

Summer School in Structural Estimation GMM, Influence Functions, and Weight Matrices

University of Michigan
Ross School of Business

Toni M. Whited

August 2021

Why am I bothering to go over something as basic as GMM?

- ▶ Recall the definition of an SMM estimator
- ▶ The simulated moments estimator of θ is defined as the solution to the minimization of

$$\hat{\theta} = \arg \min_{\theta} Q_n(\theta) \equiv g(w_i, \theta)' \hat{\Xi}_n g(w_i, \theta),$$

- ▶ $\hat{\Xi}_n$ is a positive definite matrix that converges in probability to a deterministic positive definite matrix Ξ .

We want to find an estimate of the most efficient version of Ξ

- ▶ It turns out that this estimator is also an example of a GMM estimator.
- ▶ But why do we care about the estimate of the weight matrix?

It is important to calculate the weight matrix correctly

- ▶ Far too many structural papers calculate weight matrices and standard errors incorrectly.
- ▶ Do not bootstrap the weight matrix. (Horowitz 2001). It is consistent but can be very biased in finite samples.
- ▶ Calculating the weight matrix from simulated data is in principle fine, but you are taking the model too seriously.
- ▶ I am going to use basic GMM theory to teach you how to calculate weight matrices correctly and easily.
- ▶ We will be stacking influence functions, and this will be new to most of you.

The Setup

- ▶ The following uses the notation in Wooldridge.
- ▶ Let
 - ▶ Let \mathbf{w}_i be an $(M \times 1)$ vector of random variables for observation i .
 - ▶ $\boldsymbol{\theta}$ be an $(P \times 1)$ vector of unknown coefficients.
 - ▶ $\mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})$ be an $(L \times 1)$ vector of functions $\mathbf{g} : (\mathcal{R}^M \times \mathcal{R}^P) \rightarrow \mathcal{R}^L, \quad L \geq P$
- ▶ The function $\mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})$ can be nonlinear.
- ▶ Let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$.
- ▶ Let $\hat{\boldsymbol{\theta}}$ represent an estimate of $\boldsymbol{\theta}$.
- ▶ The “hat” and “naught” notation applies to anything we might want to estimate.

Moment Restrictions

- ▶ GMM is based on what are generally called moment restrictions and sometimes called orthogonality conditions (The latter terminology comes from the rational expectations literature.)

$$E(g(w_i, \theta_0)) = 0$$

- ▶ This condition is expressed in terms of the population. The corresponding sample moment restriction is

$$\frac{1}{N} \sum_{i=1}^N g(w_i, \theta) = 0$$

- ▶ What we want to do is choose $\hat{\theta}$ to get $N^{-1} \sum_{i=1}^N g(w_i, \theta)$ as close to zero as possible.

Examples of Moment Restrictions

► IV estimation:

- Suppose you have a regression

$$y_i = x_i\beta + u_i,$$

and $E(u_i | x_i) \neq 0$.

- Or, suppose you have a **nonlinear** regression

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

and $E(u_i | x_i) \neq 0$.

- In **either** case suppose also that you have a vector of instruments z_i , that is uncorrelated with u_i , and whose dimension is at least as great as β . Then the moment restriction is

$$E(z_i u_i) = 0$$

Criterion Function

- ▶ The estimator, $\hat{\theta}$ minimizes a quadratic form:

$$Q_N(\theta) = \begin{matrix} \left[N^{-1} \sum_{i=1}^N g(w_i, \theta) \right]' & \hat{\Xi} & \left[N^{-1} \sum_{i=1}^N g(w_i, \theta) \right] \\ (1 \times L) & (L \times L) & (L \times 1) \end{matrix}$$

where $\hat{\Xi}$ is a positive definite matrix that converges in probability to Ξ_0

- ▶ In this case, Q_N converges in probability to

$$\{E[g(w_i, \theta)]\}' \Xi \{E[g(w_i, \theta)]\}$$

Exact and Overidentification

- ▶ If $L = P$, then the estimator is exactly identified, and we can find θ by solving

$$N^{-1} \sum_{i=1}^N g(w_i, \theta) = 0$$

- ▶ If $L > P$, the model is overidentified and if it is nonlinear, you usually have to use numerical techniques.
- ▶ If $g(w_i, \theta)$ has first derivatives with no closed form solutions, these numerical techniques can take a very long time.

Optimal Weighting Matrix

- ▶ The symbol Ξ represents any arbitrary, positive definite weighting matrix.
- ▶ The optimal weighting matrix is the inverse of the variance of $g(w_i, \theta)$. Call this variance

$$\Lambda \equiv E(g(w_i, \theta) g(w_i, \theta)').$$

- ▶ Estimating $\hat{\Lambda}$. Doing GMM is usually a bit circular. We want to minimize

$$Q_N(\theta) = \left[N^{-1} \sum_{i=1}^N g(w_i, \theta) \right]' \hat{\Lambda}^{-1} \left[N^{-1} \sum_{i=1}^N g(w_i, \theta) \right]$$

to get an estimate of θ . But we need an estimate of θ to estimate $\hat{\Lambda}$.

Estimating the Optimal Weighting Matrix

- ▶ You can estimate $\hat{\Lambda}$ by

$$\hat{\Lambda} \equiv \frac{1}{N} \sum_{i=1}^N [g(w_i, \theta)] [g(w_i, \theta)]'$$

- ▶ The usual procedure is as follows:
 - ▶ Estimate θ using $\hat{\Lambda} \equiv I$. (This θ is consistent but not efficient.)
 - ▶ Use this estimate of θ to estimate $\hat{\Lambda}$.
 - ▶ Re-estimate θ using the estimate of $\hat{\Lambda}$.
 - ▶ Keep going until θ converges.

Estimating the Optimal Weighting Matrix

- ▶ Researches used to use a two-step procedure.
- ▶ However, after many Monte Carlo studies over the years, most now iterate.
- ▶ Another possibility is just to minimize Q_N all at once.
- ▶ If the application permits, sometimes the optimal weighting matrix does not depend on unknown parameters, and no iteration is necessary.
- ▶ This is the case with the SMM estimators we are using in this class.

Asymptotic Distribution of GMM Estimators

- Define the following

$$\begin{aligned} G' &= \frac{\partial g(w_i, \theta)}{\partial \theta'} \\ G'_0 &= E(G') \end{aligned}$$

- Note $G' \equiv \partial g(w_i, \theta) / \partial \theta'$ is the Jacobian matrix and is a function of the data

$$\partial g(w_i, \theta) / \partial \theta' \equiv \begin{bmatrix} \partial g_1 / \partial \theta_1 & \partial g_1 / \partial \theta_2 & \dots & \partial g_1 / \partial \theta_P \\ \vdots & \vdots & \ddots & \vdots \\ \partial g_L / \partial \theta_1 & \partial g_L / \partial \theta_2 & \dots & \partial g_L / \partial \theta_P \end{bmatrix}$$

- G had dimension $P \times L$, and G' had dimension $L \times P$.

Asymptotic Distribution of GMM Estimators

- ▶ Then the asymptotic distribution of $\sqrt{N}(\hat{\theta} - \theta_0)$ is $N(0, V)$, in which

$$V \equiv [G_0 \Lambda^{-1} G_0']^{-1} \\ (P \times L)(L \times L)(L \times P)$$

- ▶ Derivation: Recall that we are trying to minimize:

$$Q_N(\theta) = \left[N^{-1} \sum_{i=1}^N g(w_i, \theta) \right]' \hat{\Xi} \left[N^{-1} \sum_{i=1}^N g(w_i, \theta) \right]$$

- ▶ How do you minimize anything? Take the derivative w.r.t θ and set the result equal to zero.

Asymptotic Distribution of GMM Estimators

- Take the derivative and set it to zero.

$$\begin{aligned} \frac{\partial Q_N(\mathbf{w}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= 0 \\ 2 \left[N^{-1} \sum_{i=1}^N \mathbf{G}(\mathbf{w}_i, \boldsymbol{\theta}) \right] \hat{\Xi} \left[N^{-1} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) \right] &= 0 \\ \left[N^{-1} \sum_{i=1}^N \mathbf{G}(\mathbf{w}_i, \boldsymbol{\theta}) \right] \hat{\Xi} \left[N^{-1} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) \right] &= 0 \end{aligned}$$

- Now take a **mean**-value (not Taylor) expansion of $\sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})$

$$\sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) = \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}}) + \sum_{i=1}^N \mathbf{G}'(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

in which $\bar{\boldsymbol{\theta}}$ is some vector between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$.

Asymptotic Distribution of GMM Estimators

- ▶ Now we substitute this mean value expansion into the first order condition.

- ▶ First order condition

$$\left[N^{-1} \sum_{i=1}^N \mathbf{G}(\mathbf{w}_i, \boldsymbol{\theta}) \right] \hat{\Xi} \left[N^{-1} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) \right] = 0$$

- ▶ Mean value expansion

$$\sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) = \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}}) + \sum_{i=1}^N \mathbf{G}'(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

- ▶ Substitution

$$\left[N^{-1} \sum_{i=1}^N \mathbf{G}(\mathbf{w}_i, \boldsymbol{\theta}) \right] \hat{\Xi} \left[N^{-1} \left(\sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}}) + \sum_{i=1}^N \mathbf{G}'(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \right) \right] = 0$$

Asymptotic Distribution of GMM Estimators

- Now replace random averages with their plims, so there should be an $o_p(1)$ term floating around.

$$\mathbf{G}_0 \Xi_0 \left[N^{-1} \left(\sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}}) \right) + \mathbf{G}'_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \right] = o_p(1)$$

and solve away.

$$\begin{aligned} \mathbf{G}_0 \Xi_0 \mathbf{G}'_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0) &= -\mathbf{G}_0 \Xi_0 \left[N^{-1} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}}) \right] + o_p(1) \\ (\boldsymbol{\theta} - \boldsymbol{\theta}_0) &= -(\mathbf{G}_0 \Xi_0 \mathbf{G}'_0)^{-1} \mathbf{G}_0 \Xi_0 \left[N^{-1} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}}) \right] + o_p(1) \\ \sqrt{N} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) &= -(\mathbf{G}_0 \Xi_0 \mathbf{G}'_0)^{-1} \mathbf{G}_0 \Xi_0 \left[N^{-1/2} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}}) \right] + o_p(1) \end{aligned}$$

- The right-hand side of this last line contains what is called an “influence function.”

Asymptotic Distribution of GMM Estimators

- So the variance of the GMM estimator can be obtained by (roughly) by covarying the influence functions!¹

$$E (\boldsymbol{\theta} - \boldsymbol{\theta}_0) (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \equiv E \left\{ (\mathbf{G}_0 \boldsymbol{\Xi}_0 \mathbf{G}_0')^{-1} \mathbf{G}_0 \boldsymbol{\Xi}_0 \left[N^{-1} \sum_{i=1}^N \mathbf{g}_0 (\mathbf{w}_i, \bar{\boldsymbol{\theta}}) \right] \left[N^{-1} \sum_{i=1}^N \mathbf{g}_0 (\mathbf{w}_i, \bar{\boldsymbol{\theta}}) \right]' \boldsymbol{\Xi}_0 \mathbf{G}_0' (\mathbf{G}_0 \boldsymbol{\Xi}_0 \mathbf{G}_0')^{-1} \right\} = (\mathbf{G}_0 \boldsymbol{\Xi}_0 \mathbf{G}_0')^{-1} \mathbf{G}_0 \boldsymbol{\Xi}_0 \boldsymbol{\Lambda} \boldsymbol{\Xi}_0 \mathbf{G}_0' (\mathbf{G}_0 \boldsymbol{\Xi}_0 \mathbf{G}_0')^{-1}$$

- Note that if we set $\boldsymbol{\Xi}_0 \equiv \boldsymbol{\Lambda}_0^{-1}$, then this mess reduces as follows:

$$\begin{aligned} (\mathbf{G}_0 \boldsymbol{\Lambda}_0^{-1} \mathbf{G}_0')^{-1} \mathbf{G}_0 \boldsymbol{\Lambda}_0^{-1} \boldsymbol{\Lambda}_0 \boldsymbol{\Lambda}_0^{-1} \mathbf{G}_0' (\mathbf{G}_0 \boldsymbol{\Lambda}_0^{-1} \mathbf{G}_0')^{-1} &= \\ (\mathbf{G}_0 \boldsymbol{\Lambda}_0^{-1} \mathbf{G}_0')^{-1} \mathbf{G}_0 \boldsymbol{\Lambda}_0^{-1} \mathbf{G}_0' (\mathbf{G}_0 \boldsymbol{\Lambda}_0^{-1} \mathbf{G}_0')^{-1} &= \\ (\mathbf{G}_0 \boldsymbol{\Lambda}_0^{-1} \mathbf{G}_0')^{-1} & \end{aligned}$$

Asymptotic Distribution of GMM Estimators

- ▶ $(\mathbf{G}_0 \mathbf{\Xi}_0 \mathbf{G}_0')^{-1} \mathbf{G}_0 \mathbf{\Xi}_0 \mathbf{\Lambda}_0 \mathbf{\Xi}_0 \mathbf{G}_0' (\mathbf{G}_0 \mathbf{\Xi}_0 \mathbf{G}_0')^{-1}$ is always greater than $(\mathbf{G}_0 \mathbf{\Lambda}_0^{-1} \mathbf{G}_0')^{-1}$, in the sense that the difference between the two is a positive definite matrix.
- ▶ So an efficient estimate of the variance of $\hat{\boldsymbol{\theta}}$ is given by

$$\frac{1}{N} \left\{ \hat{\mathbf{G}} \hat{\mathbf{\Lambda}}^{-1} \hat{\mathbf{G}}' \right\}^{-1}$$

- ▶ What if we want to calculate the variance of an $R \times 1$ dimensional function $\mathbf{r}(\hat{\boldsymbol{\theta}})$, $R \leq P$?
- ▶ We can use the “delta-method,” which gives the variance of $\mathbf{r}(\hat{\boldsymbol{\theta}})$ as

$$\left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}} \right) \left\{ \frac{1}{N} \left\{ \mathbf{G}_0 \hat{\mathbf{\Lambda}}_0^{-1} \mathbf{G}_0' \right\}^{-1} \right\} \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}} \right)'$$

Overidentifying Restrictions

- ▶ If $L > P$, the model is overidentified. We have more equations than unknowns.
- ▶ Presumably we could take different subsets of P equations and solve exactly for the P elements of θ .
- ▶ Testing the overidentifying restrictions intuitively is a matter of testing to see if different exactly identified subsets of moment restrictions have the same solution.
- ▶ If the model is correct, then each of these answers should be the same.
 - ▶ What does this idea tell you about an informal way to see if the overidentifying restrictions are rejected?

Hansen's J-Test

- ▶ The following statistic

$$N \left(\frac{1}{N} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) \right)' \hat{\boldsymbol{\Lambda}}^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) \right)$$

converges to a χ^2 statistic with $(L - P)$ degrees of freedom under the null that the overidentifying restrictions hold.

- ▶ This is what is called a portmanteau test. “Wearovercoat”
- ▶ It tests for general misspecification, not any specific sort.
- ▶ The GMM J-test therefore need not be very powerful to detect misspecification.
- ▶ I will talk about the power of the test in the context of SMM later.

Why am I torturing you with influence functions?

- ▶ Our GMM/SMM weight matrices will be very complicated. Example:
 - ▶ Moments = $M = [\text{Mean} \ \text{Variance} \ \text{RegressionSlope1} \ \text{RegressionSlope2}]$
 - ▶ Optimal weight matrix = $\text{cov}(M)^{-1}$
 - ▶ How to estimate covariance between a mean and a variance?
 - ▶ How to estimate covariance between slopes from two separate regressions?
 - ▶ Etc.
 - ▶ Especially hard if we need to cluster by firm, etc.
- ▶ Influence functions give you a simple way to estimate $\text{cov}(M)$
- ▶ Beyond structural estimation, if you know how to use influence functions, **you know how to estimate the standard error for anything!**
 - ▶ Test for serial correlation of the residuals of a nonlinear panel model?
 - ▶ Hausman test when the Hausman assumptions are not satisfied.

General Definition of an Influence Function

- ▶ Consider any estimator $\hat{\theta}$, and suppose there is a function $\phi(\mathbf{w}_i)$ such that

$$\sqrt{N}(\hat{\theta} - \theta_0) = \sum_{i=1}^N \phi(\mathbf{w}_i) / \sqrt{N} + o_p(1), \quad E(\phi(\mathbf{w}_i)) = 0, \quad E(\phi(\mathbf{w}_i)\phi(\mathbf{w}_i)') \text{ exists.}$$

- ▶ then $\phi(\mathbf{w}_i)$ is called the influence function of $\hat{\theta}$.
- ▶ In words, it gives the effect of a single observation on the estimator, up to the $o_p(1)$ remainder term.
- ▶ In different words, an influence function is a function of the data whose mean has the same asymptotic variance as the estimator.
- ▶ An estimator that has an influence function is called an asymptotically linear estimator.

Influence Function for a GMM Estimator

- Now reconsider the expression

$$\sqrt{N}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = -(\mathbf{G}_0\boldsymbol{\Xi}_0\mathbf{G}_0')^{-1}\mathbf{G}_0\boldsymbol{\Xi}_0\left[N^{-1/2}\sum_{i=1}^N\mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}})\right] + o_p(1)$$

and compare it to the general expression for an influence function:

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \sum_{i=1}^n \phi(\mathbf{w}_i)/\sqrt{N} + o_p(1)$$

- So the influence function for a GMM estimator must be:

$$-(\mathbf{G}_0\boldsymbol{\Xi}_0\mathbf{G}_0')^{-1}\mathbf{G}_0\boldsymbol{\Xi}_0\mathbf{g}(\mathbf{w}_i, \bar{\boldsymbol{\theta}})$$

- End of proof by staring. For a real, extremely brief, but logically similar proof, see Newey and McFadden (1994).

Example

- Recall the definition of an influence function :

$$- (G_0 \Xi_0 G_0')^{-1} G_0 \Xi_0 g(w_i, \bar{\theta})$$

- What is the influence function for the estimate of the mean of a random variable z_i with mean μ ?

$$g(w_i, \bar{\theta}) \equiv z_i - \mu$$

$$\Xi \equiv 1$$

$$G \equiv -1$$

$$\phi(w_i, \bar{\theta}) \equiv (z_i - \mu)$$

The sample counterpart for observation i is

$$z_i - N^{-1} \sum_{i=1}^N z_i$$

Example

- Recall the definition of an influence function :

$$- \left(\mathbf{G}_0 \mathbf{\Xi}_0 \mathbf{G}_0' \right)^{-1} \mathbf{G}_0 \mathbf{\Xi}_0 \mathbf{g} \left(\mathbf{w}_i, \bar{\boldsymbol{\theta}} \right)$$

- Consider a simple linear regression

$$y_i = x_i \beta + u_i$$

- What is the influence function for β ?

$$\begin{aligned} \mathbf{g} \left(\mathbf{w}_i, \bar{\boldsymbol{\theta}} \right) &\equiv x_i \cdot u_i = x_i \cdot (y_i - x_i \beta) \\ \mathbf{\Xi}_0 = \mathbf{\Lambda}_0^{-1} &\equiv \sigma^{-2} E \left(x_i' x_i \right)^{-1} \\ \mathbf{G}_0 &\equiv -E \left(x_i' x_i \right) \\ \phi \left(\mathbf{w}_i, \bar{\boldsymbol{\theta}} \right) &\equiv E \left(x_i' x_i \right)^{-1} (x_i \cdot (y_i - x_i \beta)) \end{aligned}$$

The sample counterpart for observation i is

$$\left(N^{-1} \sum_{i=1}^N (x_i' x_i) \right)^{-1} (x_i \cdot u_i)$$

where the operator \cdot is the Hadamard element-by-element operator.

Stacking

- ▶ What if you estimate the mean μ and the OLS coefficient β , and you want to know the covariance between these two estimates?
- ▶ Option 1: Just estimate them jointly in a big GMM system.
 - ▶ This option can be cumbersome if you have many moments.
- ▶ Option 2: Stack the influence functions and take the inner product.²
- ▶ Let $\hat{\phi}_\mu$ be the $N \times 1$ sample influence function for μ .
- ▶ Let $\hat{\phi}_\beta$ be the $N \times k$ sample influence function for β .

²The reference for this is Erickson and Whited (2002).

Stacking

- ▶ Let's define

$$\Phi_{\mu\beta} \equiv \begin{bmatrix} \left(z - N^{-1} \sum_{i=1}^N z_i \right) & \left(\left(N^{-1} \sum_{i=1}^N (x_i' x_i) \right)^{-1} (x \cdot u) \right) \end{bmatrix}$$

- ▶ Notice I dropped the i subscripts.
- ▶ The dimension of this matrix is $N \times (k + 1)$.
- ▶ The sample covariance matrix for $\begin{pmatrix} \mu \\ \beta \end{pmatrix}$ is then

$$\Phi_{\mu\beta}' \Phi_{\mu\beta} N^{-2}$$

Sample Matlab Code

```
% Mean influence function
```

```
n = size(z);
```

```
meaninflnc = z - mean(z).*ones(n,1);
```

```
% OLS influence function
```

```
b=inv(x'*x)*x'*y;
```

```
uhat = y - x*b;
```

```
olsinflnc=inv((x'*x)./n)*((x.*uhat(:,ones(size(x,2),1))));
```

```
% Big influence function
```

```
biginflnc = zeros(n,size(x,2)+1);
```

```
biginflnc(:,1) = meaninflnc;
```

```
biginflnc(:,2:(size(x,2)+1)) = olsinflnc;
```

```
% Covary the influence functions
```

```
avar = biginflnc'*biginflnc./(n^2);
```

Two-Step Estimation

- ▶ Suppose you are doing a GMM estimator, but you estimate one or more of the parameters separately via a different procedure, and then plug these estimates into your GMM moment equations.
- ▶ Why? Sometimes this type of exercise reduces the dimensionality of the problem substantially.
- ▶ How do you figure out the GMM covariance matrix?
- ▶ This is nontrivial because the GMM estimates inherit the sampling variability from the first step.

Two-Step Estimation

- ▶ Let δ be a parameter vector of dimension S that you estimate in a first step via a different procedure
- ▶ Then you plug δ into your moment vector to get

$$g(\theta, w_i, \delta)$$

and use this moment vector to estimate θ .

- ▶ The variance of the two-step estimator is

$$(G\Omega^{-1}G')^{-1}$$

- ▶ You can estimate Ω by

$$\hat{\Omega} \equiv \frac{1}{N} \sum_{i=1}^N \left[g(w_i, \theta) - \mathbb{E} \left(\frac{\partial g(\theta, w_i, \delta)}{\partial \delta} \right) \phi^\delta(\delta, w_i) \right] \left[g(w_i, \theta) - \mathbb{E} \left(\frac{\partial g(\theta, w_i, \delta)}{\partial \delta} \right) \phi^\delta(\delta, w_i) \right]'$$

in which ϕ^δ is the influence function for δ .

- ▶ A clear derivation of this estimator is in Newey and McFadden's chapter in the 4th volume of the *Handbook of Econometrics*.

Clustered Weight Matrices

- ▶ Everything I have taught you thus far is for *i.i.d.* data. Data are almost never *i.i.d.* in corporate finance.
- ▶ So how do you calculate a weight matrix and get your standard errors right if the data are not *i.i.d.*?
- ▶ We will consider the following case.
 - ▶ The sample consists of K groups (clusters) of n_k observations each ($N = n_1 + \dots + n_K$)
 - ▶ Observations are independent across groups but dependent within groups
 - ▶ $K \rightarrow \infty$, and n_k fixed for each k .

Clustered Weight Matrices

- ▶ We order observations by groups and use double-index notation so that

$$\mathbf{g}(\boldsymbol{\theta}, \mathbf{w}) \equiv \{\mathbf{g}(\boldsymbol{\theta}, \mathbf{w}_{1,1}), \dots, \mathbf{g}(\boldsymbol{\theta}, \mathbf{w}_{n_1,1}) \mid \dots \mid \mathbf{g}(\boldsymbol{\theta}, \mathbf{w}_{1,K}), \dots, \mathbf{g}(\boldsymbol{\theta}, \mathbf{w}_{n_k,K})\}$$

- ▶ Under cluster sampling, the observations $\mathbf{w}_{n,k}$ might be dependent within a cluster, k .
- ▶ I'm going to simplify notation

$$\mathbf{g}_{1,1} \equiv \mathbf{g}(\boldsymbol{\theta}, \mathbf{w}_{1,1})$$

$$\hat{\mathbf{g}}_{1,1} \equiv \mathbf{g}(\hat{\boldsymbol{\theta}}, \mathbf{w}_{1,1})$$

Clustered Weight Matrices

- Let

$$\bar{\mathbf{g}} = \sum_{j=1}^{n_k} \mathbf{g}_{j,k}$$

- Then we can define Λ as:

$$\Lambda = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^K E(\bar{\mathbf{g}}_k \bar{\mathbf{g}}_k').$$

- Note that $E(\bar{\mathbf{g}}_i \bar{\mathbf{g}}_j') = 0$ only if i and j belong to different clusters.

- Define:

$$\tilde{\mathbf{g}} = \sum_{j=1}^{n_k} \hat{\mathbf{g}}_{j,k}$$

- A consistent estimate of Λ is therefore:

$$\hat{\Lambda} = \frac{1}{N} \sum_{k=1}^K \tilde{\mathbf{g}}_k \tilde{\mathbf{g}}_k'.$$

Sample Matlab code for a balanced panel

Consider a panel with dimensions T and N :

```
k = size(biginflnc,2);
vmx = zeros(k,k);
for qq=1:capN;
    phii = sum(biginflnc((qq-1)*(capT)+1:qq*(capT),:));
    vmx = vmx + phii'*phii;
end;

vmx = vmx./((capN*capT)^2);
```

- Erickson, T., Whited, T.M., 2002. Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory* 18, 776–799.
- Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50, 1269–1286.
- Horowitz, J.L., 2001. The bootstrap, in: Heckman, J.J., Leamer, E. (Eds.), *Handbook of Econometrics*. Elsevier. volume 5 of *Handbook of Econometrics*, pp. 3159 – 3228.
- Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing, in: Engle, R., McFadden, D. (Eds.), *Handbook of Econometrics*, Vol. 4. North-Holland, Amsterdam, pp. 2111–2245.