Econometrics Exam Sheet

Michael B. Nattinger

March 10, 2021

1 Lecture 1

- If $E[g(X,\theta)] = 0 \iff \theta = \theta_0, \hat{\theta}_{mm}$ is the solution to $0 = \frac{1}{n}\sum_{i=1}^n g(x,\theta)$

2 Lecture 2

- $E[Y] = E[E[Y|X]]$ is LIE
- $Var(Y) = Var(E[Y|X]) + E[Var(Y|X)]$
- $E[\bar{X}] = \frac{1}{n}\sum_{i=1}^n E[X_i], Var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n Cov(X_i, X_j)$
- Markov: $P(|X| \geq \epsilon) \leq \frac{E[|X|]}{\epsilon}$
- Chebyshev: $P(|X - E[X]| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$
- Holder: $E[|XY|] \leq E[|X|^p]^{1/p}E[|Y|^q]^{1/q}$
- Cauchy-Schwarz: $E[|XY|]^2 \leq E[X^2]E[Y^2]$
- Convergence in Probability: W_n has plim w if for any $\epsilon > 0$, $P(|W_n - w| \leq \epsilon) \geq 1 - \epsilon$ for large enough n.
- LLN: $\bar{X}_n \to_p E[X_1]$ if X_i is i.i.d.
- continuous mapping theorem applies so long as the functions are continuous at the plim
- $\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - E[X_1]) \to_d N(0, Var(X_1))$ (CLT)
- Cramer-Wold device: W_n converge in distribution to W iff $t'W_n$ converge in distribution to $t'W$ for nonrandom t with $||t|| = 1$.
- continuous mapping works for distributions in addition to plims
- Orthogonal projection onto the column space of X: $P = X(X'X)^{-1}X', P = QQ'$ for some Q with $Q'Q = I_k$. $P = P^2, tr(P) = k$
- Also, and this is not on the slides anywhere, but Jensen's inequality might be useful:
- if $\psi(x)$ is a convex function then $\psi(E[X]) \leq E[\psi(X)]$
- FOR LLN,CMT FOLLOW STEPS THAT THEY HAVE IN OLD EXAMS i.e. 2020 or something

2.1 Block Inversion

For $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, D must be invertible. Then, block inversion says that M is invertible iff $A - BD^{-1}C := E$ is invertible, in which case:

$$M^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}BD^{-1} \\ -D^{-1}CE^{-1} & D^{-1} + D^{-1}CE^{-1}BD^{-1} \end{pmatrix}$$

2.2 Sherman-Morrison formula

Let A be invertible and square, u, v vectors, then $A + uv'$ is invertible iff $1 + v'A^{-1}u \neq 0$, in which case

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1 + v'A^{-1}u}$$

3 Lecture 3

- $(Y, X, Z)'$ random vec s.t. $Y = \beta_0 + X\beta_1 + U$ where $E[U|Z] = 0$,$Cov(Z, X) \neq 0$, $E[Y^2 + X^2 + Z^2] < \infty$
- IV estimator is the MM estimator of $Cov(Z, Y - X\beta_1) = 0$: $\hat{\beta}_1^{iv} = \frac{\hat{Cov}(Z,Y)}{\hat{Cov}(Z,X)}$, $\hat{\beta}_0^{iv} = \bar{Y} - \bar{X}\hat{\beta}_1^{iv}$
- note that each X here is a single value (not vec)

4 Lecture 4

- $Y = X'\beta_0 + U$, $E[U|Z] = 0, E[ZX']$ invertible, $E[Y^2 + ||X||^2 + ||Z||^2] < \infty$
- $E[U|Z] = 0$ assumption is called independence
- $E[ZX']$ is invertible assumption is called relevance
- Can show that $E[Z(Y - X'\beta)] = 0 \iff \beta = \beta_0 \Rightarrow \beta_0 = E[ZX']^{-1}E[ZY]$
- The IV estimator is the mm analog: $\hat{\beta}^{iv} = \left(\frac{1}{n}\sum_{i=1}^n Z_iX_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n Z_iY_i$
- $E[\hat{\beta}^{iv}|X, Z] = \beta_0 + \left(\frac{1}{n}\sum_{i=1}^n Z_iX_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n Z_iE[U_i|X, Z] \neq \beta_0$ unless X_i is also exogenous (in which case one should just use OLS anyways).
- For large sample properties we just need existence of fourth moments on top of everything.
- We can easily show that $\hat{\beta}^{iv} \to_p \beta$
- $\sqrt{n}(\hat{\beta}^{iv} - \beta) = \left(\frac{1}{n}\sum_{i=1}^n Z_iX_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_iU_i$
- by C-W we can show that this converges in distribution to $N(0, E[ZX']^{-1}E[ZZ'U^2]E[XZ']^{-1})$
- Inference: construct confidence intervals and T statistics off of asymptotic distribution, but this seems to be far off in finite samples, often.

5 Lecture 5

- One weak instrument: can and should use test inversion.
- Want to test the hypothesis $H_0 : \beta_1 = c$. Under $H_0, 0 = E[Z_1(Y - X_1 c)]$.
- $T = \frac{1}{n}\sum_{i=1}^n Z_{1i}(Y_i - X_{1i}c)$
- $\sqrt{n}T = \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{1i}U_i \to_d N(0, E[Z_1^2 U^2])$
- $S^2 = \frac{1}{n}\sum_{i=1}^n Z_{1i}^2 \hat{U}_i^2, \hat{U}_i = Y_i - cX_{1i} - Z_{1i}(\hat{\gamma}_1 - c\hat{\pi}_1)$
- where $\hat{\gamma}_1, \hat{\pi}_1$ are OLS estimators from regressions of Y_i on Z_i and X_i on Z_i, respectively.
- $AR := \frac{\sqrt{n}T}{S} \to_d N(0,1)$. Anderson-Rubin test. Does not rely on an assumption that the instrument is relevant.
- Can use as a confidence interval the region of values of c that AR test does not reject.

6 Lecture 6

- Consider a transformation of potential instrument $h(Z)$. If we use this as an instrument, the avar is the following:
- $\Omega_h = E[h(Z)X']^{-1}E[h(Z)h(Z)']E[Xh(Z)']^{-1}$
- if we impose homoscedasticity and let X be one dimensional, $\Omega_h = \frac{E[h(Z)^2]}{E[h(Z)X]^2}\sigma_U^2$
- The instrument that minimizes Ω_h is then $h^*(Z) = E[X|Z]$, this also holds if X is a vector.
- Many instruments: let X_1 just be a single value but have there be many instruments.
- 2SLS: estimate $h(Z) = E[X|Z]$ via OLS and use either (1) as an instrument or (2) as a replacement for X entirely in the second stage regression!
- $\sqrt{n}(\hat{\beta}_1^{2sls} - \beta_1 \to_d N(0, \sigma_U^2 Var(Z'\pi_1)/Cov(Z'\pi_1, X)^2$
- The variance estimator is consistent so we can use this for confidence intervals, testing, etc.
- Note: the estimated $\hat{U}_i = Y_i - X_i\hat{\beta}_1^{2sls}$, not $\tilde{U}_i = Y_i - Z_i'\hat{\pi}_1\hat{\beta}_1^{2sls}$

7 Lecture 7

- If you assume normal errors, then the maximum likelihood estimator for β_1 is called LIML. Probs not on the exam.
- With only 1 instrument, LIML, IV, and 2sls are all identical.
- Test $H_0 : \beta_1 = c$ via likelihood ratio test statistic: $-2log(\max_{\theta:\beta_1=c} L(\theta)/\max_\theta L(\theta))$
- median unbiased, larger dispersion than 2slsm asymptotically normal with same asymptotic distribution as 2sls when instruments are relevant.
- Random coefficients and endogeneity: $Y = X'\beta_0 U$

- Plim of IV is $E[ZX']^{-1}E[ZY] = E[ZX']^{-1}E[ZE[X'U|Z]]\beta_0$
- Consider a binary instrument: $X = (1, X)', Z = (1, z)', z \in \{0, 1\}$.
- plim of IV is then $Cov(Z, Y)/Cov(Z, X)$.
- Uzing binary Z we then have the plim is $\frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}$
- the above is called wald estimator.
- If we further let $X \in \{0, 1\}$ we can assume no defying and U independent of z.
- Then, the plim is $E[Y(1) - Y(0)|X(1) - X(0) = 1]$

8 Lecture 8

- MA(q): $Y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$
- MA and AR simultaneously cannot be estimated in an unbiased fashion.
- A sequence of stochastic vectors is strictly stationary if $(Z_t, \ldots, Z_{t+k}) \sim (Z_1, \ldots, Z_{1+k})\forall t, k$
- a random sample is strictly stationary, a constant sequence is strictly stationary, a trending sequence is not strictly stationary.
- If Z_t is strictly stationary and $Y_t = \phi(Z_t, Z_{t-1}, \ldots)$ then Y_t is strictly stationary.
- For an ar(1) we iterate backwards to try to apply the above and so long as the parameter does not have a unit root then we can apply it and the series is stationary.
- A time series is cov stationary if $E[Y_t] = \mu\forall t, Cov(Y_t, Y_{t+k}) = \gamma(k)\forall t$, some function θ. γ is the autocovariance function.
- Moving average processes are covariance stationary.

9 Lecture 9

- Basic setup: $Y_t = W_t'\beta + U_t$, with strictly stationary everywhere. $E[U_t|W_t] = 0, E[W_tW_t']$ and sum is invertible. Contemporaneous exogeneity and excludes multicollinearity.
- $E[\hat{\beta}^{ols}|W] = \beta + (\sum W_tW_t')^{-1}\sum WE[U_t|W]$. Note that this is not same as cont exog, rather it is strict exogeneity. Can fail in general.
- In other words, there is some bias in finite samples.
- Cont exogeneity also can fail, e.g. ARMA model.
- Ar(1) (strict exog): $\hat{\rho}_1 = \rho_1 + \frac{\sum Y_{t-1}U_t}{\sum Y_{t-1}^2}$
- $\sqrt{T}(\hat{\rho}_1 - \rho_1)$ cannot simply use LLN and CLT as they are not independent.
- For the denominator, we have strict stationary so the expectation is unbiased. If variance goes to 0 then we can apply Chebyshev and get the plim.
- $Var((1/T)\sum Y_{t-1}^2) = \frac{1}{T}\gamma(0) + \frac{2}{T}\sum_{k=1}^{T-1}(1 - k/T)\gamma(k)$
- $\gamma(0)$ is short run variance, the whole term above is long run variance.

- $\gamma(k) = \rho_1^{2k}\gamma(0)$ is the case in our example, so L-R variance can be shown to go to 0.
- Martingale CLT: Z_t strictly stationary, $E[Z_t|Z_{t-1},\dots,Z_1] = 0$, $\frac{1}{T}\sum_{t=1}^{T} \to_p E[Z_1^2], E[Z_1^2] < \infty$, then $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \to_d N(0, E[Z_1^2])$ as $T \to \infty$.
- using the above, we find that $\sqrt{T}(\hat{\rho}_1 - \rho_1) \to_d N(0, E[U_t^2]/E[Y_{t-1}^2]) = N(0, 1-\rho_1^2)$

10 Lecture 10

- Serial correlation: let $Y_t = \alpha_0 + U_t$, let Y_t be strictly stationary with $E[Y_t^2] < \infty$ and $E[U_t] = 0$.
- The OLS estimator is the sample average, and is unbiased. Therefore, if its variance goes to 0 then we can apply Chebyshev to prove consistency.
- $Var(\hat{\alpha}) = \frac{1}{T}\left(\gamma(0) + 2\sum_{k=1}^{T-1}\left(1 - \frac{k}{T}\right)\gamma(k)\right)$ ($(1/T)$ times long-run variance)
- Can use the above formula and bound it by $\frac{2\gamma(0)}{\sqrt{T}} + \max_{k \geq \sqrt{T}} |\gamma(k)| \to_{T\to\infty} 0$ IF $\max_{k \geq \sqrt{T}} |\gamma(k)|$ goes to zero as $k \to \infty$
- The avar is the limit of the long-run variance, so for it to exist $\lim_{T\to\infty}\sum_{k=1}^{T}|\gamma(k)| < \infty$
- For a MA(∞) it requires $\sum_{k=1}^{T}|\theta_k| < \infty$.
- To estimate the LR Variance, we need to estimate T covariances from T obs so there is lots of noise.
- Trade-off noise with variance: NW standard error: $\Omega_{nw} = \hat{\gamma}(0) + 2\sum_{k=1}^{b_T}\left(1 - \frac{k}{b_T+1}\right)\hat{\gamma}(k)$. This is guaranteed to be nonnegative.
- $b_T = 0.75T^{1/3}$ is a 'good' choice.
- need to show that a series is asymptotically normal be rewriting the errors and taking the limit and applying CLT when you can.

11 Lecture 11

- $Y_t = X_t'\beta_0 + \alpha + \epsilon_t$, assume $E[\epsilon_t|X_1,\dots,X_T] = 0$, invertibility of X cov mat, existence of fourth moments, random sample from distribution.
- Fixed effects: imposes no further assumptions on α so it allows for endogeneity between $X_1,\dots,X_T,\alpha$
- Random effects: imposes independence between α, X_t', ϵ_t, and a white noise structure on $\epsilon_t : Cov(\epsilon_t,\epsilon_s|X_1,\dots,X_T) = \sigma^2 1\{s=t\}$
- If we were to try OLS it is conditionally unbiased, however it does not achieve the lower bound of the variance.
- RE is essentially the optimal GLS estimator which achieves the lower bound.
- The covariance matrix of the residuals is $\Sigma = \sigma^2 I_T + \sigma_\alpha^2 1_T 1_T'$. Thus, we can use the Sherman-Morrison formula and we have:
- $\Sigma^{-1} = \frac{1}{\sigma^2}I_t - \frac{1}{\sigma^4}\frac{\sigma_\alpha^2 1_T 1_T'}{1+T\sigma_\alpha^2/\sigma^2}$

- $\sum_{s=1}^{T}(\Sigma^{-1})_{ts}X_s = X_t - X_t - \frac{T\sigma_\alpha^2}{\sigma^2+T\sigma_\alpha^2}\bar{X}$
- When $\sigma_\alpha^2, \sigma^2$ are appropriately estimated. Then,
- $\hat{\beta}^{gls} = \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}X_{it}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}Y_{it}$
- where $\tilde{X}_{it} = X_{it} - \frac{T\sigma_\alpha^2}{\sigma^2+T\sigma_\alpha^2}\bar{X}_i$
- The fixed effects estimator instead subtracts off the entire (unweighted) average as the instrument $\tilde{X}_{it}$
- Under ϵ_t homoskedastic and serially uncorrelated, the estimator has minimal conditional variance amongst IV estimators.
- $\sqrt{n}(\hat{\beta}^{FE} - \beta_0) \to_d N(0, H^{-1}\Omega H^{-1})$, $H = E\left[\sum_{t=1}^{T}(X_t - \bar{X})(X_t - \bar{X})'\right]$
- $\Omega = E\left[\left(\sum_{t=1}^{T}(X_t - \bar{X})\epsilon_t\right)\left(\sum_{s=1}^{T}(X_s - \bar{X})\epsilon_s\right)'\right]$
- To estimate we can plug in $Y_t - X_t'\hat{\beta}^{FE}$ as our estimates of ϵ_t.

12 Lecture 12

- Due to time constraints I will not be taking notes on the rest. See lecture slides if they come up (I do not anticipate them coming up).

13 Lecture 13

-