1 Basics

Distributions

Normal: $\phi(x) = (2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$

Normal CDF: $\mathbf{\Phi}(\mathbf{x}) = \int_{-\infty}^{x}\phi(y)dy$

Moments

Continuous: $m_k(X) \equiv \int_{-\infty}^{\infty} x^k f(x)dx$
Central:
$\mu_k \equiv E(X - E(X)^k) = \int_{-\infty}^{\infty}(x-\mu)^k f(x)dx$
Discrete Central:
$\mu_k \equiv E(X - E(X)^k) = \sum_{i=1}^{m} p(x)(x_i - \mu)^k$

Multivariate

Joint Density Func: $f(x_1,x_2) = \frac{\partial^2 F(x_1,x_2)}{\partial x_1 \partial x_2}$
Independence:
$F(x_1,x_2) = F(x_1,\infty)F(\infty,x_2)$
or $f(x_1,x_2) = f(x_1)f(x_2)$
Marginal Density:
$f(x_1) \equiv F_1(x_1,\infty) = \int_{-\infty}^{\infty} f(x_1,x_2)dx_2$
Conditional Density: $f(x_1|x_2) = \frac{f(x_1,x_2)}{f(x_2)}$
Law Iterate Expec: $E(E(X_1|X_2)) = E(X_1)$
Any Deterministic Func h:
$E(X_1 h(X_2) \mid X_2) = h(X_2)E(X_1 \mid X_2)$

Matrix Algebra

Dot Product: $\mathbf{a}^\top\mathbf{b} = \sum_{i=1}^{} a_i b_i$
Matrix Multiplication: $C_{ij} = \sum_{k=1}^{m} A_{ik}B_{ki}$
Transpose: $(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top$
$\|\mathbf{x}\| = (\mathbf{x}^\top\mathbf{y})^{1/2} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$
$\langle \mathbf{x},\mathbf{y}\rangle = \|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta)$
Cauchy-Schwartz: $\langle \mathbf{x},\mathbf{y}\rangle \le \|\mathbf{x}\|\|\mathbf{y}\|$
Linear dep: $\mathbf{Xb} = \mathbf{0}$, $\mathbf{b} \neq \mathbf{0}$
Singular: $\exists \mathbf{x} \neq 0 : \mathbf{Ax} = \mathbf{0}$

2 Linear Regression

OLS

Estimator: $\hat\beta = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$
$E(x_{ti}u_t) = 0 \implies \mathbf{X}^\top(\mathbf{y} - \mathbf{X}\beta) = \mathbf{0}$
SSR(β): $= \sum_{t=1}^{n}(y_i - \mathbf{X}_t\beta)^2$
$\mathbf{y}^\top\mathbf{y} = \hat\beta^\top\mathbf{X}^\top\mathbf{X}\hat\beta + (\mathbf{y}-\mathbf{X}\hat\beta)^\top(\mathbf{y}-\mathbf{X}\hat\beta)$
Projection: $\mathbf{P_X} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
$\mathbf{M_X} = \mathbf{I} - \mathbf{P_X} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
$\mathbf{P_X y} = \mathbf{X}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y})$; $\mathbf{P_X P_X} = \mathbf{P_X}$
$\mathbf{M_X y} = \hat{\mathbf{u}}$; $\mathbf{M_X X} = \mathbf{0}$; $\mathbf{M_X M_X} = \mathbf{M_X}$
$\mathbf{P_X} + \mathbf{M_X} = \mathbf{I}$; $\mathbf{P_X M_X} = \mathbf{0}$; $\mathbf{P_X}^\top = \mathbf{P_X}$
$\|\mathbf{y}\|^2 = \|\mathbf{P_X y}\|^2 + \|\mathbf{M_X y}\|^2$; $\|\mathbf{P_X y}\| \le \|\mathbf{y}\|$
Centering: $\mathbf{M}_\iota\mathbf{x} = \mathbf{z} = x - \bar{x}_\iota$; $\iota^\top\mathbf{z} = 0$
$\mathbf{P}_1 \equiv \mathbf{P}_{\mathbf{X}_1}$; $\mathbf{P}_1\mathbf{P_X} = \mathbf{P_X}\mathbf{P}_1 = \mathbf{P}_1$
FWL: β_2 from $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u}$ and $\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{X}_2\beta_2 + \text{res}$ are the same. (+ res)
Seasonal w const: $\mathbf{s}_i' = \mathbf{s}_i - \mathbf{s}_4$, $i = 1,2,3$.
Avg is const coeff. $\mathbf{M_S y}$ is deseasonalized.

$\beta^{(t)} - \hat\beta = -1\backslash 1 - h_t(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}_t^\top\hat{u}_t$ where h_t denotes the t^{th} diagonal element of $\mathbf{P_X}$.

Bias

vector of (true) model params: θ
Bias: $E(\hat\theta) - \theta_0$, $E((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}) = 0$
estimating eq: $g(\mathbf{y},\theta) = 0$ unbiased iff $\forall\mu\in\mathbb{M}, E_\mu g(\mathbf{y},\theta_\mu) = \mathbf{0}$ or $E(\mathbf{X}^\top\mathbf{u}) = \mathbf{0}$
X exogenous $\implies E(\mathbf{u}\mid\mathbf{X}) = \mathbf{0}$ and both $\hat\beta$ and estimating equations unbiased.
Make exog assumpt in cross-sec not time.
regressors predetermined: $E(\mathbf{X}^\top\mathbf{u}) = \mathbf{0}$

Stochastic Limits

Converg in prob: $\lim Pr(|Y_n - Y_\infty| > \epsilon) = 0 \Rightarrow \text{plim}\, Y_n = Y_\infty \implies$ converg dist
Converg dist: $\lim F_n(x) = F(x) \equiv Y_n \to F$
LLN: $\text{plim}\,\overline{Y_n} = \text{plim}\frac{1}{n}\sum_t^n Y_t = \mu_Y$, Y_t IID, $\overline{Y_n}$ sample mean of Y_t, μ_T pop mean.
LLN2: $\text{plim}\frac{1}{n}\sum_t^n Y_t = \lim\frac{1}{n}\sum_t^n E(Y_t)$
$\text{plim}\, Y_n Z_n = \text{plim}\, Y_n \text{plim}\, Z_n$ if converg
$\mathbf{X}^\top\mathbf{X}$ may not have plim so mult by $1\backslash n$.
consistent: $\text{plim}_\mu\hat\beta = \beta_\mu$, may be bias
$E(\mu_t \mid \mathbf{X}_t) = 0 \implies \hat\beta$ consistent.

Covariance and Precision Matrices

$\text{Cov}(b_i,b_j) \equiv E((b_i - E(b_i))(b_j - E(b_j)))$
if $i = j$, $\text{Cov}(b_i,b_j) = \text{Var}(b_i)$
$\text{Var}(\mathbf{b}) \equiv E((\mathbf{b} - E(\mathbf{b}))(\mathbf{b} - E(\mathbf{b}))^\top)$
when $E(\mathbf{b}) = \mathbf{0}$, $\text{Var}(\mathbf{b}) = E(\mathbf{bb}^\top)$ b_i, b_j indep: $\text{Cov}(b_i,b_j) = 0$, converse false
correlation: $\rho(b_i,b_j) \equiv \frac{\text{Cov}(b_i,b_j)}{(\text{Var}(b_i)\text{Var}(b_j))^{1\backslash 2}}$
$\text{Var}(\mathbf{b})$ positive semidefinite. cov and corr matrix positive definite most of the time.
positive definite: $\mathbf{x}^\top\mathbf{Ax} > 0$ for $\mathbf{x}\in k\times 1$.
$\mathbf{x}^\top\mathbf{Ax} = \sum_i^k\sum_j^k x_i x_j A_{ij}$. If $\ge 0 \Rightarrow$ semidef.
Any $\mathbf{B}^\top\mathbf{B}$ is pos semidef. If full col rank then pos def. pos def $\Rightarrow$ diag > 0 & non-singular. (pos def)$^{-1}$ $\exists$ & is pos def.
Precision mtrx: invers of cov mtrx of estmatr. $\exists$ & pos def iff cov mtrx pos def.
If u IID w Var σ^2 and cov of any pair $= 0$: $\text{Var}(\mathbf{u}) = E(\mathbf{uu}^\top) = \sigma^2\mathbf{I}$. If false, $\mathbf{\Omega}$ = err cov mtrx. If diag of $\mathbf{\Omega}$ differ, heteroskedastic. Homoskedastic: all u same Var. Autocorrelated: off-diag $\mathbf{\Omega} \neq 0$.
$\hat\beta$ unbiased & $\mathbf{\Omega} = \sigma^2\mathbf{I}$ so no hetero or autocorr, then $\text{Var}(\hat\beta) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.
Precision affected by n, σ^2, X.
Collinearity: precision for β_1 dep on $\mathbf{X}_2$.

Efficiency

$\tilde\beta$ more efficient than $\hat\beta$ iff $\text{Var}(\tilde\beta)^{-1} - \text{Var}(\hat\beta)^{-1}$ is nonzero pos semidef mtrx.
Gauss-Markov: If $E(\mathbf{u}\mid\mathbf{X}) = \mathbf{0}$ and $E(\mathbf{uu}^\top\mid\mathbf{X}) = \sigma^2\mathbf{I}$ then OLS $\hat\beta$ is BLUE (best linear unbiased estimator). Not necessary that u normally distributed.

Residuals & Disturbances

$\hat{\mathbf{u}} = \mathbf{M_X u}$ (hat resid, u dist).
If $E(\mathbf{u}\mid\mathbf{X}) = \mathbf{0} \Rightarrow E(\|\hat{\mathbf{u}}\|^2) \le E(\|\mathbf{u}\|^2)$
$\text{Var}(\hat{u}_t) < \sigma^2$; $\hat\sigma^2 \equiv \frac{1}{n}\sum_t^n \hat{u}_t^2$
$E(\hat\sigma^2) = \frac{n-k}{n}\sigma^2$
$E(\mathbf{u}^\top\mathbf{M_X u}) = E(\text{SSR}(\hat\beta)) = (n-k)\sigma^2$
unbiased: $s^2 \equiv \frac{1}{n-k}\sum_t^n\hat{u}_t^2$; s = std err.
unbias est of $\text{Var}(\hat\beta)$: $\widehat{\text{Var}}(\hat\beta) = s^2(\mathbf{X}^\top\mathbf{X})^{-1}$
s^2 unbiased and consistent.
$\text{MSE}(\hat\beta) \equiv E((\hat\beta - \beta_0)(\hat\beta - \beta_0)^\top)$
I $\hat\beta$ unbiased $\text{MSE}(\hat\beta) = \text{Var}(\hat\beta)$.

Measures of Goodness of Fit

$R_u^2 \equiv \frac{\text{ESS}}{\text{TSS}} = \frac{\|\mathbf{P_X y}\|^2}{\|\mathbf{y}\|^2} = \cos^2\theta$, where θ angle between $\mathbf{y}$ and $\mathbf{P_X y}$. $0 \le R_u^2 \le 1$.
R_c^2: center all vars first. Invalid if $\iota \notin \mathcal{S}(\mathbf{X})$.
$R_c^2 = 1 - \sum_t^n\hat{u}_t^2 \backslash \sum_t^n(y_t - \bar{y})^2$.
Adj R^2: unbiased estimators. maybe < 0.
$\bar{R}^2 \equiv 1 - \frac{\frac{1}{n-k}\sum_t^n\hat{u}_t{}^2}{\frac{1}{n-1}\sum_t^n(y_t - \bar{y})^2} = 1 - \frac{(n-1)\mathbf{y}^\top\mathbf{M_X y}}{(n-k)\mathbf{y}^\top\mathbf{M}_\iota\mathbf{y}}$
$\bar{R}^2$ does not always $\uparrow$ in regressors.

Hypothesis Testing

If u_t normal, and σ known, test $\beta = \beta_0$ w
$z = \frac{\hat\beta - \beta_0}{(\text{Var}(\hat\beta))^{1\backslash 2}} = \frac{n^{1\backslash 2}}{\sigma}(\hat\beta - \beta_0)$, $z \sim N(0,1)$
NCP: $\lambda = \frac{n^{1\backslash 2}}{\sigma}(\beta_1 - \beta_0)$, $\beta_1 \neq \beta_0$
Reject null if z large enough. 2-tail: $|z|$.
Type 1: reject true null, 2: accept false
left-tail $\Phi(-c_\alpha) = \alpha\backslash 2$, $c_\alpha = \Phi^{-1}(\alpha\backslash 2)$.
$\Phi^{-1}(.975) = 1.96$. Power: prob test rejects the null. Prob of Type $2 = 1 - P$(power).
Power $\uparrow$ with $(\beta_1 - \beta_0)\uparrow$ or $\sigma\downarrow$ or $n\uparrow$.
$p(z) = 2(1 - \Phi(|z|))$
$x \sim N(\mu,\sigma^2) \Rightarrow z = (x-\mu)\backslash\sigma$, $z \sim N(0,1)$.
Lin comb of rand vars that are jointly multivariate normal must be $\sim N$. If $\mathbf{x}$ multivar norm with 0 cov, componenets of $\mathbf{x}$ are mutually indep.
χ^2: $y \equiv \|\mathbf{z}\|^2 = \mathbf{z}^\top\mathbf{z} = \sum_i^m z_i^2$., $y \sim \chi^2(m)$ with $\mathbf{z} \sim N(\mathbf{0},\mathbf{I})$; $E(y) = m$. $\text{Var}(y) = 2m$.
$y_1 \sim \chi^2(m_1)$ & $y_2 \sim \chi^2(m_2)$ indep $\Rightarrow y_1 + y_2 \sim \chi^2(m_1 + m_2)$
$m\times 1$ $\mathbf{x} \sim N(\mathbf{0},\mathbf{\Omega})$, then $\mathbf{x}^\top\mathbf{\Omega}^{-1}\mathbf{x} \sim \chi^2(m)$
If $\mathbf{P}$ $n\times n$ w rank $r < n$ and $\mathbf{z} \sim N(\mathbf{0},\mathbf{I})$ then $\mathbf{z}^\top\mathbf{Pz} \sim \chi^2(r)$.
$z \sim N(0,1)$ & $y \sim \chi^2(m)$, z,y indep, then $t \equiv z \backslash (y\backslash m)^{1\backslash 2}$. Or $t \sim t(m)$. Only first $m-1$ moments exist. Cauchy: $t(1)$.
$\text{Var}(t) = m\backslash(m-2)$. $t(m)$ tends to std norm.
y_1, y_2 indep rand var $\sim \chi^2(m_1)$ & $\chi^2(m_2)$, then $F \equiv \frac{y_1\backslash m_1}{y_2\backslash m_2}$. $F \sim F(m_1,m_2)$.
As $m_2 \to \infty$, $F \sim 1\backslash m_1$ times $\chi^2(m_1)$.
$t \sim t(m_2) \Rightarrow t^2 \sim F(1,m_2)$.

Exact Tests ($\mathbf{u} \sim N(\mathbf{0},\sigma^2\mathbf{I})$)

$\frac{\mathbf{x}_2^\top\mathbf{M}_1\mathbf{y}}{s(\mathbf{x}_2^\top\mathbf{M}_1\mathbf{x}_2)^{1/2}} = \left(\frac{\mathbf{y}^\top\mathbf{M_X y}}{n-k}\right)^{-\frac{1}{2}}\frac{\mathbf{x}_2^\top\mathbf{M}_1\mathbf{y}}{(\mathbf{x}_2^\top\mathbf{M}_1\mathbf{x}_2)^{1\backslash 2}}$
is t-stat $t_{\beta_2} \sim t(n-k)$ for testing $\beta_2 = 0$.
$\beta_2 \in \mathbb{R} \Rightarrow$ test for $\beta_2 = \beta_{20}$: $(\hat\beta_2 - \beta_{20})/s_2$.
$F_{\beta_2} \equiv \frac{(\text{RSSR - USSR})/r}{\text{USSR}/(n-k)} = \frac{\mathbf{y}^\top\mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{y}/r}{\mathbf{y}^\top\mathbf{M_X y}/(n-k)}$
is F-stat $\sim F(r,n-k)$, used for multiple hyp on β_2. Under null, $\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{u} \Rightarrow F_{\beta_2} = \frac{\epsilon^\top\mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\epsilon/r}{\epsilon^\top\mathbf{M_X}\epsilon/(n-k)}$, where $\epsilon \equiv \mathbf{u}/\sigma$, $\mathbf{P}_{\mathbf{M}_1\mathbf{X}_2} = \mathbf{P_X} - \mathbf{P_1}$. P-value for F is $1 - F_{r,n-k}(F_{\beta_2})$. When only 1 restriction, F and 2-tailed t test are the same. If testing all $\beta = 0$, $F = \frac{n-k}{k-1}\times\frac{R_c^2}{1-R_c^2}$. If testing $\beta_1 = \beta_2$, let $\gamma = \beta_2 - \beta_1$ then $F_\gamma = \frac{(\text{RSSR}-\text{SSR}_1-\text{SSR}_2)/k}{(\text{SSR}_1+\text{SSR}_2)/(n-2k)}$

Asymptotic Theory

EDF: $\hat{F}(x) \equiv \frac{1}{n}\sum_t^n\mathbb{I}(x_t \le x)$. FTS: $\text{plim}\hat{F}(x) = F(x)$. CLT: $z_n \equiv \frac{1}{\sqrt{n}}\sum_t^n\frac{x_t - \mu}{\sigma}$ asymptotically $\sim N(0,1)$ if x_t IID.
Uncorrelated x_t with $E(x_t) = 0 \Rightarrow$ $n^{-1/2}\sum_t^n x_t$ goes to $N(0,\lim\frac{1}{n}\sum_t^n\text{Var}(x_t))$.
If $\mathbf{u} \sim IID(\mathbf{0},\sigma^2\mathbf{I})$, $E(u_t \mid \mathbf{X}_t) = 0$, $E(u_t^2 \mid \mathbf{X_t} = \sigma^2)$, $\text{plim}\frac{1}{n}\mathbf{X}^\top\mathbf{X} = S_{\mathbf{X}^\top\mathbf{X}}$ where S finite, deterministic, pos def mtrx, then $n^{1/2}(\hat\beta - \beta_0) \xrightarrow{d} N(\mathbf{0},\sigma^2\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1})$ and $\text{plim}\, s^2(n^{-1}\mathbf{X}^\top\mathbf{X})^{-1} = \sigma^2\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}$.
An estimator for cov mtrx is consistent if $\text{plim}(n\widehat{\text{Var}}(\hat\theta)) = \mathbf{V}(\theta)$, where $\mathbf{V}(\theta)$ is limiting cov mtrx of $n^{1/2}(\hat\theta - \theta_0)$
If u IID and testing $\beta_2 = \beta_2^0$, $t_{\beta_2} = \frac{\hat\beta_2 - \beta_2^0}{\sqrt{s^2(\mathbf{X}^\top\mathbf{X})_{22}^{-1}}}$ and $t_{\beta_2} \overset{a}{\sim} N(0,1) \Rightarrow$ $t_{\beta_2} = O_p(1)$. Under null $\beta_2 = \mathbf{0}$, w predetermined regressors $rF_{\beta_2} \overset{a}{\sim} \chi^2(r)$ where $r = k_2$ is dim of β_2.
$W(\hat\beta) = (\mathbf{R}\hat\beta - r)^\top(\mathbf{R}\widehat{\text{Var}}(\hat\beta)\mathbf{R})^{-1}(\mathbf{R}\hat\beta - r)$ is Wald where cov mtrx consistent. $W(\hat\beta) \overset{a}{\sim} \chi^2(r)$ under null where r is r-vector.

Multiple Testing

FWER: $\alpha_m = 1 - (1-\alpha)^m$.
Bonferroni: $Pr(\cup_i^m P_i \le \alpha/m) \le \alpha$.
Simes: $P_{(j)} \le j\alpha/m$ for increasing P. Bonf more conservative than Simes.

Power

If $\mathbf{z} \sim N(\mu,\mathbf{I})$ then $\mathbf{z}^\top\mathbf{z} \sim$ non-central $\chi^2(m,\Lambda = \mu^\top\mu)$ If $z \sim N(\mu,\mathbf{I})$ then $\mathbf{z}^\top\mathbf{z} \sim$ non-central $\chi^2(m,\Lambda = \mu^\top\mu)$. Under null $\beta_2 = \mathbf{0} \Rightarrow \Lambda = 0$.

$t(n-k,\lambda) \sim \frac{N(\lambda,1)}{(\chi^2(n-k)/(n-k))^{1/2}}$, $\lambda^2 = \Lambda$.
Pretest estimator: $\acute\beta = \mathbb{I}(F_{\gamma=0} > c_\alpha)\hat\beta + \mathbb{I}(F_{\gamma=0} \le c_\alpha)\tilde\beta$ where c_α is critical value for F test with r and $n-k-r$ df at α. $\acute\beta = \hat\beta$ when pretest rejects and $\acute\beta = \tilde\beta$ when reject.

Confidence & Sandwich Cov Matrices

$\theta_0 \in$ confidence set iff $\tau(\mathbf{y},\theta_0) \le c_\alpha$, if θ_0 true then prob is $1-\alpha$. Asymp t-stat: $(\hat\theta - \theta)/s_\theta$. Pivot: same distribution $\forall$DGP. CI exact only if τ pivot. If no stat with known finite sample dist, use Wald with k_2 vector $\hat\theta_2$ asym normal: $(\hat\theta_2 - \theta_{20})^\top(\widehat{\text{Var}}(\hat\theta_2))^{-1}(\hat\theta_2 - \theta_{20}) \le c_\alpha$.
$\text{Var}(\hat\beta) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}$
$\hat{\mathbf{\Omega}}$: $\frac{1}{n}\sum_t^n\hat{u}_t^2 x_{ti}x_{tj}$
$\widehat{\text{Var}}(\hat\beta) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{\Omega}}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}$
HC_0: Use $\hat{u}_t^2$ in diag of $\hat{\mathbf{\Omega}}$ and 0 else.
HC_1: Use $\hat{u}_t^2$ in $\hat{\mathbf{\Omega}}$ then mult by $n/(n-k)$
HC_2: $\hat{u}_t^2/(1-h_t)$ with $h_t \equiv \mathbf{X_t}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X_t}^\top$.
HC_3: $\hat{u}_t^2/(1-h_t)$ jackknife, for big varianc.
Ignore hetero for std err of sample mean.
HAC for when u_t hetero and/or autocorr.
$\hat{\mathbf{\Sigma}} = (1/n)\mathbf{X}^\top\hat{\mathbf{\Omega}}\mathbf{X}$ (Newey-West/H. White)
$f(a+h) = f(a) + hf'(a+\lambda h)$, $h = b - a$, $0 < \lambda < 1$.

Bootstrap

$\hat{p}^*(\hat\tau) = \frac{1}{B}\sum_b^B\mathbb{I}(\tau_b^* > \hat\tau)$
$\hat{p}_{et}^*(\hat\tau) = 2\min\left(\frac{1}{B}\sum_b^B\mathbb{I}(\tau_b^* \le \hat\tau), \frac{1}{B}\sum_b^B\mathbb{I}(\tau_b^* >\right.$ when biased param est and 2-tailed.
$\hat{p}^*(\hat\tau) \to p(\hat\tau)$ as $B \to \infty$. τ pivotal.

Instrumental Variables

Can define $E(u_t \mid \Omega_t) = 0$.
Err in variables: indep vars in regr model measured with err. $E(u_t \mid x_t) \neq 0$, $\text{Cov}(x_t,u_t) \neq 0$. OLS est biased and inconsist. Simultaneity: two or more endog vars jointly determined by sys of simultaneous eq.
Assume, $E(\mathbf{uu}^\top) = \sigma^2\mathbf{I}$ and at least one in $\mathbf{X}$ not predetermined wrt disturb. $n\times k$ mtrx $\mathbf{W}$ with $\mathbf{W_t} \in \mathbf{\Omega_t}$. Col of $\mathbf{W}$ are IV.
$E(u_t \mid \mathbf{W}_t) = 0$, $\mathbf{W}^\top(\mathbf{y} - \mathbf{X}\beta) = \mathbf{0}$ are unbiased est eq. $\hat\beta_{IV} \equiv (\mathbf{W}^\top\mathbf{X})^{-1}\mathbf{W}^\top\mathbf{y}$.
$\mathbf{W}^\top\mathbf{X}$ must be non-sing. $\hat\beta_{IV}$ generally biased but consistent. Assume $\mathbf{S}_{\mathbf{W}^\top\mathbf{X}} \equiv \text{plim}\frac{1}{n}\mathbf{W}^\top\mathbf{X}$ is deterministic and non-sing. Same for $\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}$. $\hat\beta_{IV}$ consistent iff $\text{plim}\frac{1}{n}\mathbf{W}^\top\mathbf{u} = 0$. Asym cov mtrx of IV est: $\sigma_0^2\text{plim}(n^{-1}\mathbf{X}^\top\mathbf{P_W X})^{-1}$. $\mathbf{J}$: full col rank, asym deterministic, min asym cov mtrx of IV est.
GIVE: $\hat\beta_{IV} = (\mathbf{X}^\top\mathbf{P_W X})^{-1}\mathbf{X}^\top\mathbf{P_W y}$ IV est:
Col of $P_W X$ should be lin indep.

IV asym normal like all est.
$\widehat{\text{Var}}(\hat\beta_{IV})=\hat\sigma^2(\mathbf{X}^\top\mathbf{P_W X})^{-1}$.
$\hat\sigma^2=\|\mathbf{y}-\mathbf{X}\hat{\beta_{IV}}\|^2/n$.

Generalized Least Squares

Consider $E(\mathbf{uu}^\top)=\mathbf{\Omega}, \mathbf{\Omega}^{-1}=\mathbf{\Psi\Psi}^\top$
$\hat\beta_{GLS}=(\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{y}$
$E(\mathbf{\Psi}^\top\mathbf{uu}^\top\mathbf{\Psi})=\mathbf{I}$.
$\text{Var}(\hat\beta_{GLS}^\top)=(\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X})^{-1}$ autocovariance of AR(1): $\mathbf{\Omega}(p)=\frac{\sigma_\epsilon^2}{1-\rho^2}\times$ mtrx with 1 diag and ρ^i increasing away from diag.
$\text{Cov}(u_t,u_{t-1})=\rho\sigma_u^2$.
$\sigma_u^2=\equiv\sigma_\epsilon^2/(1-\rho^2)$
$u_t=\rho u_{t-1}+\epsilon_t$, $\epsilon_t\sim IID(0,\sigma_\epsilon^2)$

3 Questions

Q:$X\sim N(0,1)$, $Z\equiv\mu+\sigma X$ mean and var?
A:Mean: μ, Var: σ^2.
Q:If X_1, X_2 indep, show $E(X_1\mid X_2)=0$. A:$f(x_1|x_2)=f(x_1,x_2)f(x_2)=f(x_1)f(x_2)f(x_2)=f(x_1)$.
Q:Prove $\mathbf{P_X M_X}=0$.
A:$\mathbf{P_X+M_X=I\Rightarrow P_X+P_XM_X=P_X\Rightarrow P_XM_X=0}$
Q:$\mathcal{S}(W)\neq\mathcal{S}(X)$, show $P\equiv\mathbf{X}(\mathbf{W}^\top\mathbf{X})^{-1}\mathbf{W}^\top$ idempotent but not sym. A:idem easy. Find $P^\top$. P projects onto $S(X)$. $Py=Xb$. Image of P is all $S(X)$ but image of $P^\top$ is $S(W)$. $I-P$ projects to $S^\perp(W)$. $P_W(I-P)=0$.
Q:$y=\beta_1\iota+X_2beta_2+u$ show using FWL that β_i can be written as $\ldots with M_\iota$. A:OLS resid orthog to regressors $\Rightarrow X_2^\top(M_\iota y-M_\iota X_2\beta_2)=0$. OLS of β_1 must be resid have mean 0, $\iota^\top(y-\beta_1\iota-X_2\beta_2)=0$.
Q:Show that $P_X-P_1=P_{M_1X_2}$. A:vector $\gamma\in s(M_1,X_2)$ is $M_1X_2\gamma$. Pre-multiply by P_x-P_1 and get $M_1X_2\gamma$. So invariant under projection. Now consider z orthog to $S(M_1,X_2)$, $X_2^\top M_1z=0$. Show $P_xz=P_1z$.
Q:What about $M_\iota X$ and $P_\iota X$ when $n\times3$? A:First col of $M_\iota X$ is 0. Other two are centered X. Each col of $P_\iota X$ has n copies of mean from col in X. First col is ι, 2nd: $\overline{x_2}\iota$. Know $P_\iota P_X=P_\iota$ show that $M_\iota M_X=M_X$. Finally show $P_\iota M_X=0$.
Q:Show that leverage h_t is square of cos of angle bt e_t and its proj. A:proj: P_Xe_t. $\cos^2\theta$: $e_t^\top P_Xe_t$.
Q:What is Tr(P_X)? A:Tr(P_X) $=k$ if full rank or r if less. Tr(M_X) $=n-r$.
Q:Show $(\text{Cov}(b_1,b_2))^2\le\text{Var}(b_1)\text{Var}(b_2)$. A:determinant of pos semi def matrix is nonnegative. Find det of $\text{Var}(b)=Var(b_1)Var(b_2)-(Cov(b_1,b_2))^2$.

Q:If A pos def matrix, show that A^{-1} is pos def. A:non-zero x: $x^\top A^{-1}x=(A^{-1}x)^\top A(A^{-1}x)$. quad form must be pos, and so is this.
Q:If A pos def sym the $I-A$ is pos def iff $A^{-1}-I$ is pos def. A:If $A=R^2$, $R^{-1}R^{-1}=A^{-1}$ and $R^{-1}AR^{-1}=I$. $x^\top(I-A)x=z^{top}(A^{-1}-I)z$ so left pos def iff right pos def. If $A-B$ pos def then so is $R^{-1}(A_B)R^{-1}$. Get $I-R^{-1}BR^{-1}$ pos def iff $RB^{-1}R-I$ pos def from part a). Pre and post mult by R^{-1}.
Q:Prove $E(\hat u^\top\hat u)=(n-k)\sigma_0^2$. A:$\hat u^\top\hat u=u^\top M_Xu$ which is a scalar so equal to its own trace. $E(\hat u^\top\hat u)=E(Tr(u^\top M_Xu))=E(Tr(u^\top uM_X))=\sigma_0^2Tr(M_X)$
Q:$y=X\beta+u$ and $y=X\beta+Z\gamma+u$ with $X^\top Z=0$. Show β identical. A:FWL to get estimates. Since $X^\top Z=0\Rightarrow P_ZX=0$ and $M_ZX=X$ is all you need.
Q:error $\sim IID(0,\sigma^2)$, cov of forecast errors? A:$y_*-X_*\hat\beta=u_*-X_*(X^\top X)^{-1}X^\top u$. Use def to get cov matrix.
Q:Run regression with restriction that $\beta_2+\beta_3=1$. A:sub $\beta_3=1-\beta_2$ into regression. Define $y_t'=y_t-x_{t3}$ and $z_t=x_{t2}-x_{t3}$.
Q:Suppose new regression with $y_t'=y_t+10$ A:constant goes 10 higher. R_c^2 and $\bar R^2$ unchanged becuase TSS around mean nor SSR change. But R_u^2 changes because TSS around 0 greater.
Q:If $z\sim N(0,1)$ and z test stat 2-tailed, $p(z)=2(1-\Phi(|z|))$. Show F_p, CDF of $p(z)$ is $F_p(x)=x$. For $x\ in[0,1]$. A:Pr($p(z)\le x$) $=Pr(|z|\ge\Phi^{-1}(1-x/2))$. Because $z\sim N(0,1)$, $\Phi(-\Phi^{-1}(1-x/2))+1-\Phi(\Phi^{-1}(1-x/2))$. Simplifies to x. Use symmetry of normal.
Q:$y_t=\beta_1+\beta_2x_{t1}+\beta_3x_{t2}+u_t$. Rerwite so $\beta_2-\beta_3=1$ is zero restriction. A:Sub and rearrrange so LHS is y_t+x_{t2} and add γx_{t2}. Equiv because spans of regressors are same. Becomres restricted model when $\gamma=0$.
Q:model with r restrictions $R\beta=r$. Rewrite so r zero restrictions. A:$R_1\beta_1+R_2\beta_2=r$. Solve for β_2. Sub in. Subtract $X_2R_2^{-1}r$ from y. $Z_1\equiv X_1-X_2R_2^{-1}R_1$ spans same as X. $y-X_2R_2^{-1}r=Z_1\gamma_1+X_2\gamma_2+u$ is equiv with $\beta_1=\gamma_1$. $\gamma_2=0\Rightarrow R\beta=r$.
Q:When performing Chow test A:Since n2 < k, we cannot run the regression for the second subsample. If we did, we would get SSR2 = 0. Thus, in this special case, SSR1 +SSR2 = SSR1. Therefore, RSSRSSR1 SSR2 in () must be replaced by RSSRSSR1, and SSR1 + SSR2 must be replaced by SSR1, as in the formula above. For the denominator, easy to see that the df is n1 k, since SSR1 is the SSR from a regression with n1 observations and k parameters. For the numerator, the degrees of freedom for RSSR is n k, and the degrees of freedom for SSR1 is n1 k. The difference between these is $nk(n1k)=nn1=n2$.
Q:Show r times F stat is $\chi^2(r)$ asym. A:denom of F dist as $\chi^2(n-k)$ under normality. Has mean $(n-k)$ even without normality assmpt. LLN $\Rightarrow$ denom tends to 1 asym. Let $v=n^{-1/2}X^\top\epsilon$. CLT: $v\sim N(0,S_{X^\top X})$.
Q:How is $y^\top M_Xy/\sigma_0^2$ dist? Test $\sigma=\sigma_0$. A:When specified correct: $y^\top M_Xy/\sigma_0^2=\epsilon^\top M_X\epsilon$ which $\sim\chi^2(n-k)$. Two-tailed.
Q:stat τ with CDF F where $F(-x)\neq1-F(x)$. marginal sig level satisfies one of: $F(\hat\tau)=\alpha/2$ or $1-(\alpha/2)$ Isolate α, consider $F(\tau)=1/2$.
Q:If F strict increase on $[a,b]$, show if $X\sim U(0,1)$, then $F^{-1}(X)$ is drawing from dist of which F is CDF. A:Since F increase: $Pr(F^{-1}(X)\le x)=Pr(X\le F(x))$ which is just $F(X)$ because uniform X.
Q:Var($\hat u_t$) $=(1-h_t)\sigma_0^2$. Derive alt to rescaling residuals. A:$\acute u_t=\hat u/(1-h_t)^{1/2}$. Each $\acute u_t$ has same variance σ_0^2, but not mean 0. Define $\mu_1=\frac{1}{n}\sum\acute u_t$ and $\mu_2=\frac{1}{n}\sum\acute u_t{}^2$. Answer: $\frac{\tilde s}{(\acute\mu_2-\acute\mu_1^2)^{1/2}}(\acute u-\acute\mu_1)$
Q:If $z\sim N(0,1)$ under null and $N(\lambda,1)$ under alt, show power. A:Power is $Pr(z<-c_\alpha)+Pr(z>-c_\alpha)$. $Pr(z<-c_\alpha)=\Phi(-c_\alpha-\lambda)$ and $Pr(z>-c_\alpha)=\Phi(-c_\alpha+\lambda)$
Q:$z\sim N(\mu,I)$, exp of $z^\top z$ is $m+\mu^\top\mu$ A:Write $z=\mu+x$ with $x\sim N(0,I)$. $E(z^\top z)=\mu^\top\mu+m$ because second term dist chi^2 with exp m.
Q:Using square of t-stat obtain .99 CI for β_2. A:Use simple t. $Pr(t<F_{.99}(1,n-k))=.99$. Take pos square root of both side. Write out both ineq so get rid of abs value. Sub $beta_2$ and mult -1.
Q:Explain how to construct symm bootstrap CI based on asym $t=(\hat\theta-\theta_0/s_0)$. A:Let t_j^* denote the jth bootstrap t statistic. Then, instead of sorting the t_j^* themselves, we take their absolute values and sort the them from largest to smallest. If we do this, the level α critical value can be estimated as the entry numbered (B + 1) in the sorted list. For example, if B = 999, it is entry number 50. Let us denote this estimate by c_α^*. Then the symmetric bootstrap CI is $[\hat\theta-s_\theta c_\alpha^*,\hat\theta+s_\theta c_\alpha^*]$
Q:Suppose SSR from OLS is 106.44. Under classic normal, construct .95 CI for σ^2. A:Under classic normal $y^\top M_Xy/sigma^2\sim\chi^2(n-k)$. $n-k=94$. Solve $106.44=0.025$ quant of chiσ^2, (2 equations).
Q:est θ by Least Squares. Gen 999 bootstrap and calc t for $\theta=\hat\theta$. Find .95 studentized bootstrap CI. A:$[\hat\theta-s_\theta c^*_{975},\hat\theta-s_\theta c^*_{025}]$
Q:Show that F stat that $\beta_2=\beta_{20}$ can be written as A:The denominator of (5.26) is evidently the denominator of both F statistics, since it is just the usual OLS estimator s^2 for both of the unrestricted models, (5.24) and (5.25). It is not quite so obvious that the numerator of (5.26) is the numerator of the F statistic. The usual way to construct the numerator of an F statistic is to subtract the unrestricted SSR from the restricted SSR and divide by the number of restrictions. For the model (5.25), k2 times this numerator is
Q:x_1, x_2 centered. $\hat\rho$ is sample corr. Show corr of $\hat\beta$ is $-\rho$. A:Use FWL to find $\text{Var}(\hat\beta_1)=\sigma^2(x_1^\top M_2x_1)^{-1}$ and $\text{Var}(\hat\beta_2)=\sigma^2(x_2^\top M_1x_2)^{-1}$. Rewrite $x_1^\top M_2x_1=(1-\hat\rho^2)x_1^\top x_1$. Use FWL to find Cov for β_i too. Assmpt that error are normal can be relaxed but $E(uu^\top)=\sigma^2I$ cannot.
Q:Consider .05 conf region for β_1 and β_2. Show set is circular disk, what is radius. A:exact conf region is $(\hat\beta-\beta_0)^\top X^\top X(\hat\beta-\beta_0)\le2c_{.05}s^2$. Where $c_{.05}$ is critical value for $F(2,n-2)$. Radius: $(2c_{.05}s^2)^{-1/2}$
Q:First 3 rows of matrix, left side is ι, right is $\in\mathbb{R}$. Repreated. A:Use FWL for estimate of β_2
Q:Show that minimizing the criterion function with respect to β yields generalized IV est. A:Criterion func is $Q(\beta,y)=(y-X\beta)^\top P_W(y-X\beta)$. Differentiate wrt β and set to 0.
Q:Show that plim of $\frac{1}{n}Q(\beta_0,y)$ is zero. A:plim $\frac{1}{n}Q(\beta_0,y)$ = plim $\frac{1}{n}u^\top P_Wu$. Split P_W up into three plims with W. Assumption when plim$n^{-1}W^\top u=0$ and plim$n^{-1}W^\top W$ is nonsign matrix. Get $0S_{W^\top W}^{-1}0=0$.
Q:assume asym ident condition $S_{X^\top W}(S_{W^\top W})^{-1}S_{W^\top X}$ has full rank. Show GIVE is consistent. A:Expand $\hat\beta_{IV}$ to $\beta_0+\ldots.Multiplyby$n^{-1}. plim $\hat\beta_{IV}=\beta_0+\ldots$ which goes to 0 by assumption on $n^{-1}W^\top u$ so plim $\hat\beta_{IV}=\beta_0$.
Q:Apply CLT to $n^{-1/2}W^\top u$ and get asym multivar normal with mean 0. Show $n^{1/2}(\beta_{IV}-\beta_0)$ is asym normal with mean 0. A:Use plim$n^{-1}W^\top X=S_{W^\top X}$. Show that plim$(n^{-1}X^\top P_WX)^{-1}=((S_{X^\top W})(S_{W^\top W})^{-1}S_{W^\top X})^{-1}$ Second factor: plim$(n^{-1/2}X^\top P_Wu)=S_{X^\top W}(S_{W^\top W})^{-1}$plim $(n^{-1/2}W^\top u)$. Under assmpt last term is asym normal with mean 0. So our thing is too. Asym cov matrix of $n^{-1/2}W^\top u$ is $\sigma_0^2S_{W^\top W}$.
Q:Suppose W_1 and W_2 matrices of instruments with W_2 extra col. Prove generalized IV W_2 more eff than W_1. A:$X^\top(P_{W_2}-P_{W1})X$. Since $S(W_1)\subset S(W_2)$ $P_{W_1}P_{W_2}=P_{W1}=P_{W_2}P_{W_1}$. Rewrite. Since $I-P_{W_1}$ is orthog proj matrx, it is pos semi def. Divide by n and tend to ∞.
Q:Show that simple IV unbiased when data gen by $\sigma_v=0$. A:$\hat{\beta_{IV}}=\beta_0+\sigma_u(w^\top x)^{-1}w^\top u$. If $\sigma_u=0$, then $x=w\pi_0$. Second term on right after subbing in is σ_u/π_0 times coeff est from OLS of u on w. w exog so $E((w^\top w)^{-1}w^\top u)=0$. Conclude that $E(\hat\beta_{IV})=\beta_0$. x is scalar mult of w so uncorr with u. OLS β unbiased so IV unbiased.
Q:Verify that $\hat{\beta_{IV}}=\hat\beta$ OLS when $X=[ZY]=W\Pi$. Is consistent? A:plug and chug. Yes consistent. plim of est is $\beta_0+\text{plim}(\frac{1}{n}\Pi^\top W^\top W\Pi)^{-1}\text{plim}(\frac{1}{n}\Pi^\top W^\top u)$ 1st factor in 2nd term is full rank deterministic, second factor 0. Second term is 0.
Q:Verify using assumpt that instrument in W are exog and pred and LLN. A:Need to apply LLN to $n^{-1}W^\top V$. Prob limits of n^{-1} times each term is 0. Apply CLT to second factor, asym $\sim N(0,cov)$.
Q:Prove $\frac{1}{\sigma^2}u^\top(P_{P_WX}-P_{P_WX_1})u\sim\chi^2(k_2)$ A:$P_{P_WX}-P_{P_WX_1}$ is diff btw two orthog proj matrices. So it is orthog proj. It is subspace of $S(W)$. Can be put as $Z(Z^\top Z)^{-1}Z^\top$ where Z is k_2 dimensional so proj goes to $S(Z)$. Sub it in. All cols of Z asym uncorr with u. CLT to $n^{-1/2}Z^\top u\sim a\sim N(0,\sigma_0^2Z^\top Z)$. So we have a norm sandwich and must be $\sim\chi^2(k_2)$.
Q:P proj and z not norm. Show that zPz follows $\chi^2(r)$ asym A:Expand P as Z. add $n^{-1/2}$ and n as mult by 1. Apply CLT to both sides, get normal w mean 0 and cov matrix plim $\frac{1}{n}Z^\top Z$.