Chapter 11

Generalized Method of Moments

Generalized method of moments (GMM) (Hansen, 1982) is an estimation principle that extends *method of moments*. It seeks the parameter value that minimizes a quadratic form of the moments. It is particularly useful in estimating structural economic models in which moment conditions can be derived from underlying economic theory. GMM emerges as one of the most popular estimators in modern econometrics. It includes conventional methods like the two-stage least squares (2SLS) and the three-stage least square as special cases.

11.1 Instrumental Regression

We first discuss estimation in a linear single structural equation

$$y_i = x_i'\beta + \epsilon_i$$

with K regressors. Identification is a prerequisite for structural estimation. From now on we always assume that the model is identified: there is an $L \times 1$ vector of instruments z_i such that $\mathbb{E}[z_i\epsilon_i] = 0_L$ and $\Sigma := \mathbb{E}\left[z_i x_i'\right]$ is of full column rank. Denote β_0 as the root of the equation $E\left[z_i\left(y_i - x_i'\beta\right)\right] = 0_L$, which is uniquely identified.

11.1.1 Just-identification

When $L = K$, the instrumental regression model is *just-identified*, or *exactly identified*. The orthogonality condition implies

$$\Sigma\beta_0 = \mathbb{E}\left[z_i y_i\right],$$

and we can solve express β_0 as

$$\beta_0 = \Sigma^{-1}\mathbb{E}\left[z_i y_i\right] \tag{11.1}$$

in closed form.

The closed-form solution naturally motivates an estimator in which we replace the population methods by the sample moments and this is a method-of-moments estimator. Nevertheless, we postpone the discussion of this estimator to the next section.

11.1.2 Over-identification

When $L > K$, the model is *over-identified*. The orthogonality condition still implies

$$\Sigma\beta_0 = \mathbb{E}\left[z_i y_i\right], \tag{11.2}$$

but Σ is not a square matrix so we cannot write β_0 as that in (11.1). In order to express β_0 explicitly, we define a criterion function

$$Q\left(\beta\right) = \mathbb{E}\left[z_i\left(y_i - x_i\beta\right)\right]' W \mathbb{E}\left[z_i\left(y_i - x_i\beta\right)\right],$$

where W is an $L \times L$ positive-definite non-random symmetric matrix. (The choice of W will be discussed soon.) Because of the quadratic form, $Q\left(\beta\right) \geq 0$ for all β. Identification indicates that $Q\left(\beta\right) = 0$ if and only if $\beta = \beta_0$. Therefore we conclude

$$\beta_0 = \arg\min_{\beta} Q\left(\beta\right)$$

is the unique minimizer. Since $Q(\beta)$ is a smooth function of β, the minimizer β_0 can be characterized by the first-order condition

$$0_K = \frac{\partial}{\partial \beta} Q(\beta_0) = -2\Sigma' W \mathbb{E}\left[z_i \left(y_i - x_i \beta_0\right)\right]$$

Rearranging the above equation, we have

$$\Sigma' W \Sigma \beta_0 = \Sigma' W \mathbb{E}\left[z_i y_i\right].$$

Under the rank condition $\Sigma' W \Sigma$ is invertible so that we can solve

$$\beta_0 = \left(\Sigma' W \Sigma\right)^{-1} \Sigma' W \mathbb{E}\left[z_i y_i\right]. \tag{11.3}$$

Because we have more moments (L) than the number of unknown parameters (K), we call it the *generalized* method of moments.

Remark 11.1. The above equation can be derived by pre-multiplying $\Sigma' W$ on the both sides of (11.2) without referring to the minimization problem.

Remark 11.2. Although we separate the discussion of the just-identified case and the over-identified case, the latter (11.3) actually takes (11.1) as a special case. In this sense, GMM is genuine generalization of the method of moments. to see this point, notice that when $L = K$, given any W we have

$$\begin{aligned}\beta_0 &= \left(\Sigma' W \Sigma\right)^{-1} \Sigma' W \mathbb{E}\left[z_i y_i\right] = \Sigma^{-1} W^{-1} (\Sigma')^{-1} \Sigma' W \mathbb{E}\left[z_i y_i\right] \\ &= \Sigma^{-1} W^{-1} W \mathbb{E}\left[z_i y_i\right] = \Sigma^{-1} \mathbb{E}\left[z_i y_i\right],\end{aligned}$$

as Σ is a square matrix. That is to say, in the just-identified case W plays no role because any choices of W lead to the same explicit solution of β_0.

11.2 GMM Estimator

In practice, we use the sample moments to replace the corresponding population moments. The GMM estimator mimics its population formula.

$$\begin{aligned}\widehat{\beta} &= \left(\frac{1}{n}\sum x_i z_i' W \frac{1}{n}\sum z_i x_i'\right)^{-1} \frac{1}{n}\sum x_i z_i' W \frac{1}{n}\sum z_i y_i \\ &= \left(\frac{X'Z}{n} W \frac{Z'X}{n}\right)^{-1} \frac{X'Z}{n} W \frac{Z'y}{n} \\ &= \left(X'ZWZ'X\right)^{-1} X'ZWZ'y.\end{aligned}$$

Under just-identification, this expression includes the 2SLS estimator

$$\hat{\beta} = \left(\frac{Z'X}{n}\right)^{-1} \frac{Z'y}{n} = \left(Z'X\right)^{-1} Z'y$$

as a special case.

Exercise 11.1. The same GMM estimator $\hat{\beta}$ can be obtained by minimizing

$$Q_n(\beta) = \left[\frac{1}{n}\sum_{i=1}^{n} z_i\left(y_i - x_i\beta\right)\right]' W \left[\frac{1}{n}\sum_{i=1}^{n} z_i\left(y_i - x_i\beta\right)\right] = \frac{\left(y - X\beta\right)' Z}{n} W \frac{Z'\left(y - X\beta\right)}{n},$$

or more concisely $\hat{\beta} = \arg\min_{\beta}\left(y - X\beta\right)' ZWZ'\left(y - X\beta\right)$.

Now we check the asymptotic properties of $\widehat{\beta}$. A few assumptions are in order.

Assumption 11.1 (A.1). *$Z'X/n \xrightarrow{\text{p}} \Sigma$ and $Z'\epsilon/n \xrightarrow{\text{p}} 0_L$.*

A.1 assumes that we can apply a law of large numbers, so that that the sample moments $Z'X/n$ and $Z'\epsilon/n$ converge in probability to their population counterparts.

Theorem 11.1. *Under Assumption A.1, $\widehat{\beta}$ is consistent.*

Proof. The step is similar to the consistency proof of OLS.

$$\begin{aligned}\widehat{\beta} &= \left(X'ZWZ'X\right)^{-1} X'ZWZ'\left(X'\beta_0 + \epsilon\right) \\ &= \beta_0 + \left(\frac{X'Z}{n} W \frac{Z'X}{n}\right)^{-1} \frac{X'Z}{n} W \frac{Z'\epsilon}{n} \\ &\xrightarrow{\mathrm{p}} \beta_0 + \left(\Sigma'W\Sigma\right)^{-1}\Sigma'W0 = \beta_0.\end{aligned}$$ □

To check asymptotic normality, we assume that a central limit theorem can be applied.

Assumption 11.2 (A.2). $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i\epsilon_i \xrightarrow{d} N\left(0_L, \Omega\right)$, *where* $\Omega = \mathbb{E}\left[z_i z_i'\epsilon_i^2\right]$.

Theorem 11.2 (Asymptotic Normality)**.** *Under Assumptions A.1 and A.2,*

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{d} N\left(0_K, \left(\Sigma'W\Sigma\right)^{-1}\Sigma'W\Omega W\Sigma\left(\Sigma'W\Sigma\right)^{-1}\right). \tag{11.4}$$

Proof. Multiply $\widehat{\beta} - \beta_0$ by the scaling factor $\sqrt{n}$,

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) = \left(\frac{X'Z}{n} W \frac{Z'X}{n}\right)^{-1} \frac{X'Z}{n} W \frac{Z'\epsilon}{\sqrt{n}} = \left(\frac{X'Z}{n} W \frac{Z'X}{n}\right)^{-1} \frac{X'Z}{n} W \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i'\epsilon_i.$$

The conclusion follows by the Slutsky's theorem as

$$\frac{X'Z}{n} W \frac{Z'X}{n} \xrightarrow{\mathrm{p}} \Sigma'W\Sigma$$

and

$$\frac{X'Z}{n} W \frac{1}{\sqrt{n}} \sum z_i'\epsilon_i \xrightarrow{d} \Sigma'W \times N\left(0, \Omega\right) \sim N\left(0, \Sigma'W\Omega W\Sigma\right).$$ □

11.2.1 Efficient GMM

It is clear from (11.4) that the GMM estimator's asymptotic variance depends on the choice of W. Which W makes the asymptotic variance as small as possible? The answer

is $W = \Omega^{-1}$, under which the efficient asymptotic variance is

$$\left(\Sigma'\Omega^{-1}\Sigma\right)^{-1}\Sigma'\Omega^{-1}\Omega\Omega^{-1}\Sigma\left(\Sigma'\Omega^{-1}\Sigma\right)^{-1} = \left(\Sigma'\Omega^{-1}\Sigma\right)^{-1}.$$

Proposition 11.1. *For any positive definite symmetric matrix W, the difference*

$$\left(\Sigma'W\Sigma\right)^{-1}\Sigma'W\Omega W\Sigma\left(\Sigma'W\Sigma\right)^{-1} - \left(\Sigma'\Omega^{-1}\Sigma\right)^{-1}$$

is positive semi-definite.

Proof. To simplify notation, denote $A := W\Sigma\left(\Sigma'W\Sigma\right)^{-1}$ and $B := \Omega^{-1}\Sigma\left(\Sigma'\Omega^{-1}\Sigma\right)^{-1}$ and then the difference of the two matrices becomes

$$\begin{aligned}
& \left(\Sigma'W\Sigma\right)^{-1}\Sigma'W\Omega W\Sigma\left(\Sigma'W\Sigma\right)^{-1} - \left(\Sigma'\Omega^{-1}\Sigma\right)^{-1} \\
= \;& A'\Omega A - B'\Omega B \\
= \;& (A - B + B)'\,\Omega\,(A - B + B) - B'\Omega B \\
= \;& (A - B)'\,\Omega\,(A - B) + (A - B)'\,\Omega B + B'\Omega\,(A - B).
\end{aligned}$$

Notice that

$$\begin{aligned}
B'\Omega A &= \left(\Sigma'\Omega^{-1}\Sigma\right)^{-1}\Sigma'\Omega^{-1}\Omega W\Sigma\left(\Sigma'W\Sigma\right)^{-1} \\
&= \left(\Sigma'\Omega^{-1}\Sigma\right)^{-1}\Sigma'W\Sigma\left(\Sigma'W\Sigma\right)^{-1} = \left(\Sigma'\Omega^{-1}\Sigma\right)^{-1} = B'\Omega B,
\end{aligned}$$

which implies $B'\Omega\,(A - B) = 0$ and $(A - B)'\,\Omega B = 0$. We thus conclude that

$$\left(\Sigma'W\Sigma\right)^{-1}\Sigma'W\Omega W\Sigma\left(\Sigma'W\Sigma\right)^{-1} - \left(\Sigma'\Omega^{-1}\Sigma\right)^{-1} = (A - B)'\,\Omega\,(A - B)$$

is positive semi-definite. □

11.2.2 Two-Step GMM

The *two-step GMM* is one way to construct a feasible efficient GMM estimator.

1. Choose any valid W, say $W = I_L$, to get a consistent (but inefficient in general) estimator $\hat{\beta}^\sharp = \hat{\beta}^\sharp(W)$. Save the residual $\widehat{\epsilon}_i = y_i - x_i'\hat{\beta}^\sharp$ and estimate the variance matrix $\widehat{\Omega} = \frac{1}{n}\sum z_i z_i' \widehat{\epsilon}_i^2$. Notice that this $\widehat{\Omega}$ is a consistent for Ω.

2. Set $W = \widehat{\Omega}^{-1}$ and obtain the second estimator

$$\widehat{\beta}^\natural = \widehat{\beta}^\natural(\widehat{\Omega}^{-1}) = \left(X'Z\widehat{\Omega}^{-1}Z'X\right)^{-1} X'Z\widehat{\Omega}^{-1}Z'y.$$

   This second estimator is asymptotic efficient.

Exercise 11.2. Show that if $\widehat{\Omega} \xrightarrow{p} \Omega$, then $\sqrt{n}\left(\widehat{\beta}^\natural(\widehat{\Omega}^{-1}) - \widehat{\beta}\left(\Omega^{-1}\right)\right) \xrightarrow{p} 0$. In other words, the feasible estimator $\widehat{\beta}^\natural(\widehat{\Omega}^{-1})$ is asymptotically equivalent to the infeasible efficient estimator $\widehat{\beta}\left(\Omega^{-1}\right)$.

11.2.3 Two Stage Least Squares

If we further assume conditional homoskedasticity $\mathbb{E}\left[\epsilon_i^2|z_i\right] = \sigma^2$, then

$$\Omega = \mathbb{E}\left[z_i z_i' \epsilon_i^2\right] = \mathbb{E}\left[z_i z_i' \mathbb{E}\left[\epsilon_i^2|z_i\right]\right] = \sigma^2 \mathbb{E}\left[z_i z_i'\right].$$

In the first-step of the two-step GMM we can estimate the variance of the error term by $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \widehat{\epsilon}_i^2$ and the variance matrix by $\widehat{\Omega} = \widehat{\sigma}^2 \frac{1}{n}\sum_{i=1}^n z_i z_i' = \widehat{\sigma}^2 Z'Z/n$. When we plug this $W = \widehat{\Omega}^{-1}$ into the GMM estimator,

$$\begin{aligned}
\widehat{\beta} &= \left(X'Z\left(\widehat{\sigma}^2\frac{Z'Z}{n}\right)^{-1}Z'X\right)^{-1} X'Z\left(\widehat{\sigma}^2\frac{Z'Z}{n}\right)^{-1}Z'y \\
&= \left(X'Z\left(Z'Z\right)^{-1}Z'X\right)^{-1} X'Z\left(Z'Z\right)^{-1}Z'y.
\end{aligned}$$

This is exactly the same expression of 2SLS for $L > K$. Therefore, 2SLS can be viewed as a special case of GMM with the weighting matrix $(Z'Z/n)^{-1}$. Under conditional homoskedasticity, 2SLS is the efficient estimator. 2SLS is inefficient in general cases of heteroskedasticity, despite its popularity.

Remark 11.3. 2SLS gets its name because it can be obtained using two steps: first regress X on all instruments Z, and then regress y on the fitted value along with the included exogenous variables. However, 2SLS can actually be obtained by one step using the above equation. It is a special case of GMM.

Remark 11.4. If an efficient estimator is not too difficult to implement, an econometric theorist would prefer the efficient estimator to an inefficient estimator. The benefits of using the efficient estimator is not limited to more accurate coefficient estimation. Many specification tests, for example the J-statistic we will introduce soon, count on the efficient estimator to lead to a familiar χ^2 distribution under null hypotheses. Otherwise their null asymptotic distributions will be non-standard and thereby critical values must be found by Monte Carlo simulations.

11.3 GMM in Nonlinear Model

The principle of GMM can be used in models where the parameter enters the moment conditions nonlinearly. Let $g_i(\beta) = g(w_i, \beta) \mapsto \mathbb{R}^L$ be a function of the data w_i and the parameter β. If economic theory implies $\mathbb{E}[g_i(\beta)] = 0$, which the statisticians call the *estimating equations*, we can write the GMM population criterion function as

$$Q(\beta) = \mathbb{E}[g_i(\beta)]' W \mathbb{E}[g_i(\beta)]$$

Example 11.1. Nonlinear models nest the linear model as a special case. For the linear IV model in the previous section, the data is $w_i = (y_i, x_i, z_i)$, and the moment function is

$g(w_i, \beta) = z_i'(y_i - x_i\beta)$.

In practice we use the sample moments to mimic the population moments in the criterion function

$$Q_n(\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\beta)\right)' W \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\beta)\right).$$

The GMM estimator is defined as

$$\hat{\beta} = \arg\min_{\beta} Q_n(\beta).$$

In these nonlinear models, a closed-form solution is in general unavailable, while the asymptotic properties can still be established. We state these asymptotic properties without proofs.

Theorem 11.3. *(a) If the model is identified, and*

$$\mathbb{P}\left\{\sup_{\beta\in\mathcal{B}} \left|\frac{1}{n}\sum_{i=1}^{n} g_i(\beta) - \mathbb{E}\left[g_i(\beta)\right]\right| > \varepsilon\right\} \to 0$$

for any constant $\varepsilon > 0$ where the parametric space $\mathcal{B}$ is a closed set, then $\hat{\beta} \xrightarrow{\mathrm{p}} \beta$.
(b) If in addition $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\beta_0) \xrightarrow{d} N(0, \Omega)$ and $\Sigma = \mathbb{E}\left[\frac{\partial}{\partial\beta'} g_i(\beta_0)\right]$ is of full column rank, then

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N\left(0, \left(\Sigma' W \Sigma\right)^{-1}\left(\Sigma' W \Omega W \Sigma\right)\left(\Sigma' W \Sigma\right)^{-1}\right)$$

where $\Omega = \mathbb{E}\left[g_i(\beta_0) g_i(\beta_0)'\right]$.
(c) If we choose $W = \Omega^{-1}$, then the GMM estimator is efficient, and the asymptotic variance becomes $\left(\Sigma'\Omega^{-1}\Sigma\right)^{-1}$.

Remark 11.5. The list of assumptions in the above statement is incomplete. We only lay out the key conditions but neglect some technical details.

$Q_n(\beta)$ measures how close are the moments to zeros. It can serve as a test statistic with proper scaling. Under the null hypothesis $\mathbb{E}\left[g_i(\beta)\right] = 0_L$, this Sargan-Hansen J-test

checks whether a moment condition is violated. The test statistic is

$$J\left(\widehat{\beta}\right) = nQ_n\left(\widehat{\beta}\right) = n\left(\frac{1}{n}\sum_{i=1}^{n} g_i\left(\widehat{\beta}\right)\right)' \widehat{\Omega}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g_i\left(\widehat{\beta}\right)\right)$$
$$= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i\left(\widehat{\beta}\right)\right)' \widehat{\Omega}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i\left(\widehat{\beta}\right)\right)$$

where $\widehat{\Omega}$ is a consistent estimator of Ω, and $\widehat{\beta}$ is an efficient estimator, for example, the two-step GMM estimator $\widehat{\beta}^{\natural}(\widehat{\Omega}^{-1})$. This statistic converges in distribution to a χ^2 random variable with degree of freedom $L - K$. That is, under the null,

$$J\left(\widehat{\beta}\right) \xrightarrow{d} \chi^2\left(L - K\right).$$

If the null hypothesis is false, then the test statistic tends to be large and it is more likely to reject the null.

11.4 Summary

The popularity of GMM in econometrics comes from the fact that economic theory is often not informative enough about the underlying parametric relationship amongst the variables. Instead, many economic assumptions suggest moment restrictions. From example, the *efficient market hypothesis* postulates that the future price movement Δp_{t+1} cannot be predicted by available past information set $\mathscr{I}_t$ so that $\mathbb{E}\left[\Delta p_{t+1}|\mathscr{I}_t\right] = 0$. It implies that any functions of the variables in the information set $\mathscr{I}_t$ are orthogonal to Δp_{t+1}. A plethora of moment conditions can be constructed in order to test the efficient market hypothesis.

Conceptually simple though, GMM has many practical issues in reality. There has been vast econometric literature about issues of GMM and their remedies.

Historical notes: 2SLS was attributed to Theil (1953). In the linear IV model, the J-statistic was proposed by Sargan (1958), and Hansen (1982) extended it to nonlinear

models.

Further reading: The quadratic form of GMM makes it difficult to accommodate many moments in the big data problems. *Empirical likelihood* is an alternative estimator to GMM to estimate models defined by moment restrictions. Shi (2016) solves the estimation problem of high-dimensional moments under the framework of empirical likelihood.

Zhentao Shi. Dec 3, 2020.

Bibliography

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029–1054.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, 393–415.

Shi, Z. (2016). Econometric estimation with high-dimensional moment equalities. *Journal of Econometrics 195*(1), 104–119.

Theil, H. (1953). Repeated least squares applied to complete equation systems. *The Hague: central planning bureau*.