# L8. Generalized Moment Method (GMM)

Yonghui Zhang

School of Economics, RUC

December 28, 2020

# Moment method

- Method of Moments (MM): the oldest method of finding point estimator, Karl Pearson, late 1890s
- Two fundamental features:
  - (i) it is based upon the empirical distribution that approximates the true distribution when the sample size is large enough (large sample theory)
  - (ii) it does not need to specify any sort of distribution and does not use any information about the population distribution other than its moments. (nonparametric, robust, but not efficient when the distribution is known)

# GMM

- Lars Hansen (1980): thousands of papers applying the GMM techniques have been published in the economics and finance area
- MLE, QMLE: efficiency, computation burden
- We begin by outlining the classical method of moments technique and then proceed to GMM

# Method of Moments Estimators

**Example**

Suppose $X_1, \ldots, X_n$ are i.i.d. from a gamma distribution with pdf

$$f(x, \theta_1, \theta_2) = \frac{1}{\Gamma(\theta_1)\,\theta_2^{\theta_1}} x^{\theta_1 - 1} \exp\left(-\frac{x}{\theta_2}\right)$$

where $\theta_i > 0$ for $i = 1, 2$. The likelihood function is

$$L(x, \theta_1, \theta_2) = \frac{1}{\left(\Gamma(\theta_1)\,\theta_2^{\theta_1}\right)^n} \left(\prod_{i=1}^n x_i^{\theta_1 - 1}\right) \exp\left(-\frac{1}{\theta_2} \sum_{i=1}^n x_i\right)$$

It is difficult to maximize $\ln L(x, \theta_1, \theta_2)$ w.r.t. $(\theta_1, \theta_2)$ due to the presence of the gamma function $\Gamma(\theta_1)$. Thus numerical methods must be used to maximize $\ln L(x, \theta_1, \theta_2)$.

# Method of Moments Estimators

\

- However, since $X \sim$Gamma$(\theta_1, \theta_2)$, $E(X) = \theta_1\theta_2$,and $Var(X) = \theta_1\theta_2^2$.
- Since the ecdf $F_n(x)$ converges in probability to $F(x)$, and hence their corresponding moments (if exist) should be about equal asymptotically.
- Define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

- We have

$$\theta_1\theta_2 = \bar{X} \text{ and } \theta_1\theta_2^2 = \hat{\sigma}^2$$

and obtain

$$\tilde{\theta}_1 = \frac{\bar{X}^2}{\hat{\sigma}^2} \text{ and } \tilde{\theta}_2 = \frac{\hat{\sigma}^2}{\bar{X}}.$$

# MM

- The idea of MM is to just match the population moments of a distribution to the sample moments.
- Let

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

be the $k$th sample moment and $\mu_k = E(X^k)$ the $k$th population moment.

- In general, $\mu_k = \mu_k(\theta)$. The method of moments estimator for $\theta$ is the solution to the equations

$$m_k = \mu_k(\theta) \text{ or } \frac{1}{n} \sum_{i=1}^{n} X_i^k = \mu_k(\theta)$$

for $k = 1, \dots, r$.

## Large sample theory for MM

- Consistency: $\hat{\theta}_{MM} \to_p \theta_0$ (Coming from the fact $m_k \to_p \mu_k(\theta_0)$ for $k = 1, ..., r$)
- Asymptotic distribution: Let $\mu(\theta) = (\mu_1(\theta), \ldots, \mu_r(\theta))'$.

$$\mu\left(\hat{\theta}_{MM}\right) - \mu(\theta_0) = \frac{\partial \mu(\theta_0)}{\partial \theta'}\left(\hat{\theta}_{MM} - \theta_0\right)$$

which implies that

$$\begin{aligned} \sqrt{n}\left(\hat{\theta}_{MM} - \theta_0\right) &= \left[\frac{\partial \mu(\theta_0)}{\partial \theta'}\right]^{-1} \sqrt{n}\left[\mu\left(\hat{\theta}_{MM}\right) - \mu(\theta_0)\right] \\ &= \left[\frac{\partial \mu(\theta_0)}{\partial \theta'}\right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (W_i - EW_i) \end{aligned}$$

where $W_i = \left(X_i, X_i^2, \ldots, X_i^k\right)'$.

# MM

- The most simple moment estimators: mean, variance

**Example (Poisson method of moments)**

Let $X_1, \ldots, X_n$ be i.i.d.. Poisson $(\lambda)$. Noting that $E_\lambda(X_1) = \lambda$, we can estimate $\lambda$ by

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

the sample analogue of $E_\lambda(X_1)$.

**Example (Normal moment of moments)**

Suppose $\mathbf{X} = (X_1, \ldots, X_n)'$ is a random sample from $N(\mu, \sigma^2)$. Then

$$\hat{\mu} = m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ and } \hat{\sigma}^2 = m_2 - m_1^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 .$$

# MM

**Example (OLS)**

The OLS estimator in the classical linear regression model $y_i = x_i'\beta + u_i$ is an MM estimator. Among the assumptions of the model is

$$E(x_i u_i) = 0,$$

whose sample analog is

$$\frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i = \frac{1}{n}\sum_{i=1}^{n} x_i \left(y_i - x_i'\hat{\beta}\right) = \frac{1}{n}X'\hat{u} = 0.$$

# MM

- **IV Estimator**: $E(z_i u_i) = 0$ and $\frac{1}{n}\sum_{i=1}^{n} z_i \hat{u}_i = 0$.
- **GLS Estimator**: $E\left(X'\Omega^{-1}u\right) = 0$ or $E\left(x_i^* u_i^*\right) = 0$. Then

$$\frac{1}{n}\sum_{i=1}^{n} x_i^* \hat{u}_i^* = \frac{1}{n}X^{*\prime}\hat{u}^* = \frac{1}{n}X'\Omega^{-1}\hat{u} = 0$$

- **MLE:** Let $l(\theta; X_i) = \log f(X_i, \theta)$, and $s(\theta, X_i) = \frac{\partial l(\theta, X_i)}{\partial \theta}$. Then $E\left[s(\theta, X_i)\right] = 0$ provided that

$$\begin{aligned} E\left[s(\theta, X_i)\right] &= \int \frac{\partial l(\theta, X_i)}{\partial \theta} f\left(\theta, X_i\right) dX_i \\ &= \int \frac{\partial f\left(\theta, X_i\right)}{\partial \theta} dX_i = \frac{\partial}{\partial \theta}\int f(\theta, X_i) dX_i = 0 \end{aligned}$$

Then the MLE is

$$\frac{1}{n}\sum_{i=1}^{n} s(\hat{\theta}_{MLE}, X_i) = 0.$$

# Generalized Method of Moments (GMM)

## Definition (GMM Estimator)

Consider the population moment conditions

$$Eg\left(Z_i,\theta_0\right)=0$$

where $Z_i$ is an $l\times 1$ random vector and $g\left(\cdot,\cdot\right):\mathbb{R}^l\times\mathbb{R}^k\to\mathbb{R}^m$ for $m\geq k$. Let $W_n$ be an $m\times m$ symmetric weight matrix and

$$\bar{g}\left(\theta\right)=\frac{1}{n}\sum_{i=1}^{n}g\left(Z_i,\theta\right).$$

The GMM estimator $\hat{\theta}$ of $\theta$ minimizes the criterion function

$$S_n\left(\theta\right)=\bar{g}\left(\theta\right)'W_n\bar{g}\left(\theta\right)$$

over $\theta\in\Theta\subseteq\mathbb{R}^k$.

# Generalized Method of Moments (GMM)

Example (Hall, 1978, JPE)

Suppose a representative consumer maximizes the expected utility function

$$E\left(\sum_{t=0}^{T}(1+\delta)^{-t}U\left(C_{i+t}\right)|\Omega_i\right)$$

subject to $A_i=\sum_{t=0}^{T}(1+r)^{-t}\left[W_{t+i}-C_{i+t}\right]$, where $U$ is a utility function, $\Omega_i$ is the information set at time $i$, $\delta$ is the rate of subjective time preference, $r$ is the rate of interest, $T$ is the length of economic life, $C_i$ is the consumption at time $i$, $W_i$ is the earnings at time $i$, and $A_i$ is the asset at time $i$.

## Generalized Method of Moments (GMM)

- The Euler equation

$$E\left[U'\left(C_{i+1}\right)|\Omega_i\right] = \frac{1+\delta}{1+r}U'\left(C_i\right)$$

- Let $U\left(C\right) = \frac{C^{1-\gamma}}{1-\gamma}$, we have

$$E\left[C_{i+1}^{-\gamma}|\Omega_i\right] = \frac{1+\delta}{1+r}C_i^{-\gamma}$$

or

$$E\left[1 - \frac{1+r}{1+\delta}\left(\frac{C_{i+1}}{C_i}\right)^{-\gamma}|\Omega_i\right] = 0.$$

- Now let $Q_i$ be an $m \times 1$ random vector included in $\Omega_i$, then, by the LIE, we have

$$E\left[Q_i\left\{1-\frac{1+r}{1+\delta}\left(\frac{C_{i+1}}{C_i}\right)^{-\gamma}\right\}\right]=0$$

which can be written as

$$g\left(Z_i,\theta\right)=Q_i\left\{1-\frac{1+r}{1+\delta}\left(\frac{C_{i+1}}{C_i}\right)^{-\gamma}\right\}$$

where $Z_i=\left(C_{i+1},C_i,Q_i\right)'$ and $\theta=\left(\delta,r,\gamma\right)'$.

# GMM

Example (GMM estimator of linear regression models)

Consider the following linear regression model

$$y_i = x_i'\beta + u_i,\ E\left(z_i u_i\right) = 0,\ i = 1, \ldots, n$$

where the dimensions of $x_i$ and $z_i$ are $k \times 1$ and $m \times 1$ with $m \geq k$, and the condition

$$E\left(z_i u_i\right) = 0$$

specifies $m$ moment conditions. If $m = k$, the regression model is just identified; if $m > k$, it is overidentified. We don't consider the underidentified case where $m < k$ and there are infinitely many solutions to the moment conditions. Note that the variables $z_i$ may include some components of $x_i$, but this is not required. This model falls in the definition of GMM estimator by setting

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

# GMM

- The above model includes the class of OLS and IV estimators as special cases.
- In fact, we can allow other functional form of the regression function, e.g.,

$$y_i = f(x_i, \beta) + u_i.$$

In this case, we have

$$g(y_i, x_i, z_i, \beta) = z_i \left[ y_i - f(x_i, \beta) \right].$$

## Consistency of GMM Estimators

- **Assumption G1.** $\Theta$ is compact.
- **Assumption G2.** $\bar{g}(\theta)$ converges uniformly over $\theta \in \Theta$ to a nonstochastic continuous function $g(\theta)$ in probability. That is

$$\sup_{\theta \in \Theta} \|\bar{g}(\theta) - g(\theta)\| \to_p 0.$$

- **Assumption G3.** There exists $\theta_0 \in \Theta$ such that $g(\theta_0) = 0$.
- **Assumption G4.** $\theta_0$ is the unique solution of $g(\theta_0) = 0$ over $\theta \in \Theta$.
- **Assumption G5.** $W_n \to_p W$, where $W$ is nonstochastic symmetric and nonsingular matrix.
- **Remarks:** The above assumptions are standard in the literature. They do not require the $Z_i$'s to be i.i.d.. If $Z_1, \ldots, Z_n$ are identically distributed, then

$$g(\theta) = Eg(Z_i, \theta).$$

# Consistency of GMM Estimators

**Theorem**

*Suppose that Assumptions (G1)-(G5) hold. Then GMM estimator*

$$\hat{\theta} \longrightarrow_p \theta_0.$$

- Noting that the GMM estimator is a special case of extreme estimator, we can apply the consistent result for extreme estimator to show the above theorem.

# Asymptotic Normality of GMM Estimators

- **Assumption G6.** $\theta_0$ is an interior point of $\Theta$.
- **Assumption G7**. $\{g(Z_i, \theta_0)\}$ obeys a CLT:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Z_i, \theta_0) \to_d N(0, V)$$

  where $V$ is finite and positive definite.
- **Assumption G8**. $g(Z, \theta)$ is continuously differentiable in $\theta$ for all $z$. $\{\partial g(Z_i, \theta)/\partial\theta\}$ obeys a WULLN:

$$\sup_{\theta\in\Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(Z_i, \theta)}{\partial\theta'} - G(\theta) \right\| \to_p 0,$$

  where $G(\theta) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left[\frac{\partial g(Z_i,\theta)}{\partial\theta'}\right]$ is a $m \times k$ matrix that is **continuous** in $\theta \in \Theta$ and is of full rank in the neighborhood of $\theta_0$.

# Asymptotic Normality of GMM Estimators

- G6 is required because we need to apply a Taylor series expansion.
- G7-G8 make assumptions on $g(\cdot, \theta)$ where we assume both CLT and ULLN hold directly. These are another case of "high-level assumptions". We can impose more primitive conditions on the data and $g(\cdot, \theta)$ to ensure these assumptions are met.
- When $Z_1, \ldots, Z_n$ are i.i.d., the formulae for $V$ and $G(\theta)$ can be simplified

$$V = Var\left[g(Z_i, \theta_0)\right] \text{ and } G(\theta) = E\left[\frac{\partial g(Z_i, \theta)}{\partial \theta'}\right].$$

# Asymptotic Normality of GMM Estimators

**Theorem (Asymptotic Normality)**

*Suppose Assumptions G1-G8 hold. Then*

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \rightarrow_d N\left(0,\Sigma\right),$$

*where*

$$\Sigma = \left(G_0' W G_0\right)^{-1} G_0' W V W G_0 \left(G_0' W G_0\right)^{-1} \text{ and } G_0 = G\left(\theta_0\right).$$

# Asymptotic Normality of GMM Estimators: Proofs

- By the consistency theorem, we have $\hat{\theta} \to_p \theta_0$ under Assumptions G1-G5. By Assumption G6, $\hat{\theta}$ is an interior element of $\Theta$ for sufficiently large $n$. Noting that

$$\hat{\theta} = \operatorname{argmin} \bar{g}(\theta)' W_n \bar{g}(\theta),$$

we have the FOC

$$\left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' W_n \bar{g}(\hat{\theta}) = 0.$$

- Expanding $\bar{g}(\hat{\theta})$ only around $\theta_0$ yields

$$\left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' W_n \bar{g}(\theta_0) + \left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' W_n \frac{\partial \bar{g}(\bar{\theta})}{\partial \theta'} \left(\hat{\theta} - \theta_0\right) = 0$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and $\theta_0$ and $\bar{\theta} \to_p \theta_0$.

# Asymptotic Normality of GMM Estimators: Proofs

- Hence

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right)=-\left\{\left[\frac{\partial\bar{g}(\hat{\theta})}{\partial\theta'}\right]' W_n \frac{\partial\bar{g}(\bar{\theta})}{\partial\theta'}\right\}^{-1}\left[\frac{\partial\bar{g}(\hat{\theta})}{\partial\theta'}\right]' W_n\sqrt{n}\bar{g}(\theta_0)$$

- By Assumption G8,

$$\begin{aligned}\left\|\frac{\partial\bar{g}(\hat{\theta})}{\partial\theta'}-G\left(\theta_0\right)\right\| &= \left\|\frac{\partial\bar{g}(\hat{\theta})}{\partial\theta'}-G\left(\hat{\theta}\right)\right\|+\left\|G\left(\hat{\theta}\right)-G\left(\theta_0\right)\right\| \\ &\leq \sup_{\theta\in\Theta}\left\|\frac{\partial\bar{g}(\theta)}{\partial\theta'}-G\left(\theta\right)\right\|+\left\|G\left(\hat{\theta}\right)-G\left(\theta_0\right)\right\| \\ &= o_p\left(1\right)+o_p\left(1\right),\end{aligned}$$

similarly, $\left\|\frac{\partial\bar{g}(\bar{\theta})}{\partial\theta'}-G\left(\theta_0\right)\right\|=o_p\left(1\right)$.

# Asymptotic Normality of GMM Estimators: Proofs

- By Assumptions G5 and G8 and the Slutsky theorem,

$$\left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' W_n \frac{\partial \bar{g}(\bar{\theta})}{\partial \theta'} \to_p G_0' W G_0$$

  which is positive definite.

- From Assumptions G5 and G7, and the previous results, we have

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to_d N\left(0, \left(G_0' W G_0\right)^{-1} G_0' W V W G_0 \left(G_0' W G_0\right)^{-1}\right).$$

# Optimal Weight Matrix

- What's is the **optimal** $W$ that minimizes the asymptotic covariance matrix $\Sigma$?
- The answer is $W = V^{-1}$ : weighting the moment equations according to **how precisely each of the equations is estimated**.
- The asymptotic covariance simplifies to

$$\Sigma_0 = \left(G_0' V^{-1} G_0\right)^{-1}.$$

- So it suffices to show that

$$\Sigma - \Sigma_0 \text{ is p.s.d.}$$

- Note that

$$\begin{aligned} \Sigma_0^{-1} - \Sigma^{-1} &= G_0' V^{-1} G_0 - G_0' W G_0 \left(G_0' W V W G_0\right)^{-1} G_0' W G_0 \\ &\equiv G_0' V^{-1/2} (I - P) V^{-1/2} G_0 \end{aligned}$$

where $P = V^{1/2} W G_0 \left(G_0' W V W G_0\right)^{-1} G_0' W V^{1/2}$ is a projection matrix.

# Efficient GMM Estimators

- $V$ is generally not available, yet it can be estimated consistently, say, by $\hat{V}_n^*$.
- Let $W_n = \hat{V}_n^{*-1}$ and obtain the efficient GMM estimator that has the smallest asymptotic variance-covariance matrix **given the moment conditions.**
- In practice, the efficient GMM estimator is obtained in the following two steps:
  1. Minimize $S_n(\theta)$ with $W_n = I$ to get the one-step estimator $\tilde{\theta}$ of $\theta_0$. ($\tilde{\theta}$ is generally inefficient, but it is still consistent for $\theta_0$.)
  2. Find a preliminary consistent estimator $\tilde{V}_n$ for $V = \mathtt{avar}\left[\sqrt{n}g(\theta_0)\right]$ and minimize $S_n(\theta)$ by choosing $W_n = \hat{V}_n^{-1}$. Denote the resulting estimator of $\theta_0$ as $\hat{\theta}_{EGMM}$.

- In Step 2.
  - Case (i): If $\{g(Z_i, \theta_0)\}$ forms a stationary m.d.s., then we can estimate $V$ consistently by

$$\tilde{V}_n = \frac{1}{n} \sum_{i=1}^{n} g(Z_i, \tilde{\theta}) g(Z_i, \tilde{\theta})'.$$

  - Case (ii): If $\{g(Z_i, \theta_0)\}$ is not an m.d.s., then $V = \sum_{j=-\infty}^{\infty} \Gamma(j)$ with $\Gamma(j) = E\left[g(Z_i, \theta_0) g(Z_{i-j}, \theta_0)'\right]$. A HAC estimator for $V$ is given by

$$\tilde{V}_n = \sum_{j=-n+1}^{n-1} k(j/p_n) \tilde{\Gamma}(j)$$

    with $\tilde{\Gamma}(j) = \frac{1}{n} \sum_{i=j+1}^{n} g(Z_i, \tilde{\theta}) g(Z_{i-j}, \tilde{\theta})'$ for $j \geq 0$. $k(\cdot)$ is a kernel function that satisfies certain regularity conditions and $p_n$ is the bandwidth parameter such that $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$.

# Efficient GMM Estimators

**Theorem (Asymptotic Efficiency)**

*Under Assumptions G1-G8, the two-step efficient GMM estimator has the following asymptotic distribution:*

$$\sqrt{n}\left(\hat{\theta}_{EGMM} - \theta_0\right) \to_d N\left(0, \Sigma_0\right)$$

*where*

$$\Sigma_0 = \left(G_0' V^{-1} G_0\right)^{-1}.$$

# Asymptotic Variance Estimator

- The estimation of $V$ depends crucially on the underlying data assumption.
  - m.d.s: $\hat{V}_n = n^{-1}\sum_{i=1}^{n} g(Z_i, \hat{\theta}_{EGMM})g(Z_i, \hat{\theta}_{EGMM})'$;
  - not m.d.s.: $\hat{V}_n = n^{-1}\sum_{j=-n+1}^{n-1} k\left(j/p_n\right)\hat{\Gamma}\left(j\right)$ with $\hat{\Gamma}\left(j\right) = n^{-1}\sum_{i=j+1}^{n} g(Z_i, \hat{\theta}_{EGMM})g(Z_{i-j}, \hat{\theta}_{EGMM})'$.

For both cases, we can show $\hat{V}_n$ is consistent. To cover both cases, we make the following assumption.

- **Assumption G9.** $\hat{V}_n \to_p V$
- Let

$$\hat{G}_n = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(Z_i, \hat{\theta}_{EGMM})}{\partial \theta'}.$$

**Lemma**

*Under Assumptions G1-G8,* $\hat{G}_n \to_p G_0$.

# Asymptotic Variance Estimator

- **Proof.** By the triangle inequality,

$$
\begin{aligned}
& \left\| \hat{G}_n - G_0 \right\| \\
\leq \quad & \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(Z_i, \hat{\theta}_{EGMM})}{\partial \theta'} - G\left(\hat{\theta}_{EGMM}\right) \right\| \\
& + \left\| G\left(\hat{\theta}_{EGMM}\right) - G\left(\theta_0\right) \right\| \\
\leq \quad & \sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(Z_i, \theta)}{\partial \theta'} - G\left(\theta\right) \right\| + \left\| G\left(\hat{\theta}_{EGMM}\right) - G\left(\theta_0\right) \right\| \\
\rightarrow \quad & {}_p 0.
\end{aligned}
$$

# Asymptotic Variance Estimator

**Theorem**

*Under Assumptions G1-G9, we have*

$$\hat{\Sigma}_0 = \left(\hat{G}_n' \hat{V}^{-1} \hat{G}_n\right)^{-1} \rightarrow_p \Sigma_0.$$

# Hypothesis testing

- We now consider testing the hypotheses

$$H_0 : R(\theta_0) = r \text{ vs } H_1 : R(\theta_0) \neq r$$

where $R : \mathbb{R}^k \to \mathbb{R}^q$ is a continuously differentiable function with $q \leq k$, and $r$ is a known $q \times 1$ vector.

- Assume that the first derivative $R'(\theta)$ of $R(\theta)$ is of full rank in the neighborhood of $\theta_0$. That is

$$\texttt{rank}\left[R'(\theta)\right] = q, \text{ where } R'(\theta) = \partial R(\theta)/\partial\theta'$$

- For example, if $R(\theta_0) = R\theta_0$ where $R$ is a $q \times k$ matrix of full rank $q$, we are testing linear restrictions.

# Hypothesis testing

- To test for the nonlinear restrictions specified by $H_0$, we make a Taylor series expansion of $R\left(\hat{\theta}_{EGMM}\right)$ at $\theta_0$ :

$$
\begin{aligned}
\sqrt{n}\left[R\left(\hat{\theta}_{EGMM}\right)-r\right] &= \sqrt{n}\left[R\left(\theta_0\right)-r\right]+\sqrt{n}\left[R\left(\hat{\theta}_{EGMM}\right)-R\left(\theta_0\right)\right] \\
&= \sqrt{n} R'(\bar{\theta})\left(\hat{\theta}_{EGMM}-\theta_0\right) \text{ under } H_0
\end{aligned}
$$

where $\bar{\theta} \rightarrow_p \theta_0$ under Assumptions G1-G5.

- By the continuity of $R'(\bar{\theta})$ and Slutsky theorem, we have $R'(\bar{\theta}) \rightarrow_p R'\left(\theta_0\right)$. Then

$$
\sqrt{n}\left[R\left(\hat{\theta}_{EGMM}\right)-r\right] \rightarrow_d N\left(0, R'\left(\theta_0\right) \Sigma_0\left[R'\left(\theta_0\right)\right]'\right)
$$

# Hypothesis testing

- It follows that

$$
\begin{aligned}
W_n &= \sqrt{n}\left[R\left(\hat{\theta}_{EGMM}\right)-r\right]'\left[R'\left(\hat{\theta}_{EGMM}\right)\hat{\Sigma}_0 R\left(\hat{\theta}_{EGMM}\right)\right]^{-1} \\
&\quad \times\sqrt{n}\left[R\left(\hat{\theta}_{EGMM}\right)-r\right] \\
&\rightarrow {}_d\chi^2(q)
\end{aligned}
$$

**Theorem**

*Under Assumptions G1-G9 and $H_0$,*

$$W_n \rightarrow_d \chi^2(q).$$

# An Example

- Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf $f(x) = \theta \exp(-\theta x) 1(x \geq 0)$, $\theta > 0$.
- Let $\theta_0$ be the true parameter value.
- Note that $E(X_1) = 1/\theta_0$. mm estimator: $\hat{\theta}_{mm} = \bar{X}^{-1}$, and $\sqrt{n}\left(\hat{\theta}_{mm} - \theta_0\right) \rightarrow^d N\left(0, \theta_0^2\right)$ (by delta method)
- Also, $\hat{\theta}_{ML} = \hat{\theta}_{mm}$.
- In addition, we have $E(X_1^2) = 2/\theta_0^2$.
- Then we have a GMM estimator for $\theta$ based upon the first two moment conditions

$$g_i(\theta) = \begin{pmatrix} g_1(X_i, \theta) \\ g_2(X_i, \theta) \end{pmatrix} = \begin{pmatrix} X_i^{-1} - \theta \\ X_i^{-2} - 2/\theta^2 \end{pmatrix}$$

# An Example

- The GMM estimator $\hat{\theta}$ of $\theta$ is defined as

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} S_n(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} g(X_i, \theta)\right]' W_n \left[\frac{1}{n}\sum_{i=1}^{n} g(X_i, \theta)\right]$$

where where $W_n$ is a $2 \times 2$ nonsingular weight matrix.

- By the FOC and the first order Taylor expansion

$$0 = \frac{\partial S_n\left(\hat{\theta}\right)}{\partial \theta'} = \frac{\partial S_n\left(\theta_0\right)}{\partial \theta'} + \frac{\partial^2 S_n\left(\bar{\theta}\right)}{\partial \theta' \partial \theta}\left(\hat{\theta} - \theta_0\right)$$

# An Example

- Then

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right)=-\left[\frac{\partial^2 S_n\left(\bar{\theta}\right)}{\partial\theta'\partial\theta}\right]^{-1}\sqrt{n}\frac{\partial S_n\left(\theta_0\right)}{\partial\theta'}$$

- First, $\sqrt{n}\frac{\partial S_n(\theta_0)}{\partial\theta'}=\left[\frac{1}{n}\sum_{i=1}^n\frac{\partial g_i(\theta_0)}{\partial\theta'}\right]' W_n\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i(\theta_0)$
  - $\frac{1}{n}\sum_{i=1}^n\frac{\partial g_i(\theta_0)}{\partial\theta'}=\left(\frac{1}{\theta_0^2},\frac{4}{\theta_0^3}\right)'=G_0$
  - $\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i(\theta_0)=\left(\begin{array}{c}\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(X_i^{-1}-\theta_0\right)\\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(X_i^{-2}-2/\theta_0^2\right)\end{array}\right)\to_d N\left(0,V\right),$

    where $V=E\left[g_i(\theta_0)g_i(\theta_0)'\right]=\left(\begin{array}{cc}\frac{1}{\theta_0^2}&\frac{4}{\theta_0^3}\\ \frac{4}{\theta_0^3}&\frac{20}{\theta_0^4}\end{array}\right).$

# An Example

- Second,

$$
\begin{aligned}
\frac{\partial^2 S_n\left(\bar{\theta}\right)}{\partial\theta'\partial\theta} &= \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_i\left(\bar{\theta}\right)}{\partial\theta'}\right]' W_n \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_i\left(\bar{\theta}\right)}{\partial\theta'}\right] \\
&\quad + \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 g_i\left(\bar{\theta}\right)}{\partial\theta'\partial\theta}\right]' W_n \left[\frac{1}{n}\sum_{i=1}^{n} g_i\left(\bar{\theta}\right)\right] \\
&\rightarrow {}_p G_0' W G_0 + E\left[\frac{\partial^2 g_i\left(\theta_0\right)}{\partial\theta'\partial\theta}\right]' W E\left[g_i\left(\theta_0\right)\right] \\
&= G_0' W G_0.
\end{aligned}
$$

# An Example

- It follows that $\sqrt{n}\left(\hat{\theta}-\theta_0\right) \to_d N\left(0, \left[G_0' W G_0\right]^{-1}\left(G_0' W G_0\right)^{-1} G_0' W V W G_0 \left[G_0' W G_0\right]^{-1}\right)$.
- Let $W = V^{-1}$. Then $\sqrt{n}\left(\hat{\theta}_{EGMM}-\theta_0\right) \to_d N\left(0, \left[G_0' V^{-1} G_0\right]^{-1}\right)$, where

$$\begin{aligned} G_0' V^{-1} G_0 &= \begin{pmatrix} \frac{1}{\theta_0^2} \\ \frac{4}{\theta_0^3} \end{pmatrix}' \begin{pmatrix} \frac{1}{\theta_0^2} & \frac{4}{\theta_0^3} \\ \frac{4}{\theta_0^3} & \frac{20}{\theta_0^4} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\theta_0^2} & \frac{4}{\theta_0^3} \\ \frac{4}{\theta_0^3} & \frac{20}{\theta_0^4} \end{pmatrix} \\ &= \frac{1}{\theta_0^2} \end{aligned}$$

- It follows that $\sqrt{n}\left(\hat{\theta}_{EGMM}-\theta_0\right) \to_d N\left(0, \theta_0^2\right)$.
- Note that $\hat{\theta}_{EGMM}$ has the same asymptotic distribution as $\hat{\theta}_{ML}$, the MLE of $\theta$.
- So the efficient GMM estimator is as efficient as the ML estimator in this case.

# Testing for Over-identifying Restrictions

- The null and alternative hypotheses of interest are

$$
\begin{aligned}
H_0 &: E\left[g\left(Z_i, \theta_0\right)\right] = 0 \text{ for some } \theta_0 \in \Theta \\
H_1 &: E\left[g\left(Z_i, \theta\right)\right] \neq 0 \text{ for all } \theta \in \Theta.
\end{aligned}
$$

- Suppose that $m > k$, i.e., the number of moment conditions exceeds the number of parameters to be estimated.
- Note that, in this case, we have $m - k$ over-identifying restrictions because only $k$ moment conditions are needed to identify $k$ parameters.
- We use the over-identifying restrictions to test the null hypothesis $H_0$.

# Testing for Over-identifying Restrictions

- The idea to construct the test statistic is simple. Let $\hat{\theta}$ be a GMM estimator. Under $H_0$, we expect that

$$\bar{g}\left(\hat{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta})$$

  is close to zero. (It will not be identically equal to zero unless $m = k$.)
- So we can take $\bar{g}\left(\hat{\theta}\right)$ as the basis of our test statistic and define

$$J_n = nS_n\left(\hat{\theta}_{EGMM}\right) = n\bar{g}\left(\hat{\theta}_{EGMM}\right)' \hat{V}_n^{-1}\bar{g}\left(\hat{\theta}_{EGMM}\right).$$

# Testing for Over-identifying Restrictions

To prove the next theorem for $J_n$, we need the following lemma.

**Lemma**

*Let $\xi_n \to_d N(0, I_m)$, where $I_m$ is an $m \times m$ identity matrix. Let $\Lambda$ be an $m \times m$ idempotent matrix with rank $d$. Then the quadratic form*

$$\xi_n' \Lambda \xi_n \to_d \chi^2 (d) .$$

**Theorem (Hansen's J-test for overidentifying restrictions)**

*Suppose Assumptions G1-G9 hold. Under $H_0 : E[g(Z_i, \theta_0)] = 0$ for some $\theta_0 \in \Theta$*

$$J_n \to_d \chi^2 (m - k) .$$

- **Proof.** For notational simplicity, let $\hat{\theta} = \hat{\theta}_{EGMM}$. Noting that $\hat{\theta} = \text{argmin}_{\theta}\, \bar{g}(\theta)' \tilde{V}_n^{-1} \bar{g}(\theta)$, we have the FOC:

$$\left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' \tilde{V}_n^{-1} \bar{g}(\hat{\theta}) = 0.$$

- Expanding $\bar{g}(\hat{\theta})$ around $\theta_0$ gives us

$$\left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' \tilde{V}_n^{-1} \bar{g}(\theta_0) + \left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' \tilde{V}_n^{-1} \frac{\partial \bar{g}(\bar{\theta})}{\partial \theta'} \left(\hat{\theta} - \theta_0\right) = 0.$$

- It follows that

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -\left\{\left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' \tilde{V}_n^{-1} \frac{\partial \bar{g}(\bar{\theta})}{\partial \theta'}\right\}^{-1} \left[\frac{\partial \bar{g}(\hat{\theta})}{\partial \theta'}\right]' \tilde{V}_n^{-1} \sqrt{n}\bar{g}(\theta_0$$

# Testing for Over-identifying Restrictions

- Consider the first order Taylor series expansion of $\sqrt{n}\bar{g}\left(\hat{\theta}\right)$ around $\theta_0$:

$$\sqrt{n}\bar{g}\left(\hat{\theta}\right) = \sqrt{n}\bar{g}\left(\theta_0\right) + \frac{\partial \bar{g}\left(\theta^*\right)}{\partial \theta'}\sqrt{n}\left(\hat{\theta} - \theta_0\right)$$

- Under Assumptions G1-G5, $\hat{\theta} \longrightarrow_p \theta_0$ implies that $\bar{\theta} \longrightarrow_p \theta_0$ and $\theta^* \longrightarrow_p \theta_0$.

# Testing for Over-identifying Restrictions

It follows that

$$
\begin{aligned}
\hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\hat{\theta}\right) &= \hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\theta_0\right) + \hat{V}_n^{-1/2}\frac{\partial\bar{g}\left(\theta^*\right)}{\partial\theta'}\sqrt{n}\left(\hat{\theta}-\theta_0\right) \\
&= \hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\theta_0\right) \\
&\quad -\hat{V}_n^{-1/2}\frac{\partial\bar{g}\left(\theta^*\right)}{\partial\theta'}\left\{\left[\frac{\partial\bar{g}\left(\hat{\theta}\right)}{\partial\theta'}\right]'\tilde{V}_n^{-1}\frac{\partial\bar{g}\left(\bar{\theta}\right)}{\partial\theta'}\right\}^{-1} \\
&\quad \times\left[\frac{\partial\bar{g}\left(\hat{\theta}\right)}{\partial\theta'}\right]'\tilde{V}_n^{-1}\sqrt{n}\bar{g}\left(\theta_0\right) \\
&\equiv \Psi_n\hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\theta_0\right)
\end{aligned}
$$

# Testing for Over-identifying Restrictions

- By Assumptions G1-G9 and the Slutsky theorem, $\hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\theta_0\right) \to_d N\left(0, I_m\right)$ *and*

$$\begin{aligned}\Psi_n &= I_m - \hat{V}_n^{-1/2}\frac{\partial\bar{g}\left(\theta^*\right)}{\partial\theta'}\left\{\left[\frac{\partial\bar{g}\left(\hat{\theta}\right)}{\partial\theta'}\right]'\tilde{V}_n^{-1}\frac{\partial\bar{g}\left(\bar{\theta}\right)}{\partial\theta'}\right\}^{-1} \\ &\times\left[\frac{\partial\bar{g}\left(\hat{\theta}\right)}{\partial\theta'}\right]'\tilde{V}_n^{-1}\hat{V}_n^{1/2} \\ &\to {}_p I_m - V^{-1/2}G_0\left(G_0'V^{-1}G_0\right)^{-1}G_0V^{-1/2} \equiv \Psi\end{aligned}$$

where $\Psi$ is $m \times m$ idempotent matrix with rank $m-k$.

- Note that $\hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\hat{\theta}\right) = \Psi_n\hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\theta_0\right) = \Psi\hat{V}_n^{-1/2}\sqrt{n}\bar{g}\left(\theta_0\right) + o_p\left(1\right).$
- By the above Lemma, we have $J_n \to_d \chi^2\left(k-m\right).$

# Testing for Over-identifying Restrictions

- The test is often called the **J-test** for test for overidentification in the GMM literature. If the the $J_n$ statistic is sufficiently large, it means that either the null restrictions or the other assumptions (or both) are likely to be false. If we are confident about the other assumptions, we can interpret the large $J_n$ statistic as evidence against $H_0$.
- The J-test is an **asymptotic** test. It may not perform well in small samples. In fact, the small sample properties of the test have become a concern since the middle 1990s. J-test rejects the null hypothesis too often in practice.
- In practice, we can consider testing a subset of the restrictions, e.g.,

$$H_0 : E\left[g_1\left(Z_i, \theta_0\right)\right] = 0 \text{ for some } \theta_0 \in \Theta$$

where $g_1\left(Z_i, \theta_0\right)$ is a subset of the $m \times 1$ vector. See Hayashi (2000) for details.

# GMM for linear regression models

- A1. Let $\{w_i \equiv (y_i, x_i, z_i)\}$ be a stationary ergodic process with
$$y_i = x_i'\beta_0 + u_i, i = 1, \ldots, n,$$
where $x_i$ is a $k \times 1$ vector of regressors, $\beta_0$ is an $k \times 1$ vector of parameters, $u_i$ is an unobservable error term, and $z_i$ is an $m \times 1$ vector of instruments with $m \geq k$.
- A2. $z_i$ is predetermined in the sense that it is orthogonal to the current error term. That is,
$$E\left[g_i(\beta_0)\right] = 0,$$
where $g_i(\beta) \equiv g(w_i, \beta) = z_i\left(y_i - x_i'\beta\right)$.
- A3. The $m \times k$ matrix $Q_{zx} = E\left[z_i x_i'\right]$ is finite and of full rank.
- A4. $\{g_i(\beta_0)\}$ follows a version of CLT:
$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i(\beta_0) \rightarrow_d N(0, V)$$
where $V$ is finite and positive definite.

# GMM for linear regression models

- **Rank Condition for Identification: A3**

$$Q_{zx}\beta_0 = E\left(z_i y_i\right)$$

- **Order Condition for Identification:** a necessary condition for identification

$$m(\text{number of predetermined variables}) \geq k(\text{number of regressors}).$$

# GMM for linear regression models

- In the case $m = k$, we can estimate the unknown parameter $\beta$ by the method of moments.
- The sample analogue of the population moment conditions $E[g(w_i, \beta_0)] = 0$ is

$$\bar{g}\left(\hat{\beta}_{mm}\right) = \frac{1}{n}\sum_{i=1}^{n} g\left(w_i, \hat{\beta}_{mm}\right) = \frac{1}{n}\sum_{i=1}^{n} z_i\left(y_i - x_i'\hat{\beta}_{mm}\right) = 0$$

- It follows that

$$\hat{Q}_{zx}\hat{\beta}_{mm} = \hat{Q}_{zy}$$

and

$$\hat{\beta}_{mm} = \hat{Q}_{zx}^{-1}\hat{Q}_{zy} = \left(Z'X\right)^{-1} Z'Y$$

# GMM for linear regression models

- In the case of overidentification $(m > k)$, we need to apply the general principle of GMM estimation.
- Let
$$J_n(\beta) = n\bar{g}(\beta)' W_n \bar{g}(\beta)$$
- The GMM estimator $\hat{\beta}_{GMM}$ is a minimizer of $J_n(\beta)$.
- The FOC for the GMM estimation is given by
$$\begin{aligned} \frac{\partial J_n\left(\hat{\beta}_{GMM}\right)}{\partial \beta'} &= 2n\left[\frac{\partial \bar{g}\left(\hat{\beta}_{GMM}\right)}{\partial \beta'}\right]' W_n \bar{g}\left(\hat{\beta}_{GMM}\right) \\ &= -2nX'ZW_nZ'\left(Y - X\hat{\beta}_{GMM}\right) = 0 \end{aligned}$$
which gives that $\hat{\beta}_{GMM} = \left(X'ZW_nZ'X\right)^{-1} X'ZW_nZ'Y$.
- If $m = k$, then
$\hat{\beta}_{GMM} = \left(X'ZW_nZ'X\right)^{-1} X'ZW_nZ'Y = \left(Z'X\right)^{-1} Z'Y = \hat{\beta}_{MM}$.

# GMM for linear regression models

- A5. $W_n \to_p W$, where $W$ is a symmetric, positive definite, and nonstochastic matrix.

Theorem (Asymptotic properties of the GMM estimator)

*Suppose that Assumptions A1-A3 and A5 hold. Then*

$$\hat{\beta}_{GMM} \to_p \beta_0$$

*and in addition, Assumption A4 also holds, then*

$$\sqrt{n}\left(\hat{\beta}_{GMM} - \beta_0\right) \to_d N\left(0, \Sigma\right)$$

*where*

$$\Sigma = \left(Q_{xz} W Q_{zx}\right)^{-1} Q_{xz} W V W Q_{zx} \left(Q_{xz} W Q_{zx}\right)^{-1}.$$

# GMM for linear regression models

- The optimal weight matrix $W_0$ is the one that minimizes

$$\left(Q_{xz}WQ_{zx}\right)^{-1}Q_{xz}WVWQ_{zx}\left(Q_{xz}WQ_{zx}\right)^{-1}$$

- This turns out to be $W_0 = V^{-1}$.
- Other topics
  - A two step procedure is proposed to obtain the feasible efficient GMM estimator.
  - hypothesis test
  - Overidentification test

Thanks!