution of the data is unknown. Chamberlain showed that in this context no semiparametric estimator (one which is consistent globally for the class

of models considered) can have a smaller asymptotic variance than $(\mathbf{G}'\Omega^{-1}\mathbf{G})^{-1}$ where $\mathbf{G} = \mathbb{E}\left[\frac{\partial}{\partial\beta'}g_i(\beta)\right]$. Since the GMM estimator has this asymptotic variance it is semiparametrically efficient.

The results in this section show that in the linear model no estimator has better asymptotic efficiency than the efficient linear GMM estimator. No estimator can do better (in this first-order asymptotic sense) without imposing additional assumptions.

13.9 Efficient GMM versus 2SLS

For the linear model we introduced 2SLS as a standard estimator for β . Now we have introduced GMM which includes 2SLS as a special case. Is there a context where 2SLS is efficient?

To answer this question recall that 2SLS is GMM given the weight matrix $\widehat{\mathbf{W}} = (\mathbf{Z}'\mathbf{Z})^{-1}$ or equivalently $\widehat{\mathbf{W}} = (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ since scaling doesn't matter. Since $\widehat{\mathbf{W}} \xrightarrow{p} (\mathbb{E}[ZZ'])^{-1}$ this is asymptotically equivalent to the weight matrix $\mathbf{W} = (\mathbb{E}[ZZ'])^{-1}$. In contrast, the efficient weight matrix takes the form $(\mathbb{E}[ZZ'e^2])^{-1}$. Now suppose that the structural equation error e is conditionally homoskedastic in the sense that $\mathbb{E}[e^2 | Z] = \sigma^2$. Then the efficient weight matrix equals $\mathbf{W} = (\mathbb{E}[ZZ'])^{-1}\sigma^{-2}$ or equivalently $\mathbf{W} = (\mathbb{E}[ZZ'])^{-1}$ since scaling doesn't matter. The latter weight matrix is the same as the 2SLS asymptotic weight matrix. This shows that the 2SLS weight matrix is the efficient weight matrix under conditional homoskedasticity.

Theorem 13.6 Under Assumption 12.2 and $\mathbb{E}[e^2 | Z] = \sigma^2$, $\widehat{\beta}_{2\text{sls}}$ is efficient GMM.

This shows that 2SLS is efficient under homoskedasticity. When homoskedasticity holds there is no reason to use efficient GMM over 2SLS. More broadly, when homoskedasticity is a reasonable approximation then 2SLS will be a reasonable estimator. However, this result also shows that in the general case where the error is conditionally heteroskedastic, 2SLS is inefficient relative to efficient GMM.

13.10 Estimation of the Efficient Weight Matrix

To construct the efficient GMM estimator we need a consistent estimator $\widehat{\mathbf{W}}$ of $\mathbf{W}_0 = \Omega^{-1}$. The convention is to form an estimator $\widehat{\Omega}$ of Ω and then set $\widehat{\mathbf{W}} = \widehat{\Omega}^{-1}$.

The **two-step GMM estimator** proceeds by using a one-step consistent estimator of β to construct the weight matrix estimator $\widehat{\mathbf{W}}$. In the linear model the natural one-step estimator for β is 2SLS. Set $\tilde{e}_i = Y_i - X_i'\widehat{\beta}_{2\text{sls}}$, $\tilde{g}_i = g_i(\tilde{\beta}) = Z_i\tilde{e}_i$, and $\bar{g}_n = n^{-1}\sum_{i=1}^n \tilde{g}_i$. Two moment estimators of Ω are

$$\widehat{\Omega} = \frac{1}{n} \sum_{i=1}^n \tilde{g}_i \tilde{g}_i' \quad (13.8)$$

and

$$\widehat{\Omega}^* = \frac{1}{n} \sum_{i=1}^n (\tilde{g}_i - \bar{g}_n)(\tilde{g}_i - \bar{g}_n)'. \quad (13.9)$$

The estimator (13.8) is an uncentered covariance matrix estimator while the estimator (13.9) is a centered version. Either is consistent when $\mathbb{E}[Ze] = 0$ which holds under correct specification. However under misspecification we may have $\mathbb{E}[Ze] \neq 0$. In the latter context $\widehat{\Omega}^*$ remains an estimator of $\text{var}[Ze]$

while $\hat{\Omega}$ is an estimator of $\mathbb{E}[ZZ'e^2]$. In this sense $\hat{\Omega}^*$ is a robust variance estimator. For some testing problems it turns out to be preferable to use a covariance matrix estimator which is robust to the alternative hypothesis. For these reasons estimator (13.9) is generally preferred. The uncentered estimator (13.8) is more commonly seen in practice since it is the default choice by most packages. It is also worth observing that when the model is just identified then $\bar{g}_n = 0$ so the two are algebraically identical. The choice of weight matrix may also impact covariance matrix estimation as discussed in Section 13.12.

Given the choice of covariance matrix estimator we set $\widehat{W} = \hat{\Omega}^{-1}$ or $\widehat{W} = \hat{\Omega}^{*-1}$. Given this weight matrix we construct the **two-step GMM estimator** as (13.6) using the weight matrix \widehat{W} .

Since the 2SLS estimator is consistent for β , by arguments nearly identical to those used for covariance matrix estimation we can show that $\hat{\Omega}$ and $\hat{\Omega}^*$ are consistent for Ω and thus \widehat{W} is consistent for Ω^{-1} . See Exercise 13.3.

This also means that the two-step GMM estimator satisfies the conditions for Theorem 13.4.

Theorem 13.7 Under Assumption 12.2 and $\Omega > 0$, if $\widehat{W} = \hat{\Omega}^{-1}$ or $\widehat{W} = \hat{\Omega}^{*-1}$ where the latter are defined in (13.8) and (13.9) then as $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) \xrightarrow{d} N(0, V_\beta)$ where $V_\beta = (Q'\Omega^{-1}Q)^{-1}$.

This shows that the two-step GMM estimator is asymptotically efficient.

The two-step GMM estimator of the IV regression equation can be computed in Stata using the `ivregress gmm` command. By default it uses formula (13.8). The centered version (13.9) may be selected using the `center` option.

13.11 Iterated GMM

The asymptotic distribution of the two-step GMM estimator does not depend on the choice of the preliminary one-step estimator. However, the actual value of the estimator depends on this choice and so will the finite sample distribution. This is undesirable and likely inefficient. To remove this dependence we can iterate the estimation sequence. Specifically, given $\hat{\beta}_{\text{gmm}}$ we can construct an updated weight matrix estimate \widehat{W} and then re-estimate $\hat{\beta}_{\text{gmm}}$. This updating can be iterated until convergence¹. The result is called the **iterated GMM estimator** and is a common implementation of efficient GMM.

Interestingly, B. E. Hansen and Lee (2021) show that the iterated GMM estimator is unaffected if the weight matrix is computed with or without centering. Standard errors and test statistics, however, will be affected by the choice.

The iterated GMM estimator of the IV regression equation can be computed in Stata using the `ivregress gmm` command using the `igmm` option.

13.12 Covariance Matrix Estimation

An estimator of the asymptotic variance of $\hat{\beta}_{\text{gmm}}$ can be obtained by replacing the matrices in the asymptotic variance formula by consistent estimators.

¹In practice, “convergence” obtains when the difference between the estimates at subsequent steps is smaller than a pre-specified tolerance. A sufficient condition for convergence is that the sequence is a contraction mapping. Indeed, B. Hansen and Lee (2021) have shown that the iterated GMM estimator generally satisfies this condition in large samples.

For the one-step or two-step GMM estimator the covariance matrix estimator is

$$\hat{V}_\beta = \left(\hat{Q}' \hat{W} \hat{Q} \right)^{-1} \left(\hat{Q}' \hat{W} \hat{\Omega} \hat{W} \hat{Q} \right) \left(\hat{Q}' \hat{W} \hat{Q} \right)^{-1} \quad (13.10)$$

where $\hat{Q} = \frac{1}{n} \sum_{i=1}^n Z_i X_i'$. The weight matrix is constructed using either the uncentered estimator (13.8) or centered estimator (13.9) with the residuals $\hat{e}_i = Y_i - X_i' \hat{\beta}_{\text{gmm}}$.

For the efficient iterated GMM estimator the covariance matrix estimator is

$$\hat{V}_\beta = \left(\hat{Q}' \hat{\Omega}^{-1} \hat{Q} \right)^{-1} = \left(\left(\frac{1}{n} X' Z \right) \hat{\Omega}^{-1} \left(\frac{1}{n} Z' X \right) \right)^{-1}. \quad (13.11)$$

$\hat{\Omega}$ can be computed using either the uncentered estimator (13.8) or centered estimator (13.9). Based on the asymptotic approximation the estimator (13.11) can be used as well for the two-step estimator but should use the final residuals $\hat{e}_i = Y_i - X_i' \hat{\beta}_{\text{gmm}}$.

Asymptotic standard errors are given by the square roots of the diagonal elements of $n^{-1} \hat{V}_\beta$.

It is unclear if it is preferred to use the covariance matrix estimator based on the centered or uncentered estimator of Ω to construct the covariance matrix estimator. Using the centered estimator results in a smaller covariance matrix and standard errors and thus more “significant” tests based on asymptotic critical values. In contrast the uncentered estimator of Ω will result in larger standard errors and will thus be more “conservative”.

In Stata, the default covariance matrix estimation method is determined by the choice of weight matrix. Thus if the centered estimator (13.9) is used for the weight matrix it is also used for the covariance matrix estimator.

13.13 Clustered Dependence

In Section 4.21 we introduced clustered dependence and in Section 12.25 described covariance matrix estimation for 2SLS. The methods extend naturally to GMM but with the additional complication of potentially altering weight matrix calculation.

The structural equation for the g^{th} cluster can be written as the matrix system $Y_g = X_g \beta + e_g$. Using this notation the centered GMM estimator with weight matrix W can be written as

$$\hat{\beta}_{\text{gmm}} - \beta = (X' Z W Z' X)^{-1} X' Z W \left(\sum_{g=1}^G Z'_g e_g \right).$$

The cluster-robust covariance matrix estimator for $\hat{\beta}_{\text{gmm}}$ is

$$\hat{V}_\beta = (X' Z W Z' X)^{-1} X' Z W \hat{S} W Z' X (X' Z W Z' X)^{-1} \quad (13.12)$$

with

$$\hat{S} = \sum_{g=1}^G Z'_g \hat{e}_g \hat{e}_g' Z_g \quad (13.13)$$

and the clustered residuals

$$\hat{e}_g = Y_g - X_g \hat{\beta}_{\text{gmm}}. \quad (13.14)$$

The cluster-robust estimator (13.12) is appropriate for the one-step or two-step GMM estimator. It is also appropriate for the iterated estimator when the latter uses a conventional (non-clustered) efficient weight matrix. However in the clustering context it is more natural to use a cluster-robust weight matrix

such as $\mathbf{W} = \hat{\mathbf{S}}^{-1}$ where $\hat{\mathbf{S}}$ is a cluster-robust covariance estimator as in (13.13) based on a one-step or iterated residual. This gives rise to the cluster-robust GMM estimator

$$\hat{\beta}_{\text{gmm}} = (\mathbf{X}' \mathbf{Z} \hat{\mathbf{S}}^{-1} \mathbf{Z}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \hat{\mathbf{S}}^{-1} \mathbf{Z}' \mathbf{Y}. \quad (13.15)$$

An appropriate cluster-robust covariance matrix estimator is

$$\hat{\mathbf{V}}_{\beta} = (\mathbf{X}' \mathbf{Z} \hat{\mathbf{S}}^{-1} \mathbf{Z}' \mathbf{X})^{-1}$$

where $\hat{\mathbf{S}}$ is calculated using the final residuals.

To implement a cluster-robust weight matrix use the 2SLS estimator for first step. Compute the cluster residuals (13.14) and covariance matrix (13.13). Then (13.15) is the two-step GMM estimator. Iterating the residuals and covariance matrix until convergence we obtain the iterated GMM estimator.

In Stata, using the `ivregress gmm` command with the `cluster` option implements the two-step GMM estimator using the cluster-robust weight matrix and cluster-robust covariance matrix estimator. To use the centered covariance matrix use the `center` option and to implement the iterated GMM estimator use the `igmm` option. Alternatively, you can use the `wmatrix` and `vce` options to separately specify the weight matrix and covariance matrix estimation methods.

13.14 Wald Test

For a given function $r(\beta) : \mathbb{R}^k \rightarrow \Theta \subset \mathbb{R}^q$ we define the parameter $\theta = r(\beta)$. The GMM estimator of θ is $\hat{\theta}_{\text{gmm}} = r(\hat{\beta}_{\text{gmm}})$. By the delta method it is asymptotically normal with covariance matrix $\mathbf{V}_{\theta} = \mathbf{R}' \mathbf{V}_{\beta} \mathbf{R}$ where $\mathbf{R} = \frac{\partial}{\partial \beta} r(\beta)'$. An estimator of the asymptotic covariance matrix is $\hat{\mathbf{V}}_{\theta} = \hat{\mathbf{R}}' \hat{\mathbf{V}}_{\beta} \hat{\mathbf{R}}$ where $\hat{\mathbf{R}} = \frac{\partial}{\partial \beta} r(\hat{\beta}_{\text{gmm}})'$. When θ is scalar then an asymptotic standard error for $\hat{\theta}_{\text{gmm}}$ is formed as $\sqrt{n^{-1} \hat{\mathbf{V}}_{\theta}}$.

A standard test of the hypothesis $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_1 : \theta \neq \theta_0$ is based on the Wald statistic

$$W = n(\hat{\theta} - \theta_0)' \hat{\mathbf{V}}_{\theta}^{-1} (\hat{\theta} - \theta_0).$$

Let $G_q(u)$ denote the χ_q^2 distribution function.

Theorem 13.8 Under Assumption 12.2, Assumption 7.3, and \mathbb{H}_0 , as $n \rightarrow \infty$, $W \xrightarrow{d} \chi_q^2$. For c satisfying $\alpha = 1 - G_q(c)$, $\mathbb{P}[W > c \mid \mathbb{H}_0] \rightarrow \alpha$ so the test “Reject \mathbb{H}_0 if $W > c$ ” has asymptotic size α .

For a proof see Exercise 13.5.

In Stata, the commands `test` and `testparm` can be used after `ivregress gmm` to implement Wald tests of linear hypotheses. The commands `nlcom` and `testnl` can be used after `ivregress gmm` to implement Wald tests of nonlinear hypotheses.

13.15 Restricted GMM

It is often desirable to impose restrictions on the coefficients. In this section we consider estimation subject to the linear constraints $\mathbf{R}'\beta = \mathbf{c}$. In the following section we consider nonlinear constraints.

The constrained GMM estimator minimizes the GMM criterion subject to the constraint. It is

$$\hat{\beta}_{\text{cgmm}} = \underset{\mathbf{R}'\beta = \mathbf{c}}{\operatorname{argmin}} J(\beta).$$

This is the parameter vector which makes the estimating equations as close to zero as possible with respect to the weighted quadratic distance while imposing the restriction on the parameters.

Suppose the weight matrix \mathbf{W} is fixed. Using the methods of Chapter 8 it is straightforward to derive that the constrained GMM estimator is

$$\hat{\beta}_{\text{cgmm}} = \hat{\beta}_{\text{gmm}} - (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-1} \mathbf{R} \left(\mathbf{R}' (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-1} \mathbf{R} \right)^{-1} (\mathbf{R}'\hat{\beta}_{\text{gmm}} - \mathbf{c}). \quad (13.16)$$

(For details, see Exercise 13.6.)

We derive the asymptotic distribution under the assumption that the restriction is true. Make the substitution $\mathbf{c} = \mathbf{R}'\beta$ in (13.16) and reorganize to find

$$\sqrt{n}(\hat{\beta}_{\text{cgmm}} - \beta) = \left(\mathbf{I}_k - (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-1} \mathbf{R} \left(\mathbf{R}' (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-1} \mathbf{R} \right)^{-1} \mathbf{R}' \right) \sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta). \quad (13.17)$$

This is a linear function of $\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta)$. Since the asymptotic distribution of the latter is known the asymptotic distribution of $\sqrt{n}(\hat{\beta}_{\text{cgmm}} - \beta)$ is a linear function of the former.

Theorem 13.9 Under Assumptions 12.2 and 8.3, for the constrained GMM estimator (13.16), $\sqrt{n}(\hat{\beta}_{\text{cgmm}} - \beta) \xrightarrow{d} N(0, \mathbf{V}_{\text{cgmm}})$ as $n \rightarrow \infty$, where

$$\begin{aligned} \mathbf{V}_{\text{cgmm}} = & \mathbf{V}_{\beta} - (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \mathbf{R} \left(\mathbf{R}' (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \mathbf{R} \right)^{-1} \mathbf{R}' \mathbf{V}_{\beta} \\ & - \mathbf{V}_{\beta} \mathbf{R} \left(\mathbf{R}' (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \mathbf{R} \right)^{-1} \mathbf{R}' (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \\ & + (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \mathbf{R} \left(\mathbf{R}' (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \mathbf{R} \right)^{-1} \mathbf{R}' \mathbf{V}_{\beta} \mathbf{R} \left(\mathbf{R}' (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \mathbf{R} \right)^{-1} \mathbf{R}' (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1}. \end{aligned} \quad (13.18)$$

For a proof, see Exercise 13.8. Unfortunately the asymptotic covariance matrix formula (13.18) is quite tedious!

Now suppose that the weight matrix is set as $\mathbf{W} = \hat{\Omega}^{-1}$, the efficient weight matrix from unconstrained estimation. In this case the constrained GMM estimator can be written as

$$\hat{\beta}_{\text{cgmm}} = \hat{\beta}_{\text{gmm}} - \hat{\mathbf{V}}_{\beta} \mathbf{R} (\mathbf{R}' \hat{\mathbf{V}}_{\beta} \mathbf{R})^{-1} (\mathbf{R}' \hat{\beta}_{\text{gmm}} - \mathbf{c}) \quad (13.19)$$

which is the same formula (8.25) as efficient minimum distance. (For details, see Exercise 13.7.) We find that the asymptotic covariance matrix simplifies considerably.

Theorem 13.10 Under Assumptions 12.2 and 8.3, for the efficient constrained GMM estimator (13.19), $\sqrt{n}(\hat{\beta}_{\text{cgmm}} - \beta) \xrightarrow{d} N(0, V_{\text{cgmm}})$ as $n \rightarrow \infty$, where

$$V_{\text{cgmm}} = V_{\beta} - V_{\beta} R (R' V_{\beta} R)^{-1} R' V_{\beta}. \quad (13.20)$$

For a proof, see Exercise 13.9.

The asymptotic covariance matrix (13.20) can be estimated by

$$\hat{V}_{\text{cgmm}} = \tilde{V}_{\beta} - \tilde{V}_{\beta} R (R' \tilde{V}_{\beta} R)^{-1} R' \tilde{V}_{\beta}. \quad (13.21)$$

$$\begin{aligned} \tilde{V}_{\beta} &= \left(\hat{Q}' \tilde{\Omega}^{-1} \hat{Q} \right)^{-1} \\ \tilde{\Omega} &= \frac{1}{n} \sum_{i=1}^n Z_i Z_i' \tilde{e}_i^2 \\ \tilde{e}_i &= Y_i - X_i' \hat{\beta}_{\text{cgmm}}. \end{aligned} \quad (13.22)$$

The covariance matrix (13.18) can be estimated similarly, though using (13.10) to estimate V_{β} . The covariance matrix estimator $\tilde{\Omega}$ can also be replaced with a centered version.

A constrained iterated GMM estimator can be implemented by setting $W = \tilde{\Omega}^{-1}$ where $\tilde{\Omega}$ is defined in (13.22) and then iterating until convergence. This is a natural estimator as it is the appropriate implementation of iterated GMM.

Since both $\hat{\Omega}$ and $\tilde{\Omega}$ converge to the same limit Ω under the assumption that the constraint is true the constrained iterated GMM estimator has the asymptotic distribution given in Theorem 13.10.

13.16 Nonlinear Restricted GMM

Nonlinear constraints on the parameters can be written as $r(\beta) = 0$ for some function $r : \mathbb{R}^k \rightarrow \mathbb{R}^q$. The constraint is nonlinear if $r(\beta)$ cannot be written as a linear function of β . Least squares estimation subject to nonlinear constraints was explored in Section 8.14. In this section we introduce GMM estimation subject to nonlinear constraints.

The constrained GMM estimator minimizes the GMM criterion subject to the constraint. It is

$$\hat{\beta}_{\text{cgmm}} = \underset{r(\beta)=0}{\operatorname{argmin}} J(\beta). \quad (13.23)$$

This is the parameter vector which makes the estimating equations as close to zero as possible with respect to the weighted quadratic distance while imposing the restriction on the parameters.

In general there is no explicit solution for $\hat{\beta}_{\text{cgmm}}$. Instead the solution is found numerically. Fortunately there are excellent nonlinear constrained optimization solvers implemented in standard software packages.

For the asymptotic distribution assume that the restriction $r(\beta) = 0$ is true. Using the same methods as in the proof of Theorem 8.10 we can show that (13.17) approximately holds in the sense that

$$\sqrt{n}(\hat{\beta}_{\text{cgmm}} - \beta) = \left(I_k - (X' Z W Z' X)^{-1} R' (R' (X' Z W Z' X)^{-1} R)^{-1} R' \right) \sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) + o_p(1)$$

where $R = \frac{\partial}{\partial \beta} r(\beta)'$. Thus the asymptotic distribution of the constrained estimator takes the same form as in the linear case.

Theorem 13.11 Under Assumptions 12.2 and 8.3, for the constrained GMM estimator (13.23), $\sqrt{n}(\hat{\beta}_{\text{cgmm}} - \beta) \xrightarrow{d} N(0, V_{\text{cgmm}})$ as $n \rightarrow \infty$, where V_{cgmm} equals (13.18). If $W = \hat{\Omega}^{-1}$, then V_{cgmm} equals (13.20).

The asymptotic covariance matrix in the efficient case is estimated by (13.21) with R replaced with $\hat{R} = \frac{\partial}{\partial \beta} r(\hat{\beta}_{\text{cgmm}})'$. The asymptotic covariance matrix (13.18) in the general case is estimated similarly.

To implement an iterated restricted GMM estimator the weight matrix may be set as $W = \tilde{\Omega}^{-1}$ where $\tilde{\Omega}$ is defined in (13.22), and then iterated until convergence.

13.17 Constrained Regression

Take the conventional projection model $Y = X'\beta + e$ with $E[Xe] = 0$. This is a special case of GMM as it is model (13.5) with $Z = X$. The just-identified GMM estimator equals least squares $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{ols}}$.

In Chapter 8 we discussed estimation of the projection model subject to linear constraints $R'\beta = c$, which includes exclusion restrictions. Since the projection model is a special case of GMM the constrained projection model is also constrained GMM. From the results of Section 13.15 we find that the efficient constrained GMM estimator is

$$\hat{\beta}_{\text{cgmm}} = \hat{\beta}_{\text{ols}} - \hat{V}_{\beta} R (R' \hat{V}_{\beta} R)^{-1} (R' \hat{\beta}_{\text{ols}} - c) = \hat{\beta}_{\text{emd}},$$

the efficient minimum distance estimator. Thus for linear constraints on the linear projection model efficient GMM equals efficient minimum distance. Thus one convenient method to implement efficient minimum distance is GMM.

13.18 Multivariate Regression

GMM methods can simplify estimation and inference for multivariate regressions such as those introduced in Chapter 11.

The general multivariate regression (projection) model is

$$\begin{aligned} Y_j &= X_j' \beta_j + e_j \\ E[X_j e_j] &= 0 \end{aligned}$$

for $j = 1, \dots, m$. Using the notation from Section 11.2 the equations can be written jointly as $Y = \bar{X}\beta + e$ and for the full sample as $Y = \bar{X}\beta + e$. The \bar{k} moment conditions are

$$E[\bar{X}'(Y - \bar{X}\beta)] = 0. \quad (13.24)$$

Given a $\bar{k} \times \bar{k}$ weight matrix W the GMM criterion is

$$J(\beta) = n(Y - \bar{X}\beta)' \bar{X} W \bar{X}' (Y - \bar{X}\beta).$$

The GMM estimator $\hat{\beta}_{\text{gmm}}$ minimizes $J(\beta)$. Since this is a just-identified model the estimator solves the sample equations

$$\bar{X}'(Y - \bar{X}\hat{\beta}_{\text{gmm}}) = 0.$$

The solution is

$$\hat{\beta}_{\text{gmm}} = \left(\sum_{i=1}^n \bar{X}'_i \bar{X}_i \right)^{-1} \left(\sum_{i=1}^n \bar{X}'_i Y_i \right) = (\bar{X}' \bar{X})^{-1} (\bar{X}' \mathbf{Y}) = \hat{\beta}_{\text{ols}},$$

the multivariate least squares estimator.

Thus the unconstrained GMM estimator of the multivariate regression model is least squares. The estimator does not depend on the weight matrix since the model is just-identified.

A important advantage of the GMM framework is the ability to incorporate cross-equation constraints. Consider the class of restrictions $\mathbf{R}'\beta = \mathbf{c}$. Minimization of the GMM criterion subject to this restriction has solutions as described in (13.15). The restricted GMM estimator is

$$\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{ols}} - (\bar{X}' \bar{X} \mathbf{W} \bar{X}' \bar{X})^{-1} \mathbf{R}' (\bar{X}' \bar{X} \mathbf{W} \bar{X}' \bar{X})^{-1} \mathbf{R} (\mathbf{R}' \hat{\beta}_{\text{ols}} - \mathbf{c}).$$

This estimator depends on the weight matrix because it is over-identified.

A simple choice for weight matrix is $\mathbf{W} = \bar{X}' \bar{X}$. This leads to the one-step estimator

$$\hat{\beta}_1 = \hat{\beta}_{\text{ols}} - (\bar{X}' \bar{X})^{-1} \mathbf{R}' (\mathbf{R}' (\bar{X}' \bar{X})^{-1} \mathbf{R})^{-1} (\mathbf{R}' \hat{\beta}_{\text{ols}} - \mathbf{c}).$$

The asymptotically efficient choice sets $\mathbf{W} = \hat{\Omega}^{-1}$ where $\hat{\Omega} = n^{-1} \sum_{i=1}^n \bar{X}'_i \hat{e}_i \hat{e}'_i \bar{X}_i$ and $\hat{e}_i = Y_i - \bar{X}_i \hat{\beta}_1$. This leads to the two-step estimator

$$\hat{\beta}_2 = \hat{\beta}_{\text{ols}} - (\bar{X}' \bar{X} \hat{\Omega}^{-1} \bar{X}' \bar{X})^{-1} \mathbf{R}' (\mathbf{R}' (\bar{X}' \bar{X} \hat{\Omega}^{-1} \bar{X}' \bar{X})^{-1} \mathbf{R})^{-1} (\mathbf{R}' \hat{\beta}_{\text{ols}} - \mathbf{c}).$$

When the regressors X are common across all equations the multivariate regression model can be written conveniently as in (11.3): $Y = \mathbf{B}'X + e$ with $\mathbb{E}[Xe'] = 0$. The moment restrictions can be written as the matrix system $\mathbb{E}[X(Y' - X'\mathbf{B})] = 0$. Written as a vector system this is (13.24) and leads to the same restricted GMM estimators.

These are general formula for imposing restrictions. In specific cases (such as an exclusion restriction) direct methods may be more convenient. In all cases the solution is found by minimization of the GMM criterion $J(\beta)$ subject to the restriction.

13.19 Distance Test

In Section 13.14 we introduced Wald tests of the hypothesis $\mathbb{H}_0 : \theta = \theta_0$ where $\theta = r(\beta)$ for a given function $r(\beta) : \mathbb{R}^k \rightarrow \Theta \subset \mathbb{R}^q$. When $r(\beta)$ is nonlinear an alternative is to use a criterion-based statistic. This is sometimes called the GMM Distance statistic and sometimes called a LR-like statistic (the LR is for likelihood-ratio). The idea was first put forward by Newey and West (1987a).

The idea is to compare the unrestricted and restricted estimators by contrasting the criterion functions. The unrestricted estimator takes the form

$$\hat{\beta}_{\text{gmm}} = \underset{\beta}{\operatorname{argmin}} \hat{J}(\beta)$$

where

$$\hat{J}(\beta) = n \bar{g}_n(\beta)' \hat{\Omega}^{-1} \bar{g}_n(\beta)$$

is the unrestricted GMM criterion with an efficient weight matrix estimate $\hat{\Omega}$. The minimized value of the criterion is $\hat{J} = \hat{J}(\hat{\beta}_{\text{gmm}})$.

As in Section 13.15, the estimator subject to $r(\beta) = \theta_0$ is

$$\hat{\beta}_{\text{cgmm}} = \underset{r(\beta)=\theta_0}{\operatorname{argmin}} \tilde{J}(\beta)$$

where

$$\tilde{J}(\beta) = n \bar{g}_n(\beta)' \tilde{\Omega}^{-1} \bar{g}_n(\beta)$$

which depends on an efficient weight matrix estimator, either $\hat{\Omega}$ (the same as the unrestricted estimator) or $\tilde{\Omega}$ (the iterated weight matrix from constrained estimation). The minimized value of the criterion is $\tilde{J} = \tilde{J}(\hat{\beta}_{\text{cgmm}})$.

The GMM distance (or LR-like) statistic is the difference in the criterion functions: $D = \tilde{J} - \hat{J}$. The distance test shares the useful feature of LR tests in that it is a natural by-product of the computation of alternative models.

The test has the following large sample distribution.

Theorem 13.12 Under Assumption 12.2, Assumption 7.3, and \mathbb{H}_0 , then as $n \rightarrow \infty$, $D \xrightarrow{d} \chi_q^2$. For c satisfying $\alpha = 1 - G_q(c)$, $\mathbb{P}[D > c \mid \mathbb{H}_0] \rightarrow \alpha$. The test “Reject \mathbb{H}_0 if $D > c$ ” has asymptotic size α .

The proof is given in Section 13.28.

Theorem 13.12 shows that the distance statistic has the same asymptotic distribution as Wald and likelihood ratio statistics and can be interpreted similarly. Small values of D mean that imposing the restriction does not result in a large value of the moment equations. Hence the restriction appears to be compatible with the data. On the other hand, large values of D mean that imposing the restriction results in a much larger value of the moment equations, implying that the restriction is not compatible with the data. The finding that the asymptotic distribution is chi-squared allows the calculation of asymptotic critical values and p-values.

We now discuss the choice of weight matrix. As mentioned above one simple choice is to set $\tilde{\Omega} = \hat{\Omega}$. In this case we have the following result.

Theorem 13.13 If $\tilde{\Omega} = \hat{\Omega}$ then $D \geq 0$. Furthermore, if r is linear in β then D equals the Wald statistic.

The statement that $\tilde{\Omega} = \hat{\Omega}$ implies $D \geq 0$ follows from the fact that in this case the criterion functions $\hat{J}(\beta) = \tilde{J}(\beta)$ are identical so the constrained minimum cannot be smaller than the unconstrained. The statement that linear hypotheses and $\tilde{\Omega} = \hat{\Omega}$ implies $D = W$ follows from applying the expression for the constrained GMM estimator (13.19) and using the covariance matrix formula (13.11).

The fact that $D \geq 0$ when $\tilde{\Omega} = \hat{\Omega}$ motivated Newey and West (1987a) to recommend this choice. However, $\tilde{\Omega} = \hat{\Omega}$ is not necessary. Instead, setting $\tilde{\Omega}$ to equal the constrained efficient weight matrix is natural for efficient estimation of $\hat{\beta}_{\text{cgmm}}$. In the event that $D < 0$ the test simply fails to reject \mathbb{H}_0 at any significance level.

As discussed in Section 9.17 for tests of nonlinear hypotheses the Wald statistic can work quite poorly. In particular, the Wald statistic is affected by how the hypothesis $r(\beta)$ is formulated. In contrast, the

distance statistic D is not affected by the algebraic formulation of the hypothesis. Current evidence suggests that the D statistic appears to have good sampling properties, and is a preferred test statistic relative to the Wald statistic for nonlinear hypotheses. (See B. E. Hansen (2006).)

In Stata the command `estat overid` after `ivregress gmm` can be used to report the value of the GMM criterion J . By estimating the two nested GMM regressions the values \hat{J} and \tilde{J} can be obtained and D computed.

13.20 Continuously-Updated GMM

An alternative to the two-step GMM estimator can be constructed by letting the weight matrix be an explicit function of β . This leads to the criterion function

$$J(\beta) = n \bar{g}_n(\beta)' \left(\frac{1}{n} \sum_{i=1}^n g_i(\beta) g_i(\beta)' \right)^{-1} \bar{g}_n(\beta).$$

The $\hat{\beta}$ which minimizes this function is called the **continuously-updated GMM (CU-GMM) estimator** and was introduced by L. Hansen, Heaton and Yaron (1996).

A complication is that the continuously-updated criterion $J(\beta)$ is not quadratic in β . This means that minimization requires numerical methods. It may appear that the CU-GMM estimator is the same as the iterated GMM estimator but this is not the case at all. They solve distinct first-order conditions and can be quite different in applications.

Relative to traditional GMM the CU-GMM estimator has lower bias but thicker distributional tails. While it has received considerable theoretical attention it is not used commonly in applications.

13.21 OverIdentification Test

In Section 12.31 we introduced the Sargan (1958) overidentification test for the 2SLS estimator under the assumption of homoskedasticity. L. Hansen (1982) generalized the test to cover the GMM estimator allowing for general heteroskedasticity.

Recall, overidentified models ($\ell > k$) are special in the sense that there may not be a parameter value β such that the moment condition $\mathbb{H}_0 : \mathbb{E}[Ze] = 0$ holds. Thus the model – the overidentifying restrictions – are testable.

For example, take the linear model $Y = \beta_1' X_1 + \beta_2' X_2 + e$ with $\mathbb{E}[X_1 e] = 0$ and $\mathbb{E}[X_2 e] = 0$. It is possible that $\beta_2 = 0$ so that the linear equation may be written as $Y = \beta_1' X_1 + e$. However, it is possible that $\beta_2 \neq 0$. In this case it is impossible to find a value of β_1 such that both $\mathbb{E}[X_1(Y - X_1' \beta_1)] = 0$ and $\mathbb{E}[X_2(Y - X_1' \beta_1)] = 0$ hold simultaneously. In this sense an exclusion restriction can be seen as an overidentifying restriction.

Note that $\bar{g}_n \xrightarrow{p} \mathbb{E}[Ze]$ and thus \bar{g}_n can be used to assess the hypothesis $\mathbb{E}[Ze] = 0$. Assuming that an efficient weight matrix estimator is used the criterion function at the parameter estimator is $J = J(\hat{\beta}_{\text{gmm}}) = n \bar{g}_n' \hat{\Omega}^{-1} \bar{g}_n$. This is a quadratic form in \bar{g}_n and is thus a natural test statistic for $\mathbb{H}_0 : \mathbb{E}[Ze] = 0$. Note that we assume that the criterion function is constructed with an efficient weight matrix estimator. This is important for the distribution theory.

Theorem 13.14 Under Assumption 12.2 then as $n \rightarrow \infty$, $J = J(\hat{\beta}_{\text{gmm}}) \xrightarrow{d} \chi_{\ell-k}^2$. For c satisfying $\alpha = 1 - G_{\ell-k}(c)$, $\mathbb{P}[J > c \mid \mathbb{H}_0] \rightarrow \alpha$ so the test “Reject \mathbb{H}_0 if $J > c$ ” has asymptotic size α .

The proof of the theorem is left to Exercise 13.13.

The degrees of freedom of the asymptotic distribution are the number of overidentifying restrictions. If the statistic J exceeds the chi-square critical value we can reject the model. Based on this information alone it is unclear what is wrong but it is typically cause for concern. The GMM overidentification test is a useful by-product of the GMM methodology and it is advisable to report the statistic J whenever GMM is the estimation method. When over-identified models are estimated by GMM it is customary to report the J statistic as a general test of model adequacy.

In Stata the command `estat overid` after `ivregress gmm` can be used to implement the overidentification test. The GMM criterion J and its asymptotic p-value using the $\chi^2_{\ell-k}$ distribution are reported.

13.22 Subset OverIdentification Tests

In Section 12.32 we introduced subset overidentification tests for the 2SLS estimator under the assumption of homoskedasticity. In this section we describe how to construct analogous tests for the GMM estimator under general heteroskedasticity.

Recall, subset overidentification tests are used when it is desired to focus attention on a subset of instruments whose validity is questioned. Partition $Z = (Z_a, Z_b)$ with dimensions ℓ_a and ℓ_b , respectively, where Z_a contains the instruments which are believed to be uncorrelated with e and Z_b contains the instruments which may be correlated with e . It is necessary to select this partition so that $\ell_a > k$, so that the instruments Z_a alone identify the parameters.

Given this partition the maintained hypothesis is $\mathbb{E}[Z_a e] = 0$. The null and alternative hypotheses are $\mathbb{H}_0 : \mathbb{E}[Z_b e] = 0$ and $\mathbb{H}_1 : \mathbb{E}[Z_b e] \neq 0$. The GMM test is constructed as follows. First, estimate the model by efficient GMM with only the smaller set Z_a of instruments. Let \tilde{J} denote the resulting GMM criterion. Second, estimate the model by efficient GMM with the full set $Z = (Z_a, Z_b)$ of instruments. Let \hat{J} denote the resulting GMM criterion. The test statistic is the difference in the criterion functions: $C = \hat{J} - \tilde{J}$. This is similar to the GMM distance statistic presented in Section 13.19. The difference is that the distance statistic compares models which differ based on the parameter restrictions while the C statistic compares models based on different instrument sets.

Typically $C \geq 0$. However, this is not necessary and $C < 0$ can arise. If this occurs it leads to a non-rejection of \mathbb{H}_0 .

If the smaller instrument set Z_a is just-identified so that $\ell_a = k$ then $\tilde{J} = 0$ so $C = \hat{J}$ is simply the standard overidentification test. This is why we have restricted attention to the case $\ell_a > k$.

The test has the following large sample distribution.

Theorem 13.15 Under Assumption 12.2 and $\mathbb{E}[Z_a X']$ has full rank k , then as $n \rightarrow \infty$, $C \xrightarrow{d} \chi^2_{\ell_b}$. For c satisfying $\alpha = 1 - G_{\ell_b}(c)$, $\mathbb{P}[C > c | \mathbb{H}_0] \rightarrow \alpha$. The test “Reject \mathbb{H}_0 if $C > c$ ” has asymptotic size α .

The proof of Theorem 13.15 is presented in Section 13.28.

In Stata the command `estat overid zb` after `ivregress gmm` can be used to implement a subset overidentification test where `zb` is the name(s) of the instrument(s) tested for validity. The statistic C and its asymptotic p-value using the $\chi^2_{\ell_2}$ distribution are reported.

13.23 Endogeneity Test

In Section 12.29 we introduced tests for endogeneity in the context of 2SLS estimation. Endogeneity tests are simple to implement in the GMM framework as a subset overidentification test. The model is $Y = Z_1'\beta_1 + Y_2'\beta_2 + e$ where the maintained assumption is that the regressors Z_1 and excluded instruments Z_2 are exogenous so that $\mathbb{E}[Z_1e] = 0$ and $\mathbb{E}[Z_2e] = 0$. The question is whether or not Y_2 is endogenous. The null hypothesis is $\mathbb{H}_0 : \mathbb{E}[Y_2e] = 0$ with the alternative $\mathbb{H}_1 : \mathbb{E}[Y_2e] \neq 0$.

The GMM test is constructed as follows. First, estimate the model by efficient GMM using (Z_1, Z_2) as instruments for (Z_1, Y_2) . Let \tilde{J} denote the resulting GMM criterion. Second, estimate the model by efficient GMM² using (Z_1, Z_2, Y_2) as instruments for (Z_1, Y_2) . Let \hat{J} denote the resulting GMM criterion. The test statistic is the difference in the criterion functions: $C = \hat{J} - \tilde{J}$.

The distribution theory for the test is a special case of overidentification testing.

Theorem 13.16 Under Assumption 12.2 and $\mathbb{E}[Z_2Y_2']$ has full rank k_2 , then as $n \rightarrow \infty$, $C \xrightarrow{d} \chi_{k_2}^2$. For c satisfying $\alpha = 1 - G_{k_2}(c)$, $\mathbb{P}[C > c | \mathbb{H}_0] \rightarrow \alpha$. The test “Reject \mathbb{H}_0 if $C > c$ ” has asymptotic size α .

In Stata the command `estat endogenous` after `ivregress gmm` can be used to implement the test for endogeneity. The statistic C and its asymptotic p-value using the $\chi_{k_2}^2$ distribution are reported.

13.24 Subset Endogeneity Test

In Section 12.30 we introduced subset endogeneity tests for 2SLS estimation. GMM tests are simple to implement as subset overidentification tests. The model is $Y = Z_1'\beta_1 + Y_2'\beta_2 + Y_3'\beta_3 + e$ with $\mathbb{E}[Ze] = 0$ where the instrument vector is $Z = (Z_1, Z_2)$. The $k_3 \times 1$ variables Y_3 are treated as endogenous and the $k_2 \times 1$ variables Y_2 are treated as potentially endogenous. The hypothesis to test is that Y_2 is exogenous, or $\mathbb{H}_0 : \mathbb{E}[Y_2e] = 0$ against $\mathbb{H}_1 : \mathbb{E}[Y_2e] \neq 0$. The test requires that $\ell_2 \geq (k_2 + k_3)$ so that the model can be estimated under \mathbb{H}_1 .

The GMM test is constructed as follows. First, estimate the model by efficient GMM using (Z_1, Z_2) as instruments for (Z_1, Y_2, Y_3) . Let \tilde{J} denote the resulting GMM criterion. Second, estimate the model by efficient GMM using (Z_1, Z_2, Y_2) as instruments for (Z_1, Y_2, Y_3) . Let \hat{J} denote the resulting GMM criterion. The test statistic is the difference in the criterion functions: $C = \hat{J} - \tilde{J}$.

The distribution theory for the test is a special case of the theory of overidentification testing.

Theorem 13.17 Under Assumption 12.2 and $\mathbb{E}[Z_2(Y_2', Y_3')]$ has full rank $k_2 + k_3$, then as $n \rightarrow \infty$, $C \xrightarrow{d} \chi_{k_2}^2$. For c satisfying $\alpha = 1 - G_{k_2}(c)$, $\mathbb{P}[C > c | \mathbb{H}_0] \rightarrow \alpha$. The test “Reject \mathbb{H}_0 if $C > c$ ” has asymptotic size α .

In Stata, the command `estat endogenous x2` after `ivregress gmm` can be used to implement the test for endogeneity where `x2` is the name(s) of the variable(s) tested for endogeneity. The statistic C and its asymptotic p-value using the $\chi_{k_2}^2$ distribution are reported.

²If the homoskedastic weight matrix is used this GMM estimator equals least squares, but when the weight matrix allows for heteroskedasticity the efficient GMM estimator does not equal least squares as the model is overidentified.

13.25 Nonlinear GMM

GMM applies whenever an economic or statistical model implies the $\ell \times 1$ moment condition

$$\mathbb{E}[g_i(\beta)] = 0.$$

where $g_i(\beta)$ is a possibly nonlinear function of the parameters β . Often, this is all that is known. Identification requires $\ell \geq k = \dim(\beta)$. The GMM estimator minimizes

$$J(\beta) = n \bar{g}_n(\beta)' \widehat{\mathbf{W}} \bar{g}_n(\beta)$$

for some weight matrix $\widehat{\mathbf{W}}$ where

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

The efficient GMM estimator can be constructed by setting

$$\widehat{\mathbf{W}} = \left(\frac{1}{n} \sum_{i=1}^n \widehat{g}_i \widehat{g}_i' - \bar{g}_n \bar{g}_n' \right)^{-1},$$

with $\widehat{g}_i = g_i(\widehat{\beta})$ constructed using a preliminary consistent estimator $\widehat{\beta}$, perhaps obtained with $\widehat{\mathbf{W}} = \mathbf{I}_\ell$. As in the case of the linear model the weight matrix can be iterated until convergence to obtain the iterated GMM estimator.

Proposition 13.1 Distribution of Nonlinear GMM Estimator

Under general regularity conditions, $\sqrt{n}(\widehat{\beta}_{\text{gmm}} - \beta) \xrightarrow{d} N(0, \mathbf{V}_\beta)$ where

$$\mathbf{V}_\beta = (\mathbf{Q}' \mathbf{W} \mathbf{Q})^{-1} (\mathbf{Q}' \mathbf{W} \Omega \mathbf{W} \mathbf{Q}) (\mathbf{Q}' \mathbf{W} \mathbf{Q})^{-1}$$

with $\Omega = \mathbb{E}[g_i g_i']$ and

$$\mathbf{Q} = \mathbb{E} \left[\frac{\partial}{\partial \beta'} g_i(\beta) \right].$$

If the efficient weight matrix is used then $\mathbf{V}_\beta = (\mathbf{Q}' \Omega^{-1} \mathbf{Q})^{-1}$.

The proof of this result is omitted as it uses more advanced techniques.

The asymptotic covariance matrices can be estimated by sample counterparts of the population matrices. For the case of a general weight matrix,

$$\widehat{\mathbf{V}}_\beta = (\widehat{\mathbf{Q}}' \widehat{\mathbf{W}} \widehat{\mathbf{Q}})^{-1} (\widehat{\mathbf{Q}}' \widehat{\mathbf{W}} \widehat{\Omega} \widehat{\mathbf{W}} \widehat{\mathbf{Q}}) (\widehat{\mathbf{Q}}' \widehat{\mathbf{W}} \widehat{\mathbf{Q}})^{-1}$$

where

$$\widehat{\Omega} = \frac{1}{n} \sum_{i=1}^n (g_i(\widehat{\beta}) - \bar{g})(g_i(\widehat{\beta}) - \bar{g})'$$

$$\bar{g} = n^{-1} \sum_{i=1}^n g_i(\widehat{\beta})$$

and

$$\hat{\mathbf{Q}} = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta'} g_i(\hat{\beta}).$$

For the case of the iterated efficient weight matrix,

$$\hat{\mathbf{V}}_{\beta} = (\hat{\mathbf{Q}}' \hat{\Omega}^{-1} \hat{\mathbf{Q}})^{-1}.$$

All of the methods discussed in this chapter – Wald tests, constrained estimation, distance tests, overidentification tests, endogeneity tests – apply similarly to the nonlinear GMM estimator.

13.26 Bootstrap for GMM

The bootstrap for 2SLS (Section 12.23) can be used for GMM estimation. The standard bootstrap algorithm generates bootstrap samples by sampling the triplets (Y_i^*, X_i^*, Z_i^*) independently and with replacement from the original sample. The GMM estimator is applied to the bootstrap sample to obtain the bootstrap estimates $\hat{\beta}_{\text{gmm}}^*$. This is repeated B times to create a sample of B bootstrap draws. Given these draws, bootstrap confidence intervals, including percentile, BC percentile, BC_a and percentile-t, are calculated conventionally.

For variance and standard error estimation the same cautions apply as for 2SLS. It is difficult to know if the GMM estimator has a finite variance in a given application. It is best to avoid using the bootstrap to calculate standard errors. Instead, use the bootstrap for percentile and percentile-t confidence intervals.

When the model is overidentified, as discussed for 2SLS, bootstrap GMM inference will not achieve an asymptotic refinement unless the bootstrap estimator is recentered to satisfy the orthogonality condition. We now describe the recentering recommended by Hall and Horowitz (1996).

For linear GMM with weight matrix \mathbf{W} the recentered GMM bootstrap estimator is

$$\hat{\beta}_{\text{gmm}}^{**} = (\mathbf{X}^{*'} \mathbf{Z}^* \mathbf{W}^* \mathbf{Z}^{*'} \mathbf{X}^*)^{-1} (\mathbf{X}^{*'} \mathbf{Z}^* \mathbf{W}^* (\mathbf{Z}^{*'} \mathbf{Y}^* - \mathbf{Z}^{*'} \hat{\mathbf{e}}))$$

where \mathbf{W}^* is the bootstrap version of \mathbf{W} and $\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X} \hat{\beta}_{\text{gmm}}$. For efficient GMM,

$$\mathbf{W}^* = \left(\frac{1}{n} \sum_{i=1}^n Z_i^* Z_i^{*'} (Y_i^* - X_i^{*'} \tilde{\beta}^*)^2 \right)^{-1}$$

for preliminary estimator $\tilde{\beta}^*$.

For nonlinear GMM (Section 13.25) the bootstrap criterion function is modified. The recentered bootstrap criterion is

$$\begin{aligned} J^{**}(\beta) &= n (\bar{g}_n^*(\beta) - \bar{g}_n(\hat{\beta}_{\text{gmm}}))' \mathbf{W}^* (\bar{g}_n^*(\beta) - \bar{g}_n(\hat{\beta}_{\text{gmm}})) \\ \bar{g}_n^*(\beta) &= \frac{1}{n} \sum_{i=1}^n g_i^*(\beta) \end{aligned}$$

where $\bar{g}_n(\hat{\beta}_{\text{gmm}})$ is from the sample not from the bootstrap data. The bootstrap estimator is

$$\hat{\beta}_{\text{gmm}}^{**} = \text{argmin} J^{**}(\beta).$$

The bootstrap can be used to calculate the p-value of the GMM overidentification test. For the GMM estimator with an efficient weight matrix the standard overidentification test is the Hansen J statistic

$$J = n \bar{g}_n(\hat{\beta}_{\text{gmm}})' \hat{\Omega}^{-1} \bar{g}_n(\hat{\beta}_{\text{gmm}}).$$

The recentered bootstrap analog is

$$J^{**} = n \left(\bar{g}_n^*(\hat{\beta}_{\text{gmm}}^{**}) - \bar{g}_n(\hat{\beta}_{\text{gmm}}) \right)' \hat{\Omega}^{*-1} \left(\bar{g}_n^*(\hat{\beta}_{\text{gmm}}^{**}) - \bar{g}_n(\hat{\beta}_{\text{gmm}}) \right).$$

On each bootstrap sample $J^{**}(b)$ is calculated and stored. The bootstrap p-value is

$$p^* = \frac{1}{B} \sum_{b=1}^B \mathbb{1} \{ J^{**}(b) > S \}.$$

This bootstrap p-value is asymptotically valid since J^{**} satisfies the overidentified moment conditions.

13.27 Conditional Moment Equation Models

In many contexts, an economic model implies conditional moment restriction of the form

$$\mathbb{E} [e_i(\beta) | Z_i] = 0$$

where $e_i(\beta)$ is some $s \times 1$ function of the observation and the parameters. In many cases $s = 1$. It turns out that this conditional moment restriction is more powerful than the unconditional moment equation model discussed throughout this chapter.

For example, the linear model $Y = X'\beta + e$ with instruments Z falls into this class under the assumption $\mathbb{E}[e | Z] = 0$. In this case $e_i(\beta) = Y_i - X_i'\beta$.

It is also helpful to realize that conventional regression models also fall into this class except that in this case $X = Z$. For example, in linear regression $e_i(\beta) = Y_i - X_i'\beta$, while in a nonlinear regression model $e_i(\beta) = Y_i - m(X_i, \beta)$. In a joint model of the conditional expectation $\mathbb{E}[Y | X = x] = x'\beta$ and variance $\text{var}[Y | X = x] = f(x)'\gamma$, then

$$e_i(\beta, \gamma) = \begin{cases} Y_i - X_i'\beta \\ (Y_i - X_i'\beta)^2 - f(X_i)'\gamma \end{cases}.$$

Here $s = 2$.

Given a conditional moment restriction an unconditional moment restriction can always be constructed. That is for any $\ell \times 1$ function $\phi(Z, \beta)$ we can set $g_i(\beta) = \phi(Z_i, \beta) e_i(\beta)$ which satisfies $\mathbb{E}[g_i(\beta)] = 0$ and hence defines an unconditional moment equation model. The obvious problem is that the class of functions ϕ is infinite. Which should be selected?

This is equivalent to the problem of selection of the best instruments. If $Z \in \mathbb{R}$ is a valid instrument satisfying $\mathbb{E}[e | Z] = 0$, then Z, Z^2, Z^3, \dots , etc., are all valid instruments. Which should be used?

One solution is to construct an infinite list of potent instruments and then use the first ℓ . How is ℓ to be determined? This is an area of theory still under development. One study of this problem is Donald and Newey (2001).

Another approach is to construct the **optimal instrument** which minimizes the asymptotic variance. The form was uncovered by Chamberlain (1987). Take the case $s = 1$. Let

$$R_i = \mathbb{E} \left[\frac{\partial}{\partial \beta} e_i(\beta) \middle| Z_i \right]$$

and $\sigma_i^2 = \mathbb{E}[e_i(\beta)^2 | Z_i]$. Then the optimal instrument is $A_i = -\sigma_i^{-2} R_i$. The optimal moment is $g_i(\beta) = A_i e_i(\beta)$. Setting $g_i(\beta)$ to be this choice (which is $k \times 1$, so is just-identified) yields the GMM estimator with lowest asymptotic variance. In practice A_i is unknown, but its form helps us think about construction of good instruments.

In the linear model $e_i(\beta) = Y_i - X_i'\beta$ note that $R_i = -\mathbb{E}[X_i | Z_i]$ and $\sigma_i^2 = \mathbb{E}[e_i^2 | Z_i]$. This means the optimal instrument is $A_i = \sigma_i^{-2}\mathbb{E}[X_i | Z_i]$. In the case of linear regression $X_i = Z_i$ so $A_i = \sigma_i^{-2}Z_i$. Hence efficient GMM is equivalent to GLS!

In the case of endogenous variables note that the efficient instrument A_i involves the estimation of the conditional mean of X given Z . In other words, to get the best instrument for X we need the best conditional mean model for X given Z not just an arbitrary linear projection. The efficient instrument is also inversely proportional to the conditional variance of e . This is the same as the GLS estimator; namely that improved efficiency can be obtained if the observations are weighted inversely to the conditional variance of the errors.

13.28 Technical Proofs*

Proof of Theorem 13.12 Set $\tilde{e}_i = Y_i - X_i'\hat{\beta}_{\text{cgmm}}$ and $\hat{e}_i = Y_i - X_i'\hat{\beta}_{\text{gmm}}$. By standard covariance matrix analysis $\hat{\Omega} \xrightarrow{p} \Omega$ and $\tilde{\Omega} \xrightarrow{p} \Omega$. Thus we can replace $\hat{\Omega}$ and $\tilde{\Omega}$ in the criteria without affecting the asymptotic distribution. In particular

$$\begin{aligned} \tilde{J}(\hat{\beta}_{\text{cgmm}}) &= \frac{1}{n} \tilde{e}' \mathbf{Z} \tilde{\Omega}^{-1} \mathbf{Z}' \tilde{e} \\ &= \frac{1}{n} \tilde{e}' \mathbf{Z} \hat{\Omega}^{-1} \mathbf{Z}' \tilde{e} + o_p(1). \end{aligned} \quad (13.25)$$

Now observe that

$$\mathbf{Z}' \tilde{e} = \mathbf{Z}' \hat{e} - \mathbf{Z}' \mathbf{X} (\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}}).$$

Thus

$$\begin{aligned} \frac{1}{n} \tilde{e}' \mathbf{Z} \hat{\Omega}^{-1} \mathbf{Z}' \tilde{e} &= \frac{1}{n} \hat{e}' \mathbf{Z} \hat{\Omega}^{-1} \mathbf{Z}' \hat{e} - \frac{2}{n} (\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}})' \mathbf{X}' \mathbf{Z} \hat{\Omega}^{-1} \mathbf{Z}' \hat{e} \\ &\quad + \frac{1}{n} (\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}})' \mathbf{X}' \mathbf{Z} \hat{\Omega}^{-1} \mathbf{Z}' \mathbf{X} (\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}}) \\ &= \hat{J}(\hat{\beta}_{\text{gmm}}) + \frac{1}{n} (\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}})' \mathbf{X}' \mathbf{Z} \hat{\Omega}^{-1} \mathbf{Z}' \mathbf{X} (\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}}) \end{aligned} \quad (13.26)$$

where the second equality holds because $\mathbf{X}' \mathbf{Z} \hat{\Omega}^{-1} \mathbf{Z}' \hat{e} = 0$ is the first-order condition for $\hat{\beta}_{\text{gmm}}$. By (13.16) and Theorem 13.4, under \mathbb{H}_0

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}}) &= -(\mathbf{X}' \mathbf{Z} \Omega^{-1} \mathbf{Z}' \mathbf{X})^{-1} \mathbf{R} \left(\mathbf{R}' (\mathbf{X}' \mathbf{Z} \Omega^{-1} \mathbf{Z}' \mathbf{X})^{-1} \mathbf{R} \right)^{-1} \mathbf{R}' \sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) + o_p(1) \\ &\xrightarrow{d} (\mathbf{Q}' \Omega^{-1} \mathbf{Q})^{-1} \mathbf{R} \mathbf{Z} \end{aligned} \quad (13.27)$$

where

$$\begin{aligned} \mathbf{Z} &\sim \mathbf{N}(0, \mathbf{V}_R) \\ \mathbf{V}_R &= \left(\mathbf{R} \mathbf{V}' (\mathbf{Q}' \Omega^{-1} \mathbf{Q})^{-1} \mathbf{R} \right)^{-1}. \end{aligned} \quad (13.28)$$

Putting together (13.25), (13.26), (13.27) and (13.28),

$$\begin{aligned} D &= \tilde{J}(\hat{\beta}_{\text{cgmm}}) - \hat{J}(\hat{\beta}_{\text{gmm}}) \\ &= \sqrt{n}(\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}})' \frac{1}{n} \mathbf{X}' \mathbf{Z} \hat{\Omega}^{-1} \frac{1}{n} \mathbf{Z}' \mathbf{X} \sqrt{n}(\hat{\beta}_{\text{cgmm}} - \hat{\beta}_{\text{gmm}}) \\ &\xrightarrow{d} \mathbf{Z}' \mathbf{V}_R^{-1} \mathbf{Z} \sim \chi_q^2 \end{aligned}$$

because $V_R > 0$ and Z is $q \times 1$. ■

Proof of Theorem 13.15 Let $\tilde{\beta}$ denote the GMM estimator obtained with the instrument set Z_a and let $\hat{\beta}$ denote the GMM estimator obtained with the instrument set Z . Set $\tilde{e}_i = Y_i - X_i' \tilde{\beta}$, $\hat{e}_i = Y_i - X_i' \hat{\beta}$,

$$\begin{aligned}\tilde{\Omega} &= n^{-1} \sum_{i=1}^n Z_{ai} Z_{ai}' \tilde{e}_i^2 \\ \hat{\Omega} &= n^{-1} \sum_{i=1}^n Z_i Z_i' \hat{e}_i^2.\end{aligned}$$

Let R be the $\ell \times \ell_a$ selector matrix so that $Z_a = R' Z$. Note that

$$\tilde{\Omega} = R' n^{-1} \sum_{i=1}^n Z_i Z_i' \tilde{e}_i^2 R.$$

By standard covariance matrix analysis, $\hat{\Omega} \xrightarrow{p} \Omega$ and $\tilde{\Omega} \xrightarrow{p} R' \Omega R$. Also, $\frac{1}{n} Z' X \xrightarrow{p} Q$, say. By the CLT, $n^{-1/2} Z' e \xrightarrow{d} Z$ where $Z \sim N(0, \Omega)$. Then

$$\begin{aligned}n^{-1/2} Z' \hat{e} &= \left(I_\ell - \left(\frac{1}{n} Z' X \right) \left(\frac{1}{n} X' Z \hat{\Omega}^{-1} \frac{1}{n} Z' X \right)^{-1} \left(\frac{1}{n} X' Z \right) \hat{\Omega}^{-1} \right) n^{-1/2} Z' e \\ &\xrightarrow{d} \left(I_\ell - Q(Q' \Omega^{-1} Q)^{-1} Q' \Omega^{-1} \right) Z\end{aligned}$$

and

$$\begin{aligned}n^{-1/2} Z'_a \tilde{e} &= R' \left(I_\ell - \left(\frac{1}{n} Z' X \right) \left(\frac{1}{n} X' Z R \tilde{\Omega}^{-1} R' \frac{1}{n} Z' X \right)^{-1} \left(\frac{1}{n} X' Z \right) R \tilde{\Omega}^{-1} R' \right) n^{-1/2} Z' e \\ &\xrightarrow{d} R' \left(I_\ell - Q(Q' R(R' \Omega R)^{-1} R' Q)^{-1} Q' R(R' \Omega R)^{-1} R' \right) Z\end{aligned}$$

jointly.

By linear rotations of Z and R we can set $\Omega = I_\ell$ to simplify the notation. Thus setting $P_Q = Q(Q'Q)^{-1}Q'$, $P_R = R(R'R)^{-1}R'$ and $Z \sim N(0, I_\ell)$ we have

$$\hat{J} \xrightarrow{d} Z' (I_\ell - P_Q) Z$$

and

$$\tilde{J} \xrightarrow{d} Z' (P_R - P_R Q (Q' P_R Q)^{-1} Q' P_R) Z.$$

It follows that

$$C = \hat{J} - \tilde{J} \xrightarrow{d} Z' A Z$$

where

$$A = (I_\ell - P_Q - P_R + P_R Q (Q' P_R Q)^{-1} Q' P_R).$$

This is a quadratic form in a standard normal vector and the matrix A is idempotent (this is straightforward to check). $Z' A Z$ is thus distributed χ_d^2 with degrees of freedom d equal to

$$\begin{aligned}\text{rank}(A) &= \text{tr} \left(I_\ell - P_Q - P_R + P_R Q (Q' P_R Q)^{-1} Q' P_R \right) \\ &= \ell - k - \ell_a + k = \ell_b.\end{aligned}$$

Thus the asymptotic distribution of C is $\chi_{\ell_b}^2$ as claimed. ■

13.29 Exercises

Exercise 13.1 Take the model

$$\begin{aligned} Y &= X'\beta + e \\ \mathbb{E}[Xe] &= 0 \\ e^2 &= Z'\gamma + \eta \\ \mathbb{E}[Z\eta] &= 0. \end{aligned}$$

Find the method of moments estimators $(\hat{\beta}, \hat{\gamma})$ for (β, γ) .

Exercise 13.2 Take the model $Y = X'\beta + e$ with $\mathbb{E}[e | Z] = 0$. Let $\hat{\beta}_{\text{gmm}}$ be the GMM estimator using the weight matrix $W_n = (Z'Z)^{-1}$. Under the assumption $\mathbb{E}[e^2 | Z] = \sigma^2$ show that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 (Q'M^{-1}Q)^{-1})$$

where $Q = \mathbb{E}[ZX']$ and $M = \mathbb{E}[ZZ']$.

Exercise 13.3 Take the model $Y = X'\beta + e$ with $\mathbb{E}[Ze] = 0$. Let $\tilde{e}_i = Y_i - X_i'\tilde{\beta}$ where $\tilde{\beta}$ is consistent for β (e.g. a GMM estimator with some weight matrix). An estimator of the optimal GMM weight matrix is

$$\widehat{W} = \left(\frac{1}{n} \sum_{i=1}^n Z_i Z_i' \tilde{e}_i^2 \right)^{-1}.$$

Show that $\widehat{W} \xrightarrow{p} \Omega^{-1}$ where $\Omega = \mathbb{E}[ZZ'e^2]$.

Exercise 13.4 In the linear model estimated by GMM with general weight matrix W the asymptotic variance of $\hat{\beta}_{\text{gmm}}$ is

$$V = (Q'WQ)^{-1} Q'W\Omega WQ (Q'WQ)^{-1}.$$

- Let V_0 be this matrix when $W = \Omega^{-1}$. Show that $V_0 = (Q'\Omega^{-1}Q)^{-1}$.
- We want to show that for any W , $V - V_0$ is positive semi-definite (for then V_0 is the smaller possible covariance matrix and $W = \Omega^{-1}$ is the efficient weight matrix). To do this start by finding matrices A and B such that $V = A'\Omega A$ and $V_0 = B'\Omega B$.
- Show that $B'\Omega A = B'\Omega B$ and therefore that $B'\Omega(A - B) = 0$.
- Use the expressions $V = A'\Omega A$, $A = B + (A - B)$, and $B'\Omega(A - B) = 0$ to show that $V \geq V_0$.

Exercise 13.5 Prove Theorem 13.8.

Exercise 13.6 Derive the constrained GMM estimator (13.16).

Exercise 13.7 Show that the constrained GMM estimator (13.16) with the efficient weight matrix is (13.19).

Exercise 13.8 Prove Theorem 13.9.

Exercise 13.9 Prove Theorem 13.10.

Exercise 13.10 The equation of interest is $Y = m(X, \beta) + e$ with $\mathbb{E}[Ze] = 0$ where $m(x, \beta)$ is a known function, β is $k \times 1$ and Z is $\ell \times 1$. Show how to construct an efficient GMM estimator for β .

Exercise 13.11 As a continuation of Exercise 12.7 derive the efficient GMM estimator using the instrument $Z = (X' \ X^2)'$. Does this differ from 2SLS and/or OLS?

Exercise 13.12 In the linear model $Y = X'\beta + e$ with $\mathbb{E}[Xe] = 0$ the GMM criterion function for β is

$$J(\beta) = \frac{1}{n} (Y - X\beta)' X\hat{\Omega}^{-1} X' (Y - X\beta) \quad (13.29)$$

where $\hat{\Omega} = n^{-1} \sum_{i=1}^n X_i X_i' \hat{e}_i^2$, $\hat{e}_i = Y_i - X_i' \hat{\beta}$ are the OLS residuals, and $\hat{\beta} = (X'X)^{-1} X'Y$ is least squares. The GMM estimator of β subject to the restriction $r(\beta) = 0$ is

$$\tilde{\beta} = \underset{r(\beta)=0}{\operatorname{argmin}} J_n(\beta).$$

The GMM test statistic (the distance statistic) of the hypothesis $r(\beta) = 0$ is

$$D = J(\tilde{\beta}) = \min_{r(\beta)=0} J(\beta). \quad (13.30)$$

(a) Show that you can rewrite $J(\beta)$ in (13.29) as

$$J(\beta) = n(\beta - \hat{\beta})' \hat{V}_\beta^{-1} (\beta - \hat{\beta})$$

and thus $\tilde{\beta}$ is the same as the minimum distance estimator.

(b) Show that under linear hypotheses the distance statistic D in (13.30) equals the Wald statistic.

Exercise 13.13 Take the linear model $Y = X'\beta + e$ with $\mathbb{E}[Ze] = 0$. Consider the GMM estimator $\hat{\beta}$ of β . Let $J = n \bar{g}_n(\hat{\beta})' \hat{\Omega}^{-1} \bar{g}_n(\hat{\beta})$ denote the test of overidentifying restrictions. Show that $J \xrightarrow{d} \chi_{\ell-k}^2$ as $n \rightarrow \infty$ by demonstrating each of the following.

(a) Since $\Omega > 0$, we can write $\Omega^{-1} = CC'$ and $\Omega = C'^{-1}C^{-1}$ for some matrix C .

(b) $J = n(C' \bar{g}_n(\hat{\beta}))' (C' \hat{\Omega} C)^{-1} C' \bar{g}_n(\hat{\beta})$.

(c) $C' \bar{g}_n(\hat{\beta}) = D_n C' \bar{g}_n(\beta)$ where $\bar{g}_n(\beta) = \frac{1}{n} Z' e$ and

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

(d) $D_n \xrightarrow{p} I_\ell - R(R'R)^{-1} R'$ where $R = C' \mathbb{E}[ZX']$.

(e) $n^{1/2} C' \bar{g}_n(\beta) \xrightarrow{d} u \sim N(0, I_\ell)$.

(f) $J \xrightarrow{d} u' (I_\ell - R(R'R)^{-1} R') u$.

(g) $u' (I_\ell - R(R'R)^{-1} R') u \sim \chi_{\ell-k}^2$.

Hint: $I_\ell - R(R'R)^{-1} R'$ is a projection matrix.

Exercise 13.14 Take the model $Y = X'\beta + e$ with $\mathbb{E}[Ze] = 0$, $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$, $Z \in \mathbb{R}^\ell$, $\ell \geq k$. Consider the statistic

$$J(\beta) = n\bar{m}_n(\beta)'W\bar{m}_n(\beta)$$

$$\bar{m}_n(\beta) = \frac{1}{n} \sum_{i=1}^n Z_i(Y_i - X_i'\beta)$$

for some weight matrix $W > 0$.

- Take the hypothesis $\mathbb{H}_0 : \beta = \beta_0$. Derive the asymptotic distribution of $J(\beta_0)$ under \mathbb{H}_0 as $n \rightarrow \infty$.
- What choice for W yields a known asymptotic distribution in part (a)? (Be specific about degrees of freedom.)
- Write down an appropriate estimator \widehat{W} for W which takes advantage of \mathbb{H}_0 . (You do not need to demonstrate consistency or unbiasedness.)
- Describe an asymptotic test of \mathbb{H}_0 against $\mathbb{H}_1 : \beta \neq \beta_0$ based on this statistic.
- Use the result in part (d) to construct a confidence region for β . What can you say about the form of this region? For example, does the confidence region take the form of an ellipse, similar to conventional confidence regions?

Exercise 13.15 Consider the model $Y = X'\beta + e$ with $\mathbb{E}[Ze] = 0$ and

$$R'\beta = 0 \tag{13.31}$$

with $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$, $Z \in \mathbb{R}^\ell$, $\ell > k$. The matrix R is $k \times q$ with $1 \leq q < k$. You have a random sample $(Y_i, X_i, Z_i : i = 1, \dots, n)$.

For simplicity, assume the efficient weight matrix $W = (\mathbb{E}[ZZ'e^2])^{-1}$ is known.

- Write out the GMM estimator $\widehat{\beta}$ ignoring constraint (13.31).
- Write out the GMM estimator $\widetilde{\beta}$ adding the constraint (13.31).
- Find the asymptotic distribution of $\sqrt{n}(\widetilde{\beta} - \beta)$ as $n \rightarrow \infty$ under Assumption (13.31).

Exercise 13.16 The observed data is $\{Y_i, X_i, Z_i\} \in \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^\ell$, $k > 1$ and $\ell > k > 1$, $i = 1, \dots, n$. The model is $Y = X'\beta + e$ with $\mathbb{E}[Ze] = 0$.

- Given a weight matrix $W > 0$ write down the GMM estimator $\widehat{\beta}$ for β .
- Suppose the model is misspecified. Specifically, assume that for some $\delta \neq 0$,

$$e = \delta n^{-1/2} + u \tag{13.32}$$

$$\mathbb{E}[u | Z] = 0$$

with $\mu_Z = \mathbb{E}[Z] \neq 0$. Show that (13.32) implies that $\mathbb{E}[Ze] \neq 0$.

- Express $\sqrt{n}(\widehat{\beta} - \beta)$ as a function of W , n , δ , and the variables (X_i, Z_i, u_i) .
- Find the asymptotic distribution of $\sqrt{n}(\widehat{\beta} - \beta)$ under Assumption (13.32).

Exercise 13.17 The model is $Y = Z\beta + X\gamma + e$ with $\mathbb{E}[e | Z] = 0$, $X \in \mathbb{R}$ and $Z \in \mathbb{R}$. X is potentially endogenous and Z is exogenous. Someone suggests estimating (β, γ) by GMM using the pair (Z, Z^2) as instruments. Is this feasible? Under what conditions is this a valid estimator?

Exercise 13.18 The observations are i.i.d., $(Y_i, X_i, Q_i : i = 1, \dots, n)$, where X is $k \times 1$ and Q is $m \times 1$. The model is $Y = X'\beta + e$ with $\mathbb{E}[Xe] = 0$ and $\mathbb{E}[Qe] = 0$. Find the efficient GMM estimator for β .

Exercise 13.19 You want to estimate $\mu = \mathbb{E}[Y]$ under the assumption that $\mathbb{E}[X] = 0$, where Y and X are scalar and observed from a random sample. Find an efficient GMM estimator for μ .

Exercise 13.20 Consider the model $Y = X'\beta + e$ given $\mathbb{E}[Ze] = 0$ and $R'\beta = 0$. The dimensions are $X \in R^k$ and $Z \in R^\ell$ with $\ell > k$. The matrix R is $k \times q$, $1 \leq q < k$. Derive an efficient GMM estimator for β .

Exercise 13.21 Take the linear equation $Y = X'\beta + e$ and consider the following estimators of β .

1. $\hat{\beta}$: 2SLS using the instruments Z_1 .
2. $\tilde{\beta}$: 2SLS using the instruments Z_2 .
3. $\bar{\beta}$: GMM using the instruments $Z = (Z_1, Z_2)$ and the weight matrix

$$W = \begin{pmatrix} (Z_1'Z_1)^{-1}\lambda & 0 \\ 0 & (Z_2'Z_2)^{-1}(1-\lambda) \end{pmatrix}$$

for $\lambda \in (0, 1)$.

Find an expression for $\bar{\beta}$ which shows that it is a specific weighted average of $\hat{\beta}$ and $\tilde{\beta}$.

Exercise 13.22 Consider the just-identified model $Y = X_1'\beta_1 + X_2'\beta_2 + e$ with $\mathbb{E}[Ze] = 0$ where $X = (X_1' X_2')' \in \mathbb{R}^k$ and $Z \in \mathbb{R}^k$. We want to test $\mathbb{H}_0 : \beta_1 = 0$. Three econometricians are called for advice.

- Econometrician 1 proposes testing \mathbb{H}_0 by a Wald statistic.
- Econometrician 2 suggests testing \mathbb{H}_0 by the GMM Distance Statistic.
- Econometrician 3 suggests testing \mathbb{H}_0 using the test of overidentifying restrictions.

You are asked to settle this dispute. Explain the advantages and/or disadvantages of the different procedures in this specific context.

Exercise 13.23 Take the model $Y = X'\beta + e$ with $\mathbb{E}[Xe] = 0$ and $\beta = Q\theta$, where β is $k \times 1$, Q is $k \times m$ with $m < k$, Q is known, and θ is $m \times 1$. The observations (Y_i, X_i) are i.i.d. across $i = 1, \dots, n$.

Under these assumptions what is the efficient estimator of θ ?

Exercise 13.24 Take the model $Y = \theta + e$ with $\mathbb{E}[Xe] = 0$, $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$ and (Y_i, X_i) a random sample.

- (a) Find the efficient GMM estimator of θ .
- (b) Is this model over-identified or just-identified?
- (c) Find the GMM test statistic for over-identification.

Exercise 13.25 Take the model $Y = X'\beta + e$ with $\mathbb{E}[Xe] = 0$ where X contains an intercept so $\mathbb{E}[e] = 0$. An enterprising econometrician notices that this implies the n moment conditions

$$\mathbb{E}[e_i] = 0, i = 1, \dots, n.$$

Given an $n \times n$ weight matrix W , this implies a GMM criterion

$$J(\beta) = (Y - X\beta)' W (Y - X\beta).$$

- Under i.i.d. sampling, show that the efficient weight matrix is $W = \sigma^{-2} I_n$ where $\sigma^2 = \mathbb{E}[e^2]$.
- Using the weight matrix $W = \sigma^{-2} I_n$ find the GMM estimator $\hat{\beta}$ that minimizes $J(\beta)$.
- Find a simple expression for the minimized criteria $J(\hat{\beta})$.
- Theorem 13.14 says that criterion such as $J(\hat{\beta})$ are asymptotically $\chi^2_{\ell-k}$ where ℓ is the number of moments. While the assumptions of Theorem 13.14 do not apply to this context, what is ℓ here? That is, which χ^2 distribution is the asserted asymptotic distribution?
- Does the answer in (d) make sense? Explain your reasoning.

Exercise 13.26 Take the model $Y = X'\beta + e$ with $\mathbb{E}[e | X] = 0$ and $\mathbb{E}[e^2 | X] = \sigma^2$. An econometrician more enterprising than the one in previous question notices that this implies the nk moment conditions

$$\mathbb{E}[X_i e_i] = 0, i = 1, \dots, n.$$

We can write the moments using matrix notation as $\mathbb{E}[\bar{X}'(Y - X\beta)]$ where

$$\bar{X} = \begin{pmatrix} X'_1 & 0 & \cdots & 0 \\ 0 & X'_2 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & X'_n \end{pmatrix}.$$

Given an $nk \times nk$ weight matrix W this implies a GMM criterion

$$J(\beta) = (Y - X\beta)' \bar{X} W \bar{X}' (Y - X\beta).$$

- Calculate $\Omega = \mathbb{E}[\bar{X}' e e' \bar{X}]$.
- The econometrician decides to set $W = \Omega^-$, the Moore-Penrose generalized inverse of Ω . (See Section A.6.) Note: A useful fact is that for a vector a , $(aa')^- = aa'(a'a)^{-2}$.
- Find the GMM estimator $\hat{\beta}$ that minimizes $J(\beta)$.
- Find a simple expression for the minimized criterion $J(\hat{\beta})$.
- Comment on whether the χ^2 approximation from Theorem 13.14 is appropriate for $J(\hat{\beta})$.

Exercise 13.27 Continuation of Exercise 12.22, based on the empirical work reported in Acemoglu, Johnson, and Robinson (2001).

- Re-estimate the model estimated in part (j) by efficient GMM. Use the 2SLS estimates as the first-step for the weight matrix and then calculate the GMM estimator using this weight matrix without further iteration. Report the estimates and standard errors.

- (b) Calculate and report the J statistic for overidentification.
- (c) Compare the GMM and 2SLS estimates. Discuss your findings.

Exercise 13.28 Continuation of Exercise 12.24, which involved estimation of a wage equation by 2SLS.

- (a) Re-estimate the model in part (a) by efficient GMM. Do the results change meaningfully?
- (b) Re-estimate the model in part (d) by efficient GMM. Do the results change meaningfully?
- (c) Report the J statistic for overidentification.