Hansen Lecture Notes:
The General Method of Moments

John Heilbron

Spring 2017

1 Introduction

- The General Method of Moments (GMM) is a framework that allows for parameter identification and inference testing on statistical models. It is similar in spirit to several techniques studied so far (OLS, IV, 2SLS, GLS, or MLE) but is a general formulation of which these techniques are special cases.
- This formulation allows for simultaneous treatment of multiple regressors, multiple instruments, multiple independent variables, autocorrelation and time-series dependence, heteroskedasticity, non-linear relations between variables, non- and semi-parametric specifications.
- One of the challenges of learning GMM lies in keeping well-organized the many mathematical objects involved. To this end, it is helpful to bear in mind the analogy to identification and inference techniques already mastered.
  - (2-3) We define a GMM estimator (cf. OLS, IV, 2SLS, GLS, MLE).
  - (4,6) We derive the variance-covariance matrix of the normal asymptotic sampling distribution of this GMM estimator (cf. White, Moulton, Fisher).
  - (5,7-9) We construct consistent estimators of this variance-covariance matrix.
  - (10-12) We identify a minimum standard error, or efficiency bound, that the asymptotic distribution of any GMM estimator may achieve (cf. Gauss-Markov, Cramer-Rao).
- Another challenge of learning GMM is appreciating how the framework may be applied to time-series data. This requires familiarity with stationarity, ergodicity and the Birkhoff LLN as well as martingale extraction and the Billingsley CLT.

- We treat the derivation of standard errors and construction of consistent estimators for those standard errors separately for the iid (4-5) and time-series (6-9) cases.
- Deriving the standard errors requires the ever-daunting martingale extraction (6), but the bulk of the extra consideration devoted to time-series GMM pertains to describing conditions under which we may construct a consistent standard error estimator (7-9).

2 True Parameter Values and Moment Restrictions

- We consider a statistical model that generates observable data according to a true but unknown set of parameters, $\beta_0 \in \mathbb{P} \subseteq \mathbb{R}^k$. Our aim is to use the observations to learn about value of the unknown parameters.
- The following are objects derived from a statistical model that may be used to describe the true parameter values:
  - $\beta \in \mathbb{P} \subseteq \mathbb{R}^k$ is any candidate $k \times 1$ vector of parameters in parameter space, $\mathbb{P}$.
  - $\{x_i\}_{i=1}^N \in \mathbb{R}^n$ is a sequence of $n \times 1$ iid observations OR $\{x_t\}_{t=1}^N \in \mathbb{R}^n$ is a stationary, ergodic stochastic process of $n \times 1$ observations.
  - $f : \mathbb{P} \times \mathbb{R}^n \to \mathbb{R}^r$ is a function of our observations and parameters.
- (Def.) A moment restriction is a relation between true model parameters, β_0, and observations, $\{x_i\}_{i=1}^N$ OR $\{x_t\}_{t=1}^N$, that is asserted by a statistical model and therefore known ex-ante.
  - An unconditional moment restriction takes the form $E[f(x_i \mid \beta_0)] = \underline{0}$ OR $E[f(x_t \mid \beta_0)] = \underline{0}$.
  - A conditional moment restriction takes the form $E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0}$ where $\mathcal{F}_t = \{x_t, x_{t-1}, ..., x_2, x_1\}$ is a history of previous observations.
- N.B. We may derive a moment condition in the abstract notation introduced above from a familiar statistical model like the following: $\tilde{y}_i = \tilde{x}_i' \tilde{\beta}_0 + \tilde{u}_i$, $E[\tilde{u}_i \mid \tilde{x}_i] = 0$.
  - (Claim) If $x_i = \begin{bmatrix} \tilde{y}_i \\ \tilde{x}_i \end{bmatrix}$, $\beta = \begin{bmatrix} 1 \\ -\tilde{\beta} \end{bmatrix}$, $\beta_0 = \begin{bmatrix} 1 \\ -\tilde{\beta}_0 \end{bmatrix}$, and $f(x_i \mid \beta) = x_i'\beta$ then we obtain the following moment restriction, $E[f(x_i \mid \beta_0)] = 0$.
  - (Pf.) $E[f(x_i \mid \beta_0)] \overset{f(.|.)def}{=} E[x_i'\beta_0] \overset{\beta_0 def}{=} E\left[x_i' \begin{bmatrix} 1 \\ -\tilde{\beta}_0 \end{bmatrix}\right] \overset{x_i def}{=} E[\tilde{y}_i - \tilde{x}_i'\tilde{\beta}_0] \overset{\tilde{y}_i = \tilde{x}_i'\tilde{\beta}_0 + \tilde{u}_i}{=} E[\tilde{u}_i] \overset{LIE}{=} E[E[\tilde{u}_i \mid \tilde{x}_i]] \overset{E[\tilde{u}_i | \tilde{x}_i]=0}{=} E[0] = 0$.

3 Sample Moment Restrictions, Selection Matrices, and the GMM Estimator

- (Def.) A sample moment restriction, $\frac{1}{N}\sum_{i=1}^{N} f(x_i \mid \beta) = \underline{0}$ OR $\frac{1}{N}\sum_{t=1}^{N} f(x_t \mid \beta) = \underline{0}$, is an ex-post relationship imposed on observations and candidate model parameters.
  - The sample moment restriction establishes a system of r equations (the dimension of $f(. \mid .)$) in k unknowns (the dimension of β).
  - When $r < k$, the sample moment restriction does not have a unique solution.
  - When $r > k$, the sample moment restriction may not have any solution.
- (Def.) A selection matrix, $A_{k\times r}$, is any $k \times r$ matrix ensuring that the augmented sample moment restriction, $A\frac{1}{N}\sum_{i=1}^{N} f(x_i \mid \beta) = \underline{0}$ OR $A\frac{1}{N}\sum_{t=1}^{N} f(x_t \mid \beta) = \underline{0}$, has a unique solution (up to a scalar).
  - N.B. $A\frac{1}{N}\sum_{i=1}^{N} f(x_i \mid c\beta^*) = 0 \overset{f(x_i|\beta) HOD1 in \beta}{\Longleftrightarrow} cA\frac{1}{N}\sum_{i=1}^{N} f(x_i \mid \beta^*) = 0 \overset{alg.}{\Longleftrightarrow} A\frac{1}{N}\sum_{i=1}^{N} f(x_i \mid \beta^*) = 0$, i.e. both β^* and $c\beta^*$ are solutions.
  - In such a case, we can pin down a unique solution if we know one element of β_0 by scaling the solution accordingly.
- (Def.) A GMM estimator, $\widehat{\beta}_N^{GMM}$ or $\widehat{\beta}_N^{A} \in \mathbb{P}$, is a candidate parameter that solves the sample moment restriction augmented by a selection matrix, $A\frac{1}{N}\sum_{i=1}^{N} f(x_i \mid \beta) = \underline{0}$ OR $A\frac{1}{N}\sum_{t=1}^{N} f(x_t \mid \beta) = \underline{0}$.
- (Claim) A GMM estimator is consistent, $\widehat{\beta}_N^{A} \xrightarrow{p} \beta_0$
  - (Pf.) Asserted but not proven in class.
  - N.B. This argument supposes we can get around the issue of scaling raised above.

4 The Asymptotic Sampling Distribution of the GMM Estimator (iid)

- For a sequence of iid observations, $\{x_i\}_{i=1}^{N}$, we use what is known ex-ante ($f(. \mid \beta), E[f(x_i \mid \beta_0)] = \underline{0}$), what is chosen ($A$), and what is unknown ($\beta_0$), to derive the asymptotic sampling distribution of our estimator, $\widehat{\beta}_N^{A}$.
- The objects below are useful ingredients in this derivation.
  - $A_{k\times r}$ is a selection matrix chosen to generate the GMM estimator, $\widehat{\beta}_N^{A}$.

- $D_{r\times k} = E\big[\frac{\partial}{\partial\beta'} f(x_i \mid \beta)\big|_{\beta_0}\big]$ is the expected gradient of our function evaluated at the true parameter value. (N.B. Assume this is well-defined i.e. non-infinite.)
- $V_{r\times r} = Var(f(x_i \mid \beta_0))$ is the variance-covariance matrix of our function of the data evaluated at the true parameter value. (N.B. Assume this is well-defined i.e. non-infinite.)
- $Cov(A)_{k\times k} = (AD)^{-1}AVA'(AD)'^{-1}$ is the variance-covariance matrix of the asymptotic distribution of the GMM estimator, $\widehat{\beta}_N^A$, corresponding to selection matrix A. (Its construction will be justified below.)

- (Lemma) $\sqrt{N}(\widehat{\beta}_N^A - \beta_0) \overset{p}{\to} (AD)^{-1}A\frac{1}{\sqrt{N}}\sum_{i=1}^N f(x_i \mid \beta_0)$
  - (Pf.) $\underline{0} \overset{\widehat{\beta}_N^A def.}{=} A\frac{1}{N}\sum_{i=1}^N f(x_i \mid \widehat{\beta}_N^A) \overset{MVT}{=} A\frac{1}{N}\sum_{i=1}^N f(x_i \mid \beta_0) + A\frac{1}{N}\sum_{i=1}^N \frac{\partial}{\partial\beta'} f(x_i \mid \beta)\big|_{\beta^*\in[\beta_0,\widehat{\beta}_N^A]}(\widehat{\beta}_N^A - \beta_0) \overset{lin.alg.}{\Rightarrow} \sqrt{N}(\widehat{\beta}_N^A - \beta_0) = -\Big[A\frac{1}{N}\sum_{i=1}^N \frac{\partial}{\partial\beta'} f(x_i \mid \beta)\big|_{\beta^*\in[\beta_0,\widehat{\beta}_N^A]}\Big]^{-1}\Big[A\frac{1}{\sqrt{N}}\sum_{i=1}^N f(x_i \mid \beta_0)\Big] \underset{LLN,CMT,\widehat{\beta}_N^A,\beta^*\to\beta_0,Ddef.}{\overset{p}{\to}} (AD)^{-1}A\frac{1}{\sqrt{N}}\sum_{i=1}^N f(x_i \mid \beta_0)$

- (Lemma) $\frac{1}{\sqrt{N}}\sum_{i=1}^N f(x_i \mid \beta_0) \underset{CLT}{\overset{d}{\to}} \mathcal{N}(0, V)$

- (Claim) $\sqrt{N}(\widehat{\beta}_N^A - \beta_0) \overset{d}{\to} \mathcal{N}(0, Cov(A))$
  - (Pf.) $\frac{1}{\sqrt{N}}\sum_{i=1}^N f(x_i \mid \beta_0) \underset{CLT}{\overset{d}{\to}} \mathcal{N}(0, V) \overset{CMT}{\Rightarrow} (AD)^{-1}A\frac{1}{\sqrt{N}}\sum_{i=1}^N f(x_i \mid \beta_0) \overset{d}{\to} \mathcal{N}(0, Cov(A)) \overset{CMT}{\Rightarrow} \sqrt{N}(\widehat{\beta}_N^A - \beta_0) \overset{d}{\to} \mathcal{N}(0, Cov(A))$

5 Estimating the Asymptotic Sampling Distribution of the GMM Estimator (iid)

- For a sequence of iid observations, $\{x_i\}_{i=1}^N$, we use what is known ex-ante ($f(. \mid \beta)$, $E[f(x_i \mid \beta_0)] = \underline{0}$), what is chosen ($A$), and what is revealed ex-post ($\{x_i\}_{i=1}^N$), to construct consistent estimators of the sampling distribution of the GMM estimator, $\widehat{\beta}_N^A$.
  - A is chosen. (N.B. We index the estimators below by A because A determines their construction through $\widehat{\beta}_N^A$.)
  - $\widehat{\beta}_N^A$ solves $A\frac{1}{N}\sum_{i=1}^N f(x_i \mid \beta) = \underline{0}$ and $\widehat{\beta}_N^A \overset{p}{\to} \beta_0$
  - $\widehat{D}_N^A = \frac{1}{N}\sum_{i=1}^N \frac{\partial}{\partial\beta'} f(x_i \mid \beta)\big|_{\widehat{\beta}_N^A} \underset{\widehat{\beta}_N^A\to\beta_0,LLN,CMT}{\overset{p}{\longrightarrow}} E\big[\frac{\partial}{\partial\beta'} f(x_i \mid \beta)\big|_{\beta_0}\big] = D_{r\times k}$

- $\widehat{V}_N^A = \frac{1}{N}\sum_{i=1}^N \left[f(x_i \mid \beta)f(x_i \mid \beta)'\right]\Big|_{\widehat{\beta}_N^A} \xrightarrow[\widehat{\beta}_N^A \to \beta_0, LLN, CMT]{p} E[f(x_i \mid \beta_0)f(x_i \mid \beta_0)'] \overset{E[f(x_i|\beta_0)]=\underline{0}}{=} Var(f(x_i \mid \beta_0)) = V_{r\times r}$
- $\widehat{Cov(A)}_N^A = (A\widehat{D})^{-1}A\widehat{V}A'(A\widehat{D})'^{-1} \underset{CMT}{\overset{p}{\to}} (AD)^{-1}AVA'(AD)'^{-1} = Cov(A)$

6 (DRAFT) The Asymptotic Sampling Distribution of the GMM Estimator (time-series)

- For a stochastic process, $\{x_t\}_{t=1}^N$, that is stationary and ergodic, we use what is known ex-ante $(f(. \mid \beta), E[f(x_t \mid \beta_0)] = \underline{0})$, what is chosen (A), and what is unknown (β_0), to derive the asymptotic sampling distribution of our estimator, $\widehat{\beta}_N^A$.
- The objects below are useful ingredients in this derivation.
  - $A_{k\times r}$ is a selection matrix as in the iid case.
  - $D_{r\times k}$ is an expected gradient as in the iid case.
  - $V_{r\times r} = Var(\kappa_1(x_t \mid \beta_0))$ is the variance-covariance matrix of the martingale component of the increment to the additive functional $\{Y_t\}_{t=1}^N$ where $Y_t = \sum_{j=1}^t f(x_j \mid \beta_0)$. (N.B. Assume this is well-defined i.e. non-infinite.)
  - $Cov(A)_{k\times k}$ is the variance-covariance matrix of the asymptotic distribution of the GMM estimator as in the iid case.
- (Lemma) $\sqrt{N}(\widehat{\beta}_N^A - \beta_0) \overset{p}{\to} (AD)^{-1}A\frac{1}{\sqrt{N}}\sum_{i=1}^N f(x_t \mid \beta_0)$
  - (Pf.) Same as (iid)
- DRAFT (Lemma) For a stationary, ergodic, stochastic process, $\{x_t\}_{t=1}^N$, $\frac{1}{\sqrt{N}}\sum_{t=1}^N f(x_t \mid \beta_0) \overset{d}{\to} \mathcal{N}(0, V)$ where $V = Var(\kappa_1(x_t \mid \beta_0))$
  - (Pf.) $f(x_t \mid \beta_0) = (\mathbb{I} - \mathbb{T})^{-1}f(x_t \mid \beta_0) - \mathbb{T}(\mathbb{I} - \mathbb{T})^{-1}f(x_t \mid \beta_0) \overset{add/sub}{=} (\mathbb{I}-\mathbb{T})^{-1}f(x_t \mid \beta_0) - (\mathbb{I}-\mathbb{T})^{-1}f(x_{t+1} \mid \beta_0) + \Big[(\mathbb{I}-\mathbb{T})^{-1}f(x_{t+1} \mid \beta_0) - \mathbb{T}(\mathbb{I}-\mathbb{T})^{-1}f(x_t \mid \beta_0)\Big] = (\mathbb{I}-\mathbb{T})^{-1}f(x_t \mid \beta_0) - (\mathbb{I}-\mathbb{T})^{-1}f(x_{t+1} \mid \beta_0) + \kappa_1(x_t \mid \beta_0)$
  - (Pf.) $\frac{1}{\sqrt{N}}\sum_{t=1}^N f(x_t \mid \beta_0) = \frac{1}{\sqrt{N}}\sum_{t=1}^N \kappa_1(x_t \mid \beta_0) + \frac{1}{\sqrt{N}}(\mathbb{I} - \mathbb{T})^{-1}f(x_1 \mid \beta_0) - \frac{1}{\sqrt{N}}(\mathbb{I}-\mathbb{T})^{-1}f(x_{N+1} \mid \beta_0) \underset{\frac{1}{\sqrt{N}}(\mathbb{I}-\mathbb{T})^{-1}f(x_1|\beta_0), \frac{1}{\sqrt{N}}(\mathbb{I}-\mathbb{T})^{-1}f(x_{N+1}|\beta_0)\overset{p}{\to}0, CMT}{\overset{p}{\to}} \frac{1}{\sqrt{N}}\sum_{t=1}^N \kappa_1(x_t \mid \beta_0) \underset{BillingsleyCLT}{\overset{d}{\to}} \mathcal{N}(0, V)$
- (Claim) $\sqrt{N}(\widehat{\beta}_N^A - \beta_0) \overset{d}{\to} \mathcal{N}(0, Cov(A))$ where $Cov(A) = (AD)^{-1}AVA'(AD)'^{-1}$ and $V = Var(\kappa_1(x_t \mid \beta_0))$
  - (Pf.) Same as (iid) after modifying definition of V

7 Properties of Conditional Moment Restrictions

- (Def.) A conditional moment restriction takes the form $E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0} \ \forall t$ where $\mathcal{F}_t = \{x_t, x_{t-1}, ..., x_2, x_1\}$.
- (Claim) A conditional moment restriction implies an unconditional moment restriction, $E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0} \Rightarrow E[f(x_{t+l} \mid \beta_0)] = \underline{0}$.
  - (Pf.) $E[f(x_{t+l} \mid \beta_0)] \overset{LIE}{=} E\big[E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t]\big] \overset{E[f(x_{t+l}|\beta_0)|\mathcal{F}_t]=\underline{0}}{=} E[\underline{0}] = \underline{0}$
- (Claim) A conditional moment restriction l periods in the future implies a conditional moment restriction more than l periods in the future, $E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0} \ \forall t \Rightarrow E[f(x_{t+l+j} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0} \ \forall t \ \forall j \geq 0$.
  - (Pf.) $E[f(x_{t+l+j} \mid \beta_0) \mid \mathcal{F}_t] \overset{LIE}{=} E\big[E[f(x_{t+l+j} \mid \beta_0) \mid \mathcal{F}_t, \mathcal{F}_{t+j}] \mid \mathcal{F}_t\big] \overset{\mathcal{F}_t \in \mathcal{F}_{t+j}}{=} E\big[E[f(x_{t+l+j} \mid \beta_0) \mid \mathcal{F}_{t+j}] \mid \mathcal{F}_t\big] \overset{E[f(x_{t+l}|\beta_0)|\mathcal{F}_t]=\underline{0} \ \forall t}{=} E[\underline{0} \mid \mathcal{F}_t] = \underline{0}$
- (Claim) A conditional moment restriction sets the "covariances" at times sufficiently far apart to 0, $E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0} \ \forall t \Rightarrow E[f(x_{t+n} \mid \beta_0) f(x_t \mid \beta_0)'] = 0_{r \times r} \ \forall t \ \forall n \geq l$ or $n \leq -l$.
  - (Pf.) $E[f(x_{t+n} \mid \beta_0) f(x_t \mid \beta_0)'] \overset{LIE}{=} E\big[E[f(x_{t+n} \mid \beta_0) \mid \mathcal{F}_t] f(x_t \mid \beta_0)'\big] \overset{E[f(x_{t+n}|\beta_0)|\mathcal{F}_t]=\underline{0} \ \forall t \ \forall n \geq l}{=} E[\underline{0} f(x_t \mid \beta_0)'] = E[0_{r \times r}] = 0_{r \times r}$ (symmetrically for $n \leq -l$)

8 Alternate Characterizations of the Martingale Component Variance

- (Def.) $V = Var(\kappa_1(x_t \mid \beta_0))$ where $\kappa_1(x_t \mid \beta_0)$ is the martingale component of $f(x_t \mid \beta_0)$.
  - Recall that $\kappa_1(x_t \mid \beta_0) = (\mathbb{I} - \mathbb{T})^{-1} f(x_{t+1} \mid \beta_0) - \mathbb{T}(\mathbb{I} - \mathbb{T})^{-1} f(x_t \mid \beta_0) \overset{(\mathbb{I}-\mathbb{T})^{-1}, \mathbb{T} def}{=} \sum_{j=0}^{\infty} E[f(x_{t+1+j} \mid \beta_0) \mid \mathcal{F}_{t+1}] - \sum_{j=0}^{\infty} E[f(x_{t+1+j} \mid \beta_0) \mid \mathcal{F}_t]$. One challenge to consistently estimating this last expression is that the expectations are conditional, so observations with identical histories are necessary to use Birkhoff LLN.
- (Claim) $V = \underset{N \to \infty}{p\text{-lim}} Var(\frac{1}{\sqrt{N}} \sum_{t=1}^{N} f(x_t \mid \beta_0))$
  - (Pf.) $\underset{N \to \infty}{p\text{-lim}} Var(\frac{1}{\sqrt{N}} \sum_{t=1}^{N} f(x_t \mid \beta_0)) \overset{\frac{1}{\sqrt{N}} \sum_{t=1}^{N} f(x_t|\beta_0) \overset{d}{\to} W \sim \mathcal{N}(0,V)}{=} Var(W) \overset{W \sim \mathcal{N}(0,V)}{=} V$

- This characterization eliminates conditional expectations, but observe that $\underset{N\to\infty}{p\text{-lim}} Var(\frac{1}{\sqrt{N}}\sum_{t=1}^{N} f(x_t \mid \beta_0)) = \underset{N\to\infty}{p\text{-lim}} \frac{1}{N}\sum_{t=1}^{N}\sum_{t'=1}^{N} E\big[f(x_t \mid \beta_0)f(x_{t'} \mid \beta_0)'\big]$ No matter how large we make N, we will only have one observation of $f(x_t \mid \beta)f(x_{t+N-1} \mid \beta)'$ and so cannot use Birkhoff LLN to consistently estimate the term $E\big[f(x_t \mid \beta_0)f(x_{t+N-1} \mid \beta_0)'\big]$.

- (Claim) $V = \sum_{j=-l}^{l} E[f(x_{t+j} \mid \beta_0)f(x_t \mid \beta_0)']$ if $E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0}$
  - (Pf.) $V = \underset{N\to\infty}{p\text{-lim}} Var(\frac{1}{\sqrt{N}}\sum_{t=1}^{N} f(x_t \mid \beta_0)) \overset{E[\frac{1}{\sqrt{N}}f(x_t|\beta_0)]=0, Var def.}{=} \underset{N\to\infty}{p\text{-lim}} E\big[[\frac{1}{\sqrt{N}}\sum_{t=1}^{N} f(x_t \mid \beta_0)][\frac{1}{\sqrt{N}}\sum_{t=1}^{N} f(x_t \mid \beta_0)]'\big] \overset{alg.}{=} \underset{N\to\infty}{p\text{-lim}} \frac{1}{N}\sum_{t=1}^{N}\sum_{t'=1}^{N} E\big[f(x_t \mid \beta_0)f(x_{t'} \mid \beta_0)'\big] \overset{x_t stationary}{=} \underset{N\to\infty}{p\text{-lim}} \frac{1}{N}\sum_{j=-N+1}^{N-1}(N-|j|)E\big[f(x_{t+j} \mid \beta_0)f(x_t \mid \beta_0)'\big] \overset{E[f(x_{t+n}|\beta_0)f(x_t|\beta_0)']=0_{r\times r}\ \forall t\ \forall |n|\geq l}{=} \underset{N\to\infty}{p\text{-lim}} \frac{1}{N}\sum_{j=-l}^{l}(N-|j|)E\big[f(x_{t+j} \mid \beta_0)f(x_t \mid \beta_0)'\big] \overset{\frac{N-|j|}{N}\to 1}{=} \sum_{j=-l}^{l} E[f(x_{t+j} \mid \beta_0)f(x_t \mid \beta_0)']$
  - N.B. We are summing elements of an array indexed by t and t'. (Each element is itself the expectation of a matrix.) Stationarity implies that the elements along diagonals from top left to bottom right are identical, so we sum by diagonal rather than by row. The conditional moment restriction ensures many of these diagonals have elements that are 0 matrices. Diagonals with non-zero elements approximate in length the main diagonal, which has N entries.
  - $\sum_{j=-l}^{l} E[f(x_{t+j} \mid \beta_0)f(x_t \mid \beta_0)']$ lends itself to consistent estimation by a single time-series, $\{x_t\}_{t=1}^{N}$, because for $N >> l$, we obtain many observations of $f(x_{t+j} \mid \beta_0)f(x_t \mid \beta_0)'$ for each $-l \leq j \leq l$ and may use Birkhoff LLN.

9 Estimating the Asymptotic Sampling Distribution of the GMM Estimator (time-series)

- For a stochastic process, $\{x_t\}_{t=1}^{N}$, that is stationary and ergodic, we use what is known ex-ante ($f(. \mid \beta)$, $E[f(x_{t+l} \mid \beta_0) \mid \mathcal{F}_t] = \underline{0}$), what is chosen, ($A$), and what is revealed ex-post ($\{x_i\}_{i=1}^{N}$), to construct consistent estimators of the sampling distribution of the GMM estimator, $\widehat{\beta}_N^A$.
  - A is chosen.
  - $\widehat{\beta}_N^A$ same as (iid)
  - $\widehat{D}_N^A$ same as (iid)
  - $\widehat{V}_N^A = \sum_{j=-l}^{l} \frac{1}{N-|j|}\sum_{t=max\{-j,0\}}^{min\{N-j,N\}} \big[f(x_{t+j} \mid \beta)f(x_t \mid \beta)'\big]\Big|_{\widehat{\beta}_N^A} \underset{LLN, \widehat{\beta}_N^A\to\beta_0}{\overset{p}{\to}} \sum_{j=-l}^{l} E[f(x_{t+j} \mid \beta_0)f(x_t \mid \beta_0)'] \overset{E[f(x_{t+l}|\beta_0)|\mathcal{F}_t]=\underline{0}}{=} V_{r\times r}$

- $\widehat{Cov(A)}_N^A$ same as (iid) but use modified formulation of $\widehat{V}_N^A$

10 Properties of Selection Matrices

- (Claim) For any selection matrix, $A_{k\times r}$, and non-singular matrix $B_{k\times k}$, the estimator generated by selection matrix A, $\widehat{\beta}_N^A$, and the estimator generated by selection matrix $[BA]_{k\times r}$, $\widehat{\beta}_N^{BA}$, are the same estimator, $\widehat{\beta}_N^A = \widehat{\beta}_N^{BA}$.
  - (Pf.) $\widehat{\beta}_N^A = \widehat{\beta}_N \overset{def.}{\Longleftrightarrow} \widehat{\beta}_N$ solves $A\frac{1}{N}\sum_{i=1}^N f(x_i \mid \beta) = \underline{0} \underset{mult.byB^{-1}}{\overset{mult.byB}{\Longleftrightarrow}} \widehat{\beta}^N$ solves $BA\frac{1}{N}\sum_{i=1}^N f(x_i \mid \beta) = \underline{0} \overset{def.}{\Longleftrightarrow} \widehat{\beta}_N^{BA} = \widehat{\beta}_N$
- (Claim) For any selection matrix, $A_{k\times r}$, and non-singular matrix $B_{k\times k}$, the estimator generated by selection matrix A, $\widehat{\beta}_N^A$, and the estimator generated by selection matrix $[BA]_{k\times r}$, $\widehat{\beta}_N^{BA}$, have asymptotic distributions with the same variance covariance matrix, $Cov(A) = Cov(BA)$.
  - (Pf.) $\widehat{\beta}_N^A = \widehat{\beta}_N^{BA} \Rightarrow$ identical asymptotic sampling distributions $\Rightarrow$ $Cov(BA) = Cov(A)$.
  - (Pf.) $Cov(BA) \overset{Cov(.)def.}{=} (BAD)^{-1}BAV(BA)'(BAD)'^{-1} \overset{lin.alg.}{=} (AD)^{-1}B^{-1}BAVA'B'B'^{-1}(AD)'^{-1} \overset{^{-1}def}{=} (AD)^{-1}AVA'(AD)'^{-1} \overset{Cov(.)def.}{=} Cov(A)$
- (Claim) For any selection matrix, $A_{k\times r}$, there exists a selection matrix, $X_{k\times r}$, s.t. $XD = \mathbb{I}_k$ and $Cov(A) = Cov(X) = XVX'$.
  - (Pf.) Let $X = (AD)^{-1}A$ and then $XD \overset{X def.}{=} (AD)^{-1}AD \overset{^{-1}def.}{=} \mathbb{I}_k$ and $Cov(A) \overset{Cov(A)=Cov(BA)}{=} Cov((AD)^{-1}A) \overset{X def.}{=} Cov(X) \overset{Cov(.)def.}{=} (XD)^{-1}XVX'(XD)'^{-1} \overset{XD=\mathbb{I}_k}{=} XVX'$

11 The GMM Efficiency Bound

- (Def.) An efficiency bound on a class of estimators is a variance-covariance matrix that is "(weakly) smaller than" the variance covariance matrix of the asymptotic sampling distribution of any estimator in the class (intuition NOT a technical definition).
  - The class of estimators considered here are GMM estimators, $\{\widehat{\beta}_N \mid \exists A \text{s.t.} A\frac{1}{N}\sum_{i=1}^N f(x_i \mid \widehat{\beta}_N) = \underline{0}\}$.
  - Each estimator, $\widehat{\beta}_N^A$, is indexed by the selection matrix, A, used to generate it.

- The asymptotic sampling distribution of a given estimator, $\widehat{\beta}_N^A$, has variance-covariance matrix $Cov(A)$

- (Claim) $(D'V^{-1}D)^{-1}$ is an efficiency bound of GMM estimators, $\{\widehat{\beta}_N \mid \exists A \text{s.t.} A\frac{1}{N}\sum_{i=1}^N f(x_i \mid \widehat{\beta}_N) = \underline{0}\}$
  - (Restated) For any selection matrix, A, the asymptotic distribution of the corresponding GMM estimator, $\widehat{\beta}_N^A$, has a variance covariance matrix, $Cov(A)$, s.t. $Cov(A) - (D'V^{-1}D)^{-1}$ is positive semi-definite.
  - (Pf.) $Cov(A) - (D'V^{-1}D)^{-1} \overset{\forall A \exists X s.t. XD=\mathbb{I}, Cov(A)=XVX'}{=} XVX' - (D'V^{-1}D)^{-1}$ $\overset{add/sub}{=} XVX' - (D'V^{-1}D)^{-1} - (D'V^{-1}D)^{-1} + (D'V^{-1}D)^{-1}$ $\overset{XD=\mathbb{I}, A^*:=(D'V^{-1}D)^{-1}D'V^{-1}}{=} XVX' - A^*VX' - XVA^{*\prime} + A^*VA^{*\prime}$ $\overset{alg.}{=} (X - A^*)V(X - A^*)'$ which is p.s.d. because V, a variance-covariance matrix, is p.s.d.

- (Claim) Any selection matrix of the form $A^* = BD'V^{-1}$ generates a GMM estimator, $\widehat{\beta}_N^{A^*}$, with an asymptotic distribution that attains the efficiency bound, $(D'V^{-1}D)^{-1}$.
  - (Pf.) $Cov(A^*) \overset{A^* def.}{=} Cov(BD'V^{-1}) \overset{Cov(BA)=Cov(A)}{=} Cov(D'V^{-1})$ $\overset{Cov(.) def.}{=} (D'V^{-1}D)^{-1}D'V^{-1}V(D'V^{-1})'(D'V^{-1}D)^{-1}$ $\overset{lin.alg.}{=} (D'V^{-1}D)^{-1}D'V^{-1}VV^{-1}D(D'V^{-1}D)^{-1} \overset{^{-1} def.}{=} (D'V^{-1}D)^{-1}$

12 Choosing an Efficient Selection Matrix

- We want to choose an efficient selection matrix, A^* s.t. $Cov(A^*) = (D'V^{-1}D)^{-1}$.
  - Our candidate selection matrix is $D'V^{-1}$. But we do not know the value of this object ex-ante and consistent estimation requires a consistent estimator of β_0. And though any GMM estimator, $\widehat{\beta}_N^{GMM}$, is a consistent estimator of β_0, obtaining an estimator requires us to choose a selection matrix, A. And this is the choice facing us to begin with.
  - N.B. These notes have suggested performing GMM as a two step process, first choosing a selection matrix, A, and then computing an estimator, $\widehat{\beta}_N^A$. When the goal is to choose the efficient selection matrix, A^*, as will be made clear by the techniques below, these steps are intertwined.

- Approach 1: Iteration
  - (1) Fix A^0. (2) Use A^0 and $\{x_i\}_{i=1}^N$ OR $\{x_t\}_{t=1}^N$ in the sample moment restriction to solve $\widehat{\beta}_N^{A^0}$. (3) Use $\widehat{\beta}_N^{A^0}$ and $\{x_i\}_{i=1}^N$ OR $\{x_t\}_{t=1}^N$ to

construct $\widehat{D}_N^{A^0}$ and $\widehat{V}_N^{A^0}$ according to formulae in (?) and (?). (4) Construct $A_N^1 = (\widehat{D}_N^{A^0})'(\widehat{V}_N^{A^0})^{-1}$. (5) Iterate this procedure.

  – Our candidate efficient weighting matrix is $\underset{N\to\infty}{p\text{-lim}} \lim_{s\to\infty} A_N^s$.
  – (Caveat) We may not have the result that our candidate efficient weighting matrix exists in general. (Is convergence in s guaranteed? Is convergence in N guaranteed? Is convergence independent of the initial choice of A_0 guaranteed?) We may not have the result that if the candidate matrix does exist that it is the efficient matrix, $\underset{N\to\infty}{p\text{-lim}} \lim_{s\to\infty} A_N^s = A^*$

- Approach 2: "Continuously-updated GMM"
  – Let $\beta_N^* = \arg\min_{\beta\in\mathbb{P}} \frac{1}{\sqrt{N}}\sum_{t=1}^N f(x_t \mid \beta)'\widehat{V}_N(\beta)^{-1}\frac{1}{\sqrt{N}}\sum_{t=1}^N f(x_t \mid \beta)$ where $\widehat{V}_N(\beta) = \sum_{j=-l}^{l} \frac{1}{N-|j|}\sum_{t=max\{-j,0\}}^{min\{N-j,N\}} \left[f(x_{t+j} \mid \beta) f(x_t \mid \beta)'\right]$.
  – Our candidate weighting matrix is $(\widehat{D}_N(\beta_N^*))'(\widehat{V}_N(\beta_N^*))^{-1}$
  – (Claim) $\beta_N^* \xrightarrow{p} \beta_0$ (asserted not proven in class), which guarantees consistency of our candidate weighting matrix.

13 Chi-Squared Distributions

- Chi-Squared Statistics
- Another phrasing of the problem

14 GMM Tests

- Testing parameters
- Testing restrictions