ECON 721:

Generalized Method of Moments (aka Minimum Distance Estimation)

Fall, 2016

Main reference: Hansen (1982, *Econometrica*), Hansen & Singleton (1982, 1988)

Suppose that we have some data $\{x_t\}$, $t = 1..T$ and we want to test some hypotheses about $E(x_t) = \mu$. For example,

$$H_0 : \mu = \mu_0$$

How do we proceed? By a central limit theorem,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu) \tilde{} N(0, V_0)$$

$$\begin{aligned} V_0 &= E((x_t - \mu)(x_t - \mu)') \text{ if } x_t \text{ iid} \\ &= lim_{T \to \infty} E \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu)' \right) \text{ in general case.} \end{aligned}$$

Because V_0 is symmetric, we can decompose:

$$V_0 = QDQ'$$

where D is a matrix of eigenvalues and Q is an orthogonal matrix of eigenvectors.

$QQ' = I$

$Q^{-1} = Q'$.

We can write:

$$QD^{1/2}D^{1/2}Q' = V_0.$$

Under null, H_0,

$$\begin{aligned} &\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \mu_0)\right]' V_0^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \mu_0)\right] \\ = &\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \mu_0)\right]' QD^{-1/2}D^{-1/2}Q'\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \mu_0)'\right] \tilde{\ }\chi^2(n), \end{aligned}$$

where n is the number of moment conditions (length of vector of moments).

Note that the previous equation is a quadratic form in

$$D^{-1/2}Q'\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \mu_0)\right] \tilde{\ } N(0, I_n)$$

How does the test statistic behave under the alternative ($\mu \neq \mu_0$)?

$$
\begin{aligned}
&\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\mu_0)\right]' V_0^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\mu_0)\right] \\
&= \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\mu)\right]' V_0^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\mu)\right] + \\
&\qquad \frac{2}{\sqrt{T}}\sum_{t=1}^{T}(\mu-\mu_0)'V_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\mu) + \\
&\qquad \cdots + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mu-\mu_0)'V_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mu-\mu_0)
\end{aligned}
$$

The last term is $O(T)$ and gets large under the alternative (as we want).

Problems:

(i) V_0 is not known a priori.

Estimate $V_T \to V_0$

In an iid setting, use sample VCV matrix. More generally, approximate limit by finite T.

(ii) μ not known.

Suppose we want to test

$$H_0 : \mu = \varphi(\beta)$$

where φ is specified, but β is unknown.

We can estimate β by minimum χ^2 estimation:

$$min_{\beta\in B}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\varphi(\beta))\right]' V_T^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\varphi(\beta))\right] \tilde{} \chi^2(n-k),$$

where k = dimension of β and n = number of moments.

We will show that searching over k dimensions, you lose k degrees of freedom.

Let's find distribution theory for $\hat{\beta}$. First order condition (FOC) is:

$$\sqrt{T}\frac{\partial\varphi}{\partial\beta}\bigg|_{\hat{\beta}_T}' V_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\varphi(\hat{\beta}_T))=0$$

Taylor expand $\varphi(\hat{\beta}_T)$ around $\varphi(\hat{\beta}_0)$:

$$\varphi(\hat{\beta}_T)=\varphi(\beta_0)+\frac{\partial\varphi}{\partial\beta}\bigg|_{\hat{\beta}^*}(\hat{\beta}_T-\beta_0)\text{ , for }\beta^*\text{ between }\beta_0\text{ and }\hat{\beta}_T$$

Plug into FOC (for $\varphi(\hat{\beta}_T)$) to get

$$\sqrt{T}\frac{\partial\varphi}{\partial\beta}\bigg|_{\hat{\beta}_T}' V_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[x_t-\varphi(\hat{\beta}_0)-\frac{\partial\varphi}{\partial\beta}\bigg|_{\hat{\beta}^*}(\hat{\beta}_T-\beta_0)\right]=0$$

Rearrange to solve for $(\hat{\beta}_T-\beta_0)$:

$$\left[\sqrt{T}\frac{\partial\varphi}{\partial\beta}\bigg|_{\hat{\beta}_T}' V_T^{-1}(\frac{T}{\sqrt{T}})\frac{\partial\varphi}{\partial\beta}\bigg|_{\beta^*}\right](\hat{\beta}_T-\beta_0)=\sqrt{T}\frac{\partial\varphi}{\partial\beta}\bigg|_{\hat{\beta}_T}' V_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\varphi(\beta_0))$$

Apply CLT to last term. If

$$\frac{\partial\varphi}{\partial\beta}\bigg|_{\hat{\beta}_T}\rightarrow_p D_0$$

$$V_T\rightarrow_p V_0$$

$$\frac{\partial\varphi}{\partial\beta}\bigg|_{\beta^*}\rightarrow_p D_0$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\varphi(\beta_0))\tilde{\ }N(0,V_0)$$

Then,

$$
\begin{aligned}
\sqrt{T}(\hat{\beta}_T - \beta_0) &\sim N(0, (D_0' V_0^{-1} D_0)^{-1} D_0' V_0^{-1} V_0 V_0^{-1'} D_0 (D_0' V_0^{-1} D_0)^{-1'}) \\
&\sim N(0, (D_0' V_0^{-1} D_0)^{-1})
\end{aligned}
$$

where D_0' is $k \times n$ and V_0 is $n \times n$. Why is the limiting distribution of the criterion $\chi^2(n-k)$?

Write

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \varphi(\hat{\beta}_T)) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu_0) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\mu_0 - \varphi(\hat{\beta}_T))
$$

Recall that

$$
\begin{aligned}
\varphi(\hat{\beta}_T) &= \varphi(\beta_0) + \left.\frac{\partial \varphi}{\partial \beta}\right|_{\hat{\beta}*} (\hat{\beta}_T - \beta_0) \\
&= \mu_0 + \left.\frac{\partial \varphi}{\partial \beta}\right|_{\beta^*} (\hat{\beta}_T - \beta_0)
\end{aligned}
$$

which implies

$$
\mu_0 - \varphi(\hat{\beta}_T) = -\left.\frac{\partial \varphi}{\partial \beta}\right|_{\beta^*} (\hat{\beta}_T - \beta_0)
$$

Plug in the expression for $\hat{\beta}_T - \beta_0$ that we derived earlier to get:

$$
\mu_0 - \varphi(\hat{\beta}_T) = -\left.\frac{\partial \varphi}{\partial \beta}\right|_{\beta^*} \left\{ \left.\frac{\partial \varphi}{\partial \beta}\right|_{\hat{\beta}_T}' V_T^{-1} \left.\frac{\partial \varphi}{\partial \beta}\right|_{\beta^*} \right\}^{-1} \left.\frac{\partial \varphi}{\partial \beta}\right|_{\hat{\beta}_T}' V_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu_0)
$$

Note that this term is a linear combination of $\sum_{t=1}^{T}(x_t - \mu_0)$, so

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \varphi(\hat{\beta}_T)) = B \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu_0)
$$

where

$$plimB = I - \left.\frac{\partial\varphi}{\partial\beta}\right|_{\beta_0}\left[\left.\frac{\partial\varphi}{\partial\beta}\right|_{\beta_0}^{\prime}V_0^{-1}\left.\frac{\partial\varphi}{\partial\beta}\right|_{\beta_0}\right]^{-1}\left.\frac{\partial\varphi}{\partial\beta}\right|_{\beta_0}^{\prime}V_0^{-1} = B_0$$

Note that $\left.\frac{\partial\varphi}{\partial\beta}\right|_{\beta_0}^{\prime}V_0^{-1}B_0 = 0$. This tells us that certain linear combinations of B_0 will give a degenerate distribution (along k dimensions), which needs to be taken into account when testing.

Recall that

$$V_0^{-1} = QD^{-1/2}D^{-1/2}Q'$$

so

$$D^{-1/2}Q'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t - \varphi(\hat{\beta}_T) = D^{-1/2}Q'B\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \mu_0)$$

where

$$\begin{aligned}
D^{-1/2}Q'B &= D^{-1/2}Q' - D^{-1/2}Q'\left.\frac{\partial\varphi}{\partial\beta}\right|_{\hat{\beta}_T}\left[\left.\frac{\partial\varphi}{\partial\beta}\right|_{\hat{\beta}_T}^{\prime}QD^{-1/2}D^{-1/2}Q'\left.\frac{\partial\varphi}{\partial\beta}\right|_{\hat{\beta}_T}\right]^{-1}\left.\frac{\partial\varphi}{\partial\beta}\right|_{\beta_0}^{\prime}QD^{-1/2}D^{-1/2}Q' \\
&= (I - A(A'A)^{-1}A')D^{-1/2}Q' = M_A D^{-1/2}Q'
\end{aligned}$$

M_A is an idempotent matrix.

$$A = D^{-1/2}Q'\left.\frac{\partial\varphi}{\partial\beta}\right|_{\hat{\beta}_T}$$

Thus,

$$D^{-1/2}Q'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \varphi(\hat{\beta}_T)) = M_A D^{-1/2}Q'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \mu_0)$$

The matrix A accounts for the fact that we performed the minimization over β.

How is the distribution theory affected? We have a quadratic form in normal random variables with an idempotent matrix (M_A). For example

$$\hat{\varepsilon}'\hat{\varepsilon} = \varepsilon' M_X \varepsilon$$

$$M_X = I - X(X'X)^{-1}X'.$$

Some Facts (see, e.g., Seber)

(i) Let $Y \tilde{} N(\theta, \sigma^2 I_n)$

Let P be a symmetric matrix of rank r.

Then

$$Q = \frac{(Y-\theta)'P(Y-\theta)}{\sigma^2} \tilde{} \chi^2(r) \text{ iff } P = P^2$$

i.e. if P idempotent.

(ii) If $Q_i \tilde{} \chi^2(r_i)$, $i = 1, 2, r_1 > r_2$, and $Q = Q_1 - Q_2$ is independent of Q_2, then $Q \tilde{} \chi^2(r_1 - r_2)$.

Apply these results to get the distribution of

$$\frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2} = \frac{\varepsilon' M_X \varepsilon}{\sigma^2}$$

The rank of an idempotent matrix is equal to its trace and

$$tr(M_X) = \sum_{t=1}^{n} \lambda_i$$

for an idempotent matrix, λ_i are all 0 or 1.

$$
\begin{aligned}
rank(I_{n\times n} - X(X'X)^{-1}X') &= rank(I) - rank(X(X'X)^{-1}X') \\
&= n - trace(X(X'X)^{-1}X') \\
&= n - trace((X'X)^{-1}X'X) \\
&= n - k
\end{aligned}
$$

using the fact that $tr(AB) = tr(BA)$ (trace is invariant to cyclical perturbations).

Thus,

$$\frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2} \sim \chi^2(n-k)$$

By the same reasoning, under the null

$$\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \varphi(\hat{\beta}_T))\right]' V_T^{-1} \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \varphi(\hat{\beta}_T))\right] \sim \chi^2(n-k)$$

We preserve the χ^2 distribution but lose degrees of freedom in estimating β.

In the case where $n = k$ (just identified case), we can estimate β but we have no degrees of freedom left to perform the test.

Would GMM provide a method for estimating β if we use a weighting matrix that is different from V_T^{-1}?

Suppose we replace V_0^{-1} by W_0

$$min_{\beta\in B}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \varphi(\hat{\beta}_T))\right]' W_0 \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \varphi(\hat{\beta}_T))\right]$$

- Could choose, for example, $W_0 = I$, which avoids the need to estimate the weighting matrix.

- The result will be that the asymptotic covariance is altered, but $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t - \varphi(\hat{\beta}_T))$ will still be normally distributed. The asymptotic criterion distribution will be different.

What is the advantage of focusing on min-χ^2 estimation?

It turns out that $W_0 = V_0^{-1}$ gives the smallest covariance matrix. You get a more efficient estimator for β and the most powerful test of restrictions.

We can show this by showing that the following is positive semi-definite:

$$(D_0' W_0 D_0)^{-1}(D_0' W_0 V_0 W_0 D_0)(D_0' W_0 D_0)^{-1} - (D_0' V_0^{-1} D_0)^{-1}$$

where the first term is the covariance matrix for $\sqrt{T}(\hat{\beta}_T - \beta_0)$ when a general weighting matrix is used.

Equivalently, we can show that

$$(D_0' V_0^{-1} D_0) - (D_0' W_0 D_0)(D_0' W_0 V_0 W_0 D_0)^{-1}(D_0' W_0 D_0)$$

is positive semidefinite.

$$\begin{aligned}
&\alpha' \left[(D_0' V_0^{-1} D_0) - (D_0' W_0 D_0)((D_0' W_0 V_0 W_0 D_0)^{-1}(D_0' W_0 D_0)\right] \alpha \\
&= \alpha' D_0' V_0^{-1/2} \left[I - V^{1/2} W_0' D_0 (D_0' W_0 V_0^{1/2} V_0^{1/2} W_0 D_0)^{-1} D_0' W_0 V^{1/2}\right] V_0^{-1/2} D_0 \alpha \\
&= \alpha' D_0' V_0^{-1/2} \left[I - \tilde{V}(\tilde{V}'\tilde{V})\tilde{V}'\right] V^{1/2} D_0 \alpha \geq 0
\end{aligned}$$

The last term equals 0 if $W_0 = V_0^{-1}$.

The term in brackets is idempotent, so all eigenvalues are either 0 or 1.

We showed that it can be written in quadratic form. The quadratic form $x'Ax$ is positive (negative) semidefinite iff all eigenvalues of A are nonnegative (nonpositive) and at least one is zero.

Therefore, $W_0 = V_0^{-1}$ is the optimal choice for the weighting matrix.

Many standard estimators can be interpreted as GMM estimators

Some examples:

(i) *OLS*

$$y_t = x_t\beta + u_t$$

$$E(u_t x_t) = 0 \Rightarrow E((y_t - x_t\beta)x_t) = 0$$

$$min_{\beta \in B} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (y_t - x_t\beta)x_t \right)' V_T^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (y_t - x_t\beta)x_t \right)$$

where V_T is an estimator for $E(x_t u_t u_t' x_t')$ for the iid case, $= \sigma^2 E(x_t x_t')$ if homoskedastic.

(ii) *Instrumental variables*

$$y_t = x_t\beta_t + u_t$$

$$E(u_t x_t) \neq 0$$

$$E(u_t z_t) = 0$$

$$E(x_t z_t) \neq 0$$

where there are k_1 instruments.

$$min_{\beta \in B} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (y_t - x_t\beta)z_t \right)' V_T^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (y_t - x_t\beta)z_t \right)$$

where $V_T = \hat{E}(z_t u_t u_t' z_t')$ in iid case.

Suppose $E(u_t u_t' | z_t) = \sigma^2 I$

Then $V_0 = \sigma^2 E(z_t z_t')^{-1}$, $V_0^{-1} = (\sigma^2)^{-1} E(z_t z_t')$.

It is possible to verify that 2SLS and GMM give the same estimator:

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

In first stage, regress X on Z, get $\hat{X} = Z(Z'Z)^{-1}Z'X$ and regress Y on Z, get $\hat{Y} = Z(Z'Z)^{-1}Z'Y$. In the second stage, regress $\hat{Y}$ on $\hat{X}$

$$\begin{aligned} Var(\hat{\beta}_{2SLS}) &= \left[E(x_i z_i')E(z_i z_i')^{-1}E(x_i z_i')'\right]^{-1} \times E(x_i z_i)E(z_i z_i')^{-1}E(z_i u_i u_i z_i')E(z_i z_i')^{-1}E(x_i z_i')' \\ &\quad \times \left[E(x_i z_i')E(z_i z_i')^{-1}E(x_i z_i')'\right]^{-1} \\ &= \left[E(x_i z_i')E(z_i z_i')^{-1}E(x_i z_i')'\right]^{-1} \sigma^2 \end{aligned}$$

Under GMM,

$$Var(\hat{\beta}_{GMM}) = (D_0'V_0D_0)^{-1}$$
$$D_0 = \frac{\partial \varphi}{\partial \beta}\bigg|_{\beta_0} = plim \frac{1}{n}\sum_{i=1}^{n} x_i z_i = E(x_i z_i')$$
$$V_0 = \sigma^2 E(z_i z_i')$$

$$\begin{aligned} (D_0'V_0^{-1}D_0')^{-1} &= (E(x_i z_i')'(\sigma^2 E(z_i z_i'))^{-1}E(x_i z_i'))^{-1} \\ &= \sigma^2(E(x_i z_i')E(z_i z_i')^{-1}E(x_i z_i')')^{-1} \end{aligned}$$

(iii) *Nonlinear least squares*

$$y_t = \varphi(x_t;\beta) + u_t$$
$$E(u_t\varphi(x_t;\beta)) = 0$$

$$min_{\beta\in B}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t - \varphi(x_t;\beta))\varphi(x_t;\beta))'\right) V_T^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t - \varphi(x_t;\beta))\varphi(x_t;\beta)\right)$$

(iv) *General method of moments*

Any moment condition (usually derived from an economic model)

$E(f_t(\beta)) = 0$

$$min_{\beta \in B} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\beta) \right)' V_T^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\beta) \right)$$

where $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\beta_0) \tilde{} N(0, V_0)$.