ECON 468 Cheat Sheet
Tiffany Yong Page:1

1 Basics

Distributions

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

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

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:
$\mu_k \equiv E(X-E(X)^k) = \sum_{i=1}^m p(x_i)(x_i-\mu)^k$

Multivariate

Joint Density Func: $f(x_1,x_2) = \frac{\partial^2 F(x_1,x_2)}{\partial x_1 \partial x_2}$

Indep: $F(x_1,x_2)=F(x_1,\infty)F(\infty,x_2)$ 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

Symmetric: $\mathbf{A}=\mathbf{A}^\top$

Dot Product: $\mathbf{a}^\top\mathbf{b} = \sum_{i=1}^n a_i b_i$

Matrix Mult: $C_{ij} = \sum_{k=1}^m A_{ik}B_{ki}$

Invertible: $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$

$\|\mathbf{x}\| = (\mathbf{x}^\top\mathbf{y})^{1/2} = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$

$\mathbf{X}^\top\mathbf{X}_{ij} = \sum_{t=1}^n x_{ti}x_{tj}$

$\langle \mathbf{x},\mathbf{y}\rangle = \|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta)$

$\sin(x\pm y) = \sin x\cos y \pm \cos x \sin y$

$\cos(x\pm y) = \cos x\cos y \mp \sin x\sin y$

$\sin 2\theta = 2\sin\theta\cos\theta$

$\cos 2\theta = 1-2\sin^2\theta = 2\cos^2\theta - 1$

Cauchy-Schwartz: $\langle\mathbf{x},\mathbf{y}\rangle \le \|\mathbf{x}\|\|\mathbf{y}\|$

Linear Indep: $\nexists \mathbf{b}\neq\mathbf{0}$ s.t. $\mathbf{Xb}=\mathbf{0}$

Singular: $\exists \mathbf{x}\neq 0: \mathbf{Ax}=\mathbf{0}$

$Tr(ABC)=Tr(CAB)=Tr(BCA)$

$Tr(P_X) = \text{rank}(X)$

2 Linear Regression

OLS

Information set: Ω_t, for conditioning

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: $\mathbf{P_X} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$

$\mathbf{M_X} = \mathbf{I}-\mathbf{P_X} = \mathbf{I}-\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$

$\mathbf{P_X y} = \mathbf{X}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}) = \mathbf{X}\hat\beta$; $\mathbf{P_X P_X} = \mathbf{P_X}$

$\mathbf{M_X y} = \hat{\mathbf{u}}$; $\mathbf{M_X X} = \mathbf{0}$; $\mathbf{M_X M_X} = \mathbf{M_X}$

$\mathbf{P_X}+\mathbf{M_X} = \mathbf{I}$; $\mathbf{P_X M_X} = \mathbf{0}$; $\mathbf{P_X}^\top = \mathbf{P_X}$

$\|\mathbf{y}\|^2 = \|\mathbf{P_X y}\|^2 + \|\mathbf{M_X y}\|^2$; $\|\mathbf{P_X y}\| \le \|\mathbf{y}\|$

$\mathbf{A}$, $k\times k$ nonsingular, $\mathbf{P_{XA}} = \mathbf{P_X}$

$\mathbf{X}_1,\mathbf{X}_2$ orthogonal, $\mathbf{X}_1^\top\mathbf{X}_2 = \mathbf{O}$

Centering: $\mathbf{M}_\iota\mathbf{x} = \mathbf{z} = \mathbf{x} - \bar{x}\iota$; $\iota^\top\mathbf{z} = 0$

$\mathbf{P_1} \equiv \mathbf{P_{X_1}}$; $\mathbf{P_1 P_X} = \mathbf{P_X P_1} = \mathbf{P_1}$

Res in regression w/ constant or dummy variables sum to 0.

FWL: β_2 from $\mathbf{y} = \mathbf{X_1}\beta_1 + \mathbf{X_2}\beta_2 + \mathbf{u}$ and $\mathbf{M_1 y} = \mathbf{M_1 X_2}\beta_2 + \text{res}$ are the same. (+ res)

FWL β: $(\mathbf{X}_2^\top\mathbf{M_1 X_2})^{-1}\mathbf{X}_2^\top\mathbf{M_1 y}$

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. $\mathbf{D}$ has G dummy vars for fixed effects.

$\mathbf{M_D x} = \mathbf{x} - [\iota_{n_1}\bar{x}_1 \dots \iota_{n_G}\bar{x}_G]$

$\hat\eta = [\bar{y}_1 - \bar{\mathbf{X}}_1\hat\beta \dots \bar{y}_G - \bar{\mathbf{X}}_G\hat\beta]$

$h_t = \mathbf{e_t}^\top\mathbf{P_X e_t} = \|\mathbf{P_X e_t}\|^2$

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

Bias

vector of (true) model params: θ

Bias: $E(\hat\theta)-\theta_0$, $E((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}) = 0$

$E_\mu(\hat\beta) = \beta_\mu + E(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u})$

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, finite $E(Y_n)^2$ LLN2: $\text{plim}\frac{1}{n}\sum_t^n Y_t = \lim\frac{1}{n}\sum_t^n E(Y_t)$

$\text{plim}\eta(Y_n) = \text{plim}\eta(Y_\infty)$ if converg

$\text{plim}\, Y_n Z_n = \text{plim}\, Y_n \text{plim} Z_n$ if converg

$\mathbf{X}^\top\mathbf{X}$ may not have plim so mult by $1\setminus n$.

Then, $\text{plim}1\setminus n\mathbf{X}^\top\mathbf{X} = \mathbf{S_{X^\top X}}$

same order: $\forall\epsilon > 0\exists K,N : P(|a_n\setminus n^r| > K) < \epsilon \forall n > N \Rightarrow f(n) = O_p(n^r)$

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: $\rho(b_i,b_j) \equiv \frac{\text{Cov}(b_i,b_j)}{(\text{Var}(b_i)\text{Var}(b_j))^{1\setminus 2}}$

$\text{Var}(\mathbf{b})$ 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.

$\mathbf{A}$ pos semidef $\Rightarrow \det(A)\ge 0$

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.

If $k\times k$ $\mathbf{A}$, symmetric, pos def, $\exists k\times k$ $\mathbf{B}: \mathbf{A} = \mathbf{B}^\top\mathbf{B}$, $\mathbf{B}$ not unique

Precision mtrx: invers of cov mtrx of estmatr. $\exists$ & pos def iff cov mtrx pos def.

If u w/ Var σ^2 and cov of any pair $=0$: $\text{Var}(\mathbf{u}) = E(\mathbf{uu}^\top) = \sigma^2\mathbf{I}$ — white noise. If false, $\mathbf{\Omega}$ = err cov mtrx. If diag of $\mathbf{\Omega}$ differ, heteroskedastic. Homoskedastic: all μ same Var. Autocorrelated: off-diag $\mathbf{\Omega}\neq 0$.

$\text{Var}(\hat\beta) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega X}(\mathbf{X}^\top\mathbf{X})^{-1}$

$\hat\beta$ unbiased & $\mathbf{\Omega} = \sigma^2\mathbf{I}$ so no hetero or autocorr, then $\text{Var}(\hat\beta) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.

FWL Variance: $\text{Var}(\hat\beta_1) = \sigma_0^2(\mathbf{x}_1^\top\mathbf{M_2 x_1})^{-1}$

Precision affected by n,σ^2, X.

Collinearity: precision for β_1 dep on $\mathbf{X_2}$.

Efficiency

$\tilde\beta$ more effic than $\hat\beta$ iff $\text{Var}(\tilde\beta)^{-1} - \text{Var}(\hat\beta)^{-1}$ nonzero pos semidef mtrx $\Rightarrow \text{Var}(\hat\beta) - \text{Var}(\tilde\beta)$ nonzero pos semidef mtrx

Lin. estmtr: $\tilde\beta = \mathbf{Ay} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} + \mathbf{Cy}$

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.

Unnecessary for $\mu\sim N$.

Residuals & Disturbances

$\hat\mu = \mathbf{M_X u}$ (hat resid, μ dist).

If $E(\mathbf{u}\mid\mathbf{X}) = \mathbf{0} \Rightarrow E(\hat{u}_t|\mathbf{X}) = 0 \Rightarrow E(\|\hat{\mathbf{u}}\|^2|\mathbf{X}) \le E(\|\mathbf{u}\|^2|\mathbf{X})$

$\text{Var}(\hat{u}_t) = (1-h_t)\sigma^2 < \sigma^2$; $\hat\sigma^2 \equiv \frac{1}{n}\sum_t^n\hat{u}_t^2$

$E(\hat\sigma^2) = \frac{n-k}{n}\sigma^2$

$E(\mathbf{u}^\top\mathbf{M_X u}) = E(\text{SSR}(\hat\beta)) = (n-k)\sigma^2$

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.

$\text{MSE}(\tilde\beta) \equiv E((\tilde\beta - \beta_0)(\tilde\beta-\beta_0)^\top)$

If $\tilde\beta$ unbiased $\text{MSE}(\tilde\beta) = \text{Var}(\tilde\beta)$.

Measures of Goodness of Fit

TSS = ESS + SSR

$R_u^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\|\mathbf{P_X y}\|^2}{\|\mathbf{y}\|^2} = \cos^2\theta$, 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 $\iota\notin\mathcal{S}(\mathbf{X})$.

$R_c^2 = 1 - \sum_t^n\hat{u}_t^2 \setminus \sum_t^n(y_t-\bar{y})^2$.

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

$z = \frac{\hat\beta - \beta_0}{(\text{Var}(\hat\beta))^{1\setminus 2}} = \frac{n^{1\setminus 2}}{\sigma}(\hat\beta - \beta_0)$, $z\sim N(0,1)$

NCP: $\lambda = \frac{n^{1\setminus 2}}{\sigma}(\beta_1 - \beta_0)$, $\beta_1 \neq \beta_0$

Reject null if z large enough. 2-tail: $|z|$. Type 1: reject true null, 2: accept false

left-tail $\Phi(-c_\alpha) = \alpha\setminus 2$, $c_\alpha = \Phi^{-1}(\alpha\setminus 2)$.

$\Phi^{-1}(.975) = 1.96$, $\Phi^{-1}(.95) = 1.645$.

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 $n\uparrow$.

$p(z) = 2(1-\Phi(|z|))$

$x\sim N(\mu,\sigma^2) \Rightarrow z = (x-\mu)\setminus\sigma$, $z\sim N(0,1)$.

Density of $\text{N}(\mu,\sigma^2) = 1\setminus\sigma\phi(x-\mu\setminus\sigma)$

Lin comb of rand vars that are jointly multivariate normal must be $\sim N$. If $\mathbf{x}$ multivar norm where $\mathbf{x} = \mathbf{Az}$, $\mathbf{x}\sim\text{N}(\mu,\mathbf{\Omega})$, $\mathbf{\Omega} = \text{Var}(\mathbf{x}) = \mathbf{AA}^\top$, $\mathbf{A}$ lower-triangular. If $\mathbf{x}$ multivar norm vector w 0 covs, $\mathbf{x}$ components 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 $\Rightarrow y_1+y_2\sim\chi^2(m_1+m_2)$

$\chi^2(m)\xrightarrow{d}\text{N}(m,2m)$

$m\times 1$ $\mathbf{x}\sim N(\mathbf{0},\mathbf{\Omega})$, then $\mathbf{x}^\top\mathbf{\Omega^{-1}x}\sim\chi^2(m)$

If $\mathbf{P}$ $n\times n$ orthogonal projection 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 $t\equiv z\setminus(y\setminus m)^{1\setminus 2}$. Or $t\sim t(m)$. Only first $m-1$ moments exist. Cauchy: $t(1)$. $\text{Var}(t) = m\setminus(m-2)$. $t(m)$ converges in distribution to std norm

y_1,y_2 indep rand var $\sim\chi^2(m_1)$ & $\chi^2(m_2)$, then $F\equiv\frac{y_1\setminus m_1}{y_2\setminus m_2}$. $F\sim F(m_1,m_2)$. As $m_2\to\infty$, $F\sim 1\setminus m_1$ times $\chi^2(m_1)$. $t\sim t(m_2)\Rightarrow t^2\sim F(1,m_2)$.

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

$$\frac{\mathbf{x}_2^\top\mathbf{M_1 y}}{s(\mathbf{x}_2^\top\mathbf{M_1 x_2})^{1/2}} = \left(\frac{\mathbf{y}^\top\mathbf{M_X y}}{n-k}\right)^{-\frac{1}{2}}\frac{\mathbf{x}_2^\top\mathbf{M_1 y}}{(\mathbf{x}_2^\top\mathbf{M_1 x_2})^{1\setminus 2}}$$

is t-stat $t_{\beta_2}\sim t(n-k)$ for testing $\beta_2 = 0$.

$\beta_2\in\mathbb{R}\Rightarrow$ test for $\beta_2 = \beta_{20}$: $(\hat\beta_2 - \beta_{20})/s_2$.

USSR $= \mathbf{y}^\top\mathbf{M_1 y} - \mathbf{y}^\top\mathbf{M_1 P_{M_1X_2} M_1 y}$

RSSR $= \mathbf{y}^\top\mathbf{M_1 y}$

$F_{\beta_2} \equiv \frac{\text{(RSSR - USSR)}/r}{\text{USSR}/(n-k)} = \frac{\mathbf{y}^\top\mathbf{P_{M_1X_2}y}/r}{\mathbf{y}^\top\mathbf{M_X y}/(n-k)}$ is F-stat $\sim F(r,n-k)$, used for multiple hyp on β_2. Under null, $\mathbf{M_1 y} = \mathbf{M_1 u}\Rightarrow F_{\beta_2} = \frac{\epsilon^\top\mathbf{P_{M_1X_2}}\epsilon/r}{\epsilon^\top\mathbf{M_X}\epsilon/(n-k)}$, where $\epsilon\equiv\mathbf{u}/\sigma$, $\mathbf{P_{M_1X_2}} = \mathbf{P_X} - \mathbf{P_1}$.

$\mathbf{y} = \mathbf{P_1 y} + \mathbf{P_{M_1X_2}y} + \mathbf{M_X y}$

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 $\beta = 0$, $F = \text{ESS}\setminus(\text{TSS - ESS}) = \frac{n-k}{k-1}\times\frac{R_c^2}{1-R_c^2}$. 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: $\hat{F}(x) \equiv \frac{1}{n}\sum_t^n\mathbb{I}(x_t\le x)$. Fund Thm Stats: $\text{plim}\hat{F}(x) = F(x)$. CLT: $z_n\equiv\frac{1}{\sqrt{n}}\sum_t^n\frac{x_t-\mu}{\sigma}\xrightarrow{d}N(0,1)$ if x_t IID.

Uncorrelated x_t with $E(x_t) = 0\Rightarrow n^{-1/2}\sum_t^n x_t$ goes to $N(0,\lim\frac{1}{n}\sum_t^n\text{Var}(x_t))$.

Same for multiv norm.

If $\mathbf{u}\sim IID(\mathbf{0},\sigma^2\mathbf{I})$, $E(u_t\mid\mathbf{X_t}) = 0$, $E(u_t^2\mid\mathbf{X_t}) = \sigma^2$ (innovations), $\text{plim}\frac{1}{n}\mathbf{X}^\top\mathbf{X} = S_{X^\top X}$ where S finite, deterministic, pos def mtrx, then $n^{1/2}(\hat\beta - \beta_0)\xrightarrow{d}N(\mathbf{0},\sigma^2\mathbf{S}_{\mathbf{X^\top X}}^{-1})$ and $\text{plim}\, s^2(n^{-1}\mathbf{X}^\top\mathbf{X})^{-1} = \sigma^2\mathbf{S}_{\mathbf{X^\top X}}^{-1}$.

$\hat\beta$ is root-n consistent, since $O_p(n^{1\setminus 2})$.

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 & $F_{\beta_2}\overset{a}{\sim}F(r,n-k)$. Wald: $\mathbf{R}\beta = \mathbf{r}$, r is vec of lin. restricts. $W(\hat\beta) = (n^{1\setminus 2}\mathbf{R}\hat\beta - r)^\top(\mathbf{R}n\widehat{\text{Var}}(\hat\beta)\mathbf{R}^\top)^{-1}(n^{1\setminus 2}\mathbf{R}\hat\beta - r)$ 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: $\alpha_m = 1-(1-\alpha)^m$ (reject if any test rejects).

Bonferroni: $Pr(\cup_i^m P_i\le\alpha/m)\le\alpha$ (reject if $P_i\le\alpha/m$).

Simes: Reject if any $P_{(j)}\le j\alpha/m$ 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 $\mathbf{z}\sim N(\mu,\mathbf{I})$ then $\mathbf{z}^\top\mathbf{z}\sim$ non-central $\chi^2(m,\Lambda = \mu^\top\mu)$. If $\mathbf{x}\sim\text{N}(\mu,\mathbf{\Omega})$, $\mathbf{x}^\top\mathbf{\Omega x}\sim\chi^2(m,\mu^\top\Omega\mu)$.

$\Lambda = (1\setminus\sigma^2)\beta_2^\top\mathbf{X}_2^\top\mathbf{M_1 X_2}\beta_2$. Under null $\beta_2 = \mathbf{0}\Rightarrow\Lambda = 0$.

Power of χ^2 or F test in r $\downarrow$, in $\Lambda\uparrow$.

$t(n-k,\lambda)\sim\frac{N(\lambda,1)}{(\chi^2(n-k)/(n-k))^{1/2}}$, $\lambda = (1\setminus\sigma)(\mathbf{x}_2^\top\mathbf{M_1 x_2})^{1/2}\beta_2 = \sqrt{\Lambda}$.

Pretest estimator: $\acute\beta = \mathbb{I}(F_{\gamma=0} > c_\alpha)\hat\beta + \mathbb{I}(F_{\gamma=0}\le c_\alpha)\tilde\beta = \tilde\beta + \hat\lambda(\hat\beta - \tilde\beta)$ 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

$\tilde{\beta}=\tilde{\beta}$ when not reject.
$\tilde{\beta}-\beta=-\mathbf{Q}(\hat{\lambda}\hat{\gamma}-\gamma)+(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}$
$\mathrm{MSE}(\tilde{\beta})=\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}+\mathbf{Q}\mathrm{MSE}(\hat{\lambda}\hat{\gamma})\mathbf{Q}^\top$

Confidence & Sandwich Cov Matrices

Coverage: P(conf. set includes true value) Test stat: $\tau(\mathbf{y},\theta_0)$. $\theta_0\in$ 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. Asymmetric CI: reject $\hat{\tau}$ if $\hat{\tau}<c_\alpha^-$ or $\hat{\tau}>c_\alpha^+$. $s^2=\mathbf{y}^\top\mathbf{M}_\mathbf{X}\mathbf{y}\backslash(n-k)$
Conf ellipse centr. at $(\hat{\beta_1},\hat{\beta_2})$. Points can be in CI & outside ellipse, and vice versa. Conf. reg. formula: $(\hat{\beta}_2-\beta_{20})^\top\mathbf{X}_2^\top\mathbf{M}_1\mathbf{X}_2(\hat{\beta}_2-\beta_{20})\le c_\alpha k_2 s^2$
$\mathrm{Corr}(X_1,X_2)=\mathbf{x}_1^\top\mathbf{x}_2\backslash(\mathbf{x}_1^\top\mathbf{x}_1)^{1\backslash2}(\mathbf{x}_2^\top\mathbf{x}_2)^{1\backslash2}$
$\mathrm{Corr}(\hat{\beta_1},\hat{\beta_2})$ = -$\mathrm{Corr}(X_1,X_2)$
If no stat with known finite sample dist, use Wald with k_2 vector $\hat{\theta}_2$ asym normal: $(\hat{\theta}_2-\theta_{20})^\top(\widehat{\mathrm{Var}}(\hat{\theta}_2))^{-1}(\hat{\theta}_2-\theta_{20})\le c_\alpha$.
What if disturbances are not IID? $\widehat{\mathrm{Var}}(\hat{\beta})\ne s^2(\mathbf{X}^\top\mathbf{X})^{-1}$. Assume disturbances indep & exogen. regres.
$\mathrm{Var}(\hat{\beta})=(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}$: Sandwich cov matrix. OLS estmtr inefficient, treats all diag. ω_t^2 the same even though some have more weight.
$(1\backslash n)\mathbf{X}^\top\hat{\mathbf{\Omega}}\mathbf{X}=\lim\frac{1}{n}\sum_t^n\omega_t^2x_{ti}x_{tj}$
$\widehat{\mathrm{Var}}(\hat{\beta})=(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{\Omega}}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}$
HC_0: $\hat{u}_t^2$ in diag of $\hat{\mathbf{\Omega}}$ and 0 else.
$\mathrm{plim}\frac{1}{n}\sum_t^n u_t^2x_{ti}x_{tj}\overset{a}{=}\frac{1}{n}\sum_t^n\hat{u}_t^2x_{ti}x_{tj}\to\lim\frac{1}{n}\sum_t^n\omega_t^2x_{ti}x_{tj}$ w/ LLN on $v_t=u_t^2-\varpi_t^2$
$n^{1\backslash2}(\hat{\beta}-\beta_0)\to N(\mathbf{0},\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}(\lim\frac{1}{n}\mathbf{X}^\top\mathbf{\Omega}\mathbf{X})\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1})$
$\lim\frac{1}{n}\widehat{\mathrm{Var}}_h(\hat{\beta})=\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}(\lim\frac{1}{n}\mathbf{X}^\top\mathbf{\Omega}\mathbf{X})\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}$
Then, $\hat{\beta}$ is root-n consistent, asymp. norm
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 $h_t\equiv\mathbf{X_t}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}_\mathbf{t}^\top$.
HC_3: $\hat{u}_t^2/(1-h_t)$ jackknife, for big variance (small residuals).
Ignore hetero for std err of sample mean, since $\lim(1\backslash n)\sum_{t=1}^n(\omega+t^2-\sigma^2)x_{ti}x_{tj}=0$; $1\backslash n\sum_{t=1}^n\omega_t^2\to\sigma^2$. All hetero affects efficiency, but only hetero related to squares and cross products of x_{ti} affects validity of $\hat{\beta}$.
HAC for when u_t hetero and/or autocorr.
$\Sigma=\lim1\backslash n\sum_{t=1}^n\sum_{s=1}^n E(u_tu_s\mathbf{X}_\mathbf{t}^\top\mathbf{X_s})$
Autocovariance matrices of $\mathbf{X}_t^\top u_t$:
$\mathbf{\Gamma}(j)=\frac{1}{n}\sum_{t=j+1}^n E(u_tu_{t-j}\mathbf{X}_\mathbf{t}^\top\mathbf{X}_{t-j})$ $j\ge0$ or
$\mathbf{\Gamma}(j)=\frac{1}{n}\sum_{t=-j+1}^n E(u_tu_{t+j}\mathbf{X}_\mathbf{t}^\top\mathbf{X}_{t+j})$ $j<0$.

However, replacing u_t and u_{t-j} with $\hat{u}_t,\hat{u}_{t-j}$ gives us an inconsistent estimator of $\Gamma(j)$, since j can be very close to n and no LLN will apply to it. To solve this, we only estimate models where $\Gamma(j)\to0$ as $j\to\infty$. Then beyond some threshold, we can assume autocovariance is 0.
$\hat{\Sigma}_{HW}=\hat{\mathbf{\Gamma}}(0)+\sum_{j=1}^p(\hat{\mathbf{\Gamma}}(j)+\hat{\mathbf{\Gamma}}(j))$, $p\to\infty$ with $O(n^{1\backslash4})$ as $n\to\infty$. But HW might not be pos def. Thus, NW estimator.
$\hat{\Sigma}_{NW}=\hat{\mathbf{\Gamma}}(0)+\sum_{j=1}^p(1-\frac{j}{p+1})(\hat{\mathbf{\Gamma}}(j)+\hat{\mathbf{\Gamma}}(j))$
$p\to\infty$ with $O(n^{1\backslash3})$ as $n\to\infty$, since HW underestimates cov matrices, esp for larger vals of j.
$\hat{\Sigma}=(1/n)\mathbf{X}^\top\hat{\mathbf{\Omega}}\mathbf{X}$ (Newey-West/H. White)
Error components model: $u_{gi}=v_g+\epsilon_{gi}$
Cluster: disturbances uncorrelated across clusters but corrlted and hetero within clusters. Block-diag sandwich matrix:
$(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}=(\mathbf{X}^\top\mathbf{X})^{-1}(\sum_{g=1}^G\mathbf{X}_\mathbf{g}^\top\mathbf{\Omega}_\mathbf{g}\mathbf{X}_\mathbf{g})(\mathbf{X}^\top\mathbf{X})^{-1}$
$\mathrm{Var}(\hat{\beta}_2)\backslash\mathrm{Var}_c(\hat{\beta}_2)=1+(n_g-1)\rho$
Group fixed effects: if regressors don't vary within clusters, fixed effects will explain all variation, so we can't tell coef we're interested in. Intra-cluster corr. also comes from data collection, misspecification; more complex than error-components model
$CV_1=\frac{G(n-1)}{(G-1)(n-k)}(\mathbf{X}^\top\mathbf{X})^{-1}(\sum_{g=1}^G\mathbf{X}_\mathbf{g}^\top\hat{\mathbf{u}}_\mathbf{g}\hat{\mathbf{u}}_\mathbf{g}^\top\mathbf{X}_\mathbf{g})(\mathbf{X}^\top\mathbf{X})^{-1}$; cannot exceed G (rank 1 sum).
When $n_g=1\forall g$ s.t. $G=n$, $CV_1=HC_1$. dof adjust for #regressors and #clusters.
For asymp. construct, $G\to\infty$, or $\hat{\beta}$ inconsistent
$M_{gg}=\mathbf{I}_{n_g}-\mathbf{X}_g(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}_g^\top$, diag blocks of $\mathbf{M}_\mathbf{X}$ not proj matrices, pos def
$CV_2=\dot{\mathbf{u}}_g=M_{gg}^{-1/2}\hat{\mathbf{u}}$; $CV_3=\ddot{\mathbf{u}}_g=M_{gg}^{-1}\hat{\mathbf{u}}$
Req to use CV: approp. cluster level, large no, of clusters, disturbances homo across clusters.
$y_{gti}=\eta_g+\lambda_t+u_{gti}$, which represent group fixed effect, time fixed effect and idiosyncratic shock
$y_{gti}=\beta_1+\beta_2D_{gti}^2+\beta_3D_{gti}^2+\delta D_{gti}^bD_{gti}^2+u_{gti}$; $\beta_1=\eta_\alpha+\lambda_1,\beta_2=\eta_\beta-\eta_\alpha,\beta_3=\lambda_2-\lambda_1$
No clustered disturbances: collin. betw jurisdic. dummy and treatm. dummy if $\exists$ jurisd. treated/untreated every period. Can't tell what are treatm. effects vs. cluster effects
MVT = $f(a+h)=f(a)+hf'(a+\lambda h)$ $0<\lambda<1$
First order taylor: $f(a+h)\cong f(a)+hf'(a)$, $h=b-a$. Second order taylor: $f(a+h)=f(a)+hf'(a)+\frac{1}{2}h^2f''(a)$

$f(a+h)=f(a)+\sum_{i=1}^{p-1}\frac{h^i}{i!}f^{(i)}(a)+\frac{hp}{p!}f^{(p)}(a+\lambda h)=f(a)+\sum_{i=1}^p\frac{h^i}{i!}f^{(i)}(a)$, when $\lambda=0$
$f(\mathbf{x}+\mathbf{h})\cong f(\mathbf{x})+\sum_{j=1}^m h_jf_j(\mathbf{x}+\lambda\mathbf{h})$
$n^{1\backslash2}(\hat{\gamma}-\gamma_0)\overset{a}{=}g_0'n^{1\backslash2}(\hat{\theta}-\theta_0)$, and asymp. distribution follows, with $s_\gamma=|g'(\hat{\theta})|s_\theta$
DMI = $[\hat{\gamma}-s_\gamma z_{1-\alpha/2},\hat{\gamma}+s_\gamma z_{1-\alpha/2}]$. Transform I: $[g(\hat{\theta}-s_\theta z_{1-\alpha/2}),g(\hat{\theta}+s_\theta z_{1-\alpha/2})]$.
Use DMI if γ normal, use TI if θ norm.
$n^{1\backslash2}(\hat{\gamma}-\gamma_0)\overset{a}{\sim}N(0,\mathbf{G_0}\mathbf{V}^\infty(\hat{\theta})\mathbf{G}_\mathbf{0}^\top)$, $\mathbf{G_0}$ $l\times k$ is the Jacobian with $\partial g_i(\theta)\backslash\partial\theta_j$
$\widehat{\mathrm{Var}}(\hat{\gamma})=\hat{\mathbf{G}}\widehat{\mathrm{Var}}(\hat{\theta})\hat{\mathbf{G}}^\top$, $\hat{\mathbf{G}}=\mathbf{G}(\hat{\theta})$

Bootstrap

Monte Carlo: Choose DGP in H_0 and sim. samples to get val of (asymp.) pivotal test stat. Sim. samples indep $\Rightarrow$ sim test stats draw from EDF; consistent CDF estimate
$\hat{F}^*(x)=1\backslash B\sum_{b=1}^B\mathbb{I}(\tau_b^*\le x)$
$\hat{p}^*(\hat{\tau})=1-\hat{F}^*(\hat{\tau})=\frac{1}{B}\sum_b^B\mathbb{I}(\tau_b^*>\hat{\tau})$
$\hat{p}_s^*(\hat{\tau})=\frac{1}{B}\sum_b^B\mathbb{I}(|\tau_b^*|>|\hat{\tau}|)$
$\hat{p}_{et}^*(\hat{\tau})=2\min\left(\frac{1}{B}\sum_b^B\mathbb{I}(\tau_b^*\le\hat{\tau}),\frac{1}{B}\sum_b^B\mathbb{I}(\tau_b^*>\hat{\tau})\right)$ when biased param est and 2-tailed.
$\hat{p}^*(\hat{\tau})\to p(\hat{\tau})$ as $B\to\infty$. τ pivotal.
If $\alpha(\beta+1)\in\mathbb{Z}$, MC test is exact. $\hat{p}^*(\hat{\tau})=r\backslash B$. Reject if $r<\alpha\beta\Rightarrow\lfloor\alpha\beta\rfloor+1$ vals where rej. $P(\mathrm{rej.})=(\lfloor\alpha\beta\rfloor+1)\backslash(B+1)$.
This equals α, so $\alpha(\beta+1)=\lfloor\alpha\beta\rfloor+1\in\mathbb{Z}$
For unknown parametric DGP, estimate a bootstrap DGP to draw sim. samples by regressing assuming CNLM.
$y_t=\mathbf{X_t}\tilde{\beta}+\tilde{\delta}y_{t-1}^*+u_t^*,u_t^*\sim NID(0,\tilde{s}^2)$
Resampling: without assuming CNLM bootstrap DGP, get disturbances from EDF of original sample residuals. This EDF is first centered ($-\bar{u}\iota$), then rescaled by multiplying by $(n\backslash(n-k))^{1\backslash2}$
GR1: Bootstrap DGP $\in\mathbb{M}_0$
GR2: Unless the test stat is pivotal for $\mathbb{M}_0$, bootstrap DGP should be best possible estimate of the true DGP, assuming true DGP $\mu\in\mathbb{M}_0$
Loss of power due to finite B
Pairs: $y_t^*=\mathbf{X}_{s1}\hat{\beta}_1+\hat{u}_s$, $\hat{u}_s$ from unres. mdl, s resampling index, $\tilde{\beta}$ from res. mdl
Wild: $y_t^*=\mathbf{X}_t\tilde{\beta}+s_t^*\tilde{u}_t$, $s_t^*=-1$ w/ prob 1/2 and 1 w/ prob 1/2 (Rademacher)
Block: Resample l-length blocks of rescaled residuals or $[\mathbf{y},\mathbf{X}]$ pairs. $l=O(n^{1\backslash3})$
Stationary: l random. Moving block: $n-l+1$ overlapping blocks, shift by 1
Block-of-blocks: Resample from $n-l+1$ moving blocks
$\mathrm{AR}(p)$ process: $y_t=\rho_0+\sum_{i=1}^p\rho_iy_{t-i}+u_t,u_t\sim IID(0,\sigma^2)$. $\hat{u}_t=\sum_{i=1}^p\rho_i\hat{u}_{t-i}+\epsilon_t$, $p=t+1\dots n$. We obtain $\hat{\sigma}_\epsilon^2$ and $\hat{\rho}_i$.
Sieve: $u_t^*=\sum_{i=1}^p\hat{\rho}_iu*_{t-i}+\epsilon_t^*$, where $\epsilon_t^*\sim N(0,\hat{\sigma}_\epsilon^2)$ or resampled
$t_b=(\theta_b^*-\hat{\theta})\backslash s_b^*$. Change null to $\theta_0=\hat{\theta}$
CI = $[\hat{\theta}-s_\theta c_{1-\alpha\backslash2}^*,\hat{\theta}-s_\theta c_{\alpha\backslash2}^*]$, $c_{\alpha\backslash2}^*=t$-val for $r_{\alpha\backslash2}=\lceil\alpha\beta\backslash2\rceil$-th val of t_b^*
$\tau_b^*=(\theta_b^*-\hat{\theta})^\top(\mathrm{Var}^*(\theta_b^*))^{-1}(\theta_b^*-\hat{\theta})$, and CR is Wald with $c_\alpha^*=(B+1)(1-\alpha)$-th τ_b^*

Instrumental Variables

Can define $E(u_t\mid\Omega_t)=0$.
Err in variables: indep vars in regr model measured with err. $u_t=u_t^\circ+v_{2t}-\beta_2v_{1t}\Rightarrow E(u_t\mid x_t)\ne0$, $\mathrm{Cov}(x_t,u_t)=E(x_tu_t)\ne0$. OLS est biased and inconsist. Simultaneity: two or more endog vars jointly determined by sys of simultaneous eq.
Assume, $E(\mathbf{u}\mathbf{u}^\top)=\sigma^2\mathbf{I}$ 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 $\mathbf{S}_{\mathbf{W}^\top\mathbf{X}}\equiv\mathrm{plim}\frac{1}{n}\mathbf{W}^\top\mathbf{X}$ is deterministic and non-sing. Same for $\mathbf{S}_{\mathbf{W}^\top\mathbf{W}}$. $\hat{\beta_{IV}}$ consistent iff $\mathrm{plim}\frac{1}{n}\mathbf{W}^\top\mathbf{u}=0$ (disturb. asymp. uncorr w/ instr.). Asym cov mtrx of IV est: $\sigma_0^2\mathrm{plim}(n^{-1}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}$.
If overidentified ($l>k$), we aim to find $\mathbf{WJ}$ s.t. $\mathbf{J}$: full col rank, asym deterministic, min asym cov mtrx of IV est.
$\mathbf{X}=\bar{\mathbf{X}}+\mathbf{V}$, $E(\mathbf{V}_t|\Omega_t)=\mathbf{0}$, $\bar{\mathbf{X}}_t=E(\mathbf{X}|\Omega_t)$
$\bar{X}$ optimal instr, by LLN & $E(\mathbf{V}^\top\mathbf{W})=\mathbf{O}$, and since $n^{-1}\bar{\mathbf{X}}^\top\mathbf{M}_\mathbf{W}\bar{\mathbf{X}}$ pos semidef. All exo/predet explan. vars shld be in $\mathbf{W}$.
GIVE: $\hat{\beta}_{IV}=(\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{y}$ IV est: Col of $\mathbf{P}_\mathbf{W}\mathbf{X}$ should be lin indep. Asymptotically, $\mathbf{S}_{\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X}}$ determ. & non-sing.
IV asym normal like all est. $n^{1/2}(\hat{\beta}_{IV}-\beta_0)=(n^{-1}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}n^{-1/2}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{u}$, apply CLT. $\widehat{\mathrm{Var}}(\hat{\beta}_{IV})=\hat{\sigma}^2(\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}$, $\hat{\sigma}^2=(1\backslash n)\|\mathbf{y}-\mathbf{X}\hat{\beta}_\mathbf{IV}\|^2$ since σ_0^2 unknown
$\hat{\beta}_{IV}-\beta_0=\frac{\sigma_u\mathbf{w}^\top(\rho\mathbf{v}+\mathbf{u_1})}{\pi_0+\sigma_v\mathbf{w}^\top\mathbf{v}}=\frac{\rho\sigma_uz}{\sigma_v(a+z)}$, $z=\mathbf{w}^\top\mathbf{v},a=\pi_0/\sigma_v$ Unbias. when $\rho=0$, otherwise moments don't exist.
W_{β_2}: $\mathbf{y}-\mathbf{X}_1\tilde{\beta}_1=\mathbf{P}_\mathbf{W}\mathbf{X}_1\mathbf{b}_1+\mathbf{P}_\mathbf{W}\mathbf{X}_2\mathbf{b}_2$+res ESS divided by consistent est. of σ^2, ESS $O_p(1)=\mathbf{y}^\top\mathbf{P}_{\mathbf{P}_\mathbf{W}\mathbf{X}}-\mathbf{P}_{\mathbf{P}_\mathbf{W}\mathbf{X}_1}\mathbf{y}/\sigma^2$.
$\mathbf{M}_{\mathbf{P}_\mathbf{W}\mathbf{X}_1}\mathbf{P}_\mathbf{W}=\mathbf{P}_\mathbf{W}-\mathbf{P}_{\mathbf{P}_\mathbf{W}\mathbf{X}_1}$, $\mathbf{P}_{\mathbf{P}_\mathbf{W}\mathbf{X}}=\mathbf{P}_{\mathbf{P}_\mathbf{W}\mathbf{X}_1}+\mathbf{P}_{(\mathbf{P}_\mathbf{W}-\mathbf{P}_{\mathbf{P}_\mathbf{W}\mathbf{X}_1})}$, RHS orthog. Test stat is $nR_u^2\overset{a}{\sim}\chi^2(l-k)$. Can't use F stat: oblique proj, so TSS $\ne$ ESS + SSR. $Q(\hat{\beta},\mathbf{y})=(\mathbf{y}-\mathbf{X}\hat{\beta})^\top\mathbf{P}_\mathbf{W}(\mathbf{y}-\mathbf{X}\hat{\beta})$. Use HAC/HCCME for hetero/autocorr.
Test overidentification: check $\mathbf{W}^{*\top}\mathbf{u}\ne0$. Large val means either misspec. mdl or omitted instr.s from regression func
$\hat{\beta}_{IV}-\hat{\beta}_{OLS}=(\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{M}_\mathbf{X}\mathbf{y}$.
DWH: Test if $\mathbf{Y}^\top\mathbf{P}_\mathbf{W}\mathbf{M}_\mathbf{X}\mathbf{y}=\mathbf{0}$ w/ regression $\mathbf{y}=\mathbf{X}\beta+\mathbf{P}_\mathbf{W}\mathbf{Y}\delta+\mathbf{u}$, and F test if $\hat{\delta}=0$. If H_0 rej, $\mathbf{Y}$ endog or $\mathbf{P}_\mathbf{W}\mathbf{Y}$ good expl pwr.
Bootstrap w/ red. form eqs, get $\tilde{\beta}$&$\tilde{\mathbf{u}}$, then get $\hat{\Pi}_2$&$\hat{\mathbf{V}}_2$. For resamp, draw rows $\hat{\mathbf{V}}_t$. For param., estimate $\hat{\sigma}=1/n\hat{\mathbf{V}}^\top\hat{\mathbf{V}}$

Generalized Least Squares

Consider $E(\mathbf{u}\mathbf{u}^\top)=\mathbf{\Omega}$, $\mathbf{\Omega}^{-1}=\mathbf{\Psi}\mathbf{\Psi}^\top$
$\hat{\beta}_{GLS}=(\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{y}$
$E(\mathbf{\Psi}^\top\mathbf{u}\mathbf{u}^\top\mathbf{\Psi})=\mathbf{I}$.
$\mathrm{Var}(\hat{\beta}_{GLS})=(\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X})^{-1}$ G-M assump.
GLS Criterion: $(\mathbf{y}-\mathbf{X}\beta)^\top\mathbf{\Omega}^{-1}(\mathbf{y}-\mathbf{X}\beta)$
$\widehat{\mathrm{Var}}(\hat{\beta}_{GLS})=s^2(\mathbf{X}^\top\mathbf{\Delta}^{-1}\mathbf{X})^{-1}$, $\mathbf{\Omega}=\sigma^2\mathbf{\Delta}$
Feasible: $E(u_t)^2=\exp(\mathbf{Z_t}\gamma)$ regress $\log\hat{u}_t^2=\mathbf{Z_t}\gamma+v_t$ for $\hat{\gamma}$, $\hat{\omega}_t=(\exp(\mathbf{Z_t}\hat{\gamma}))^{1/2}$. $\hat{\beta}$ consis $\Rightarrow\hat{\gamma}$ consis.
$\hat{u}_t^2=b_\delta+\mathbf{Z_t}\mathbf{b}_\gamma$ + res, w/ δ methd. $H_0:\mathbf{b}_\gamma=\mathbf{0}$, F stat or nR_c^2 asymp $\chi^2(r)$
$u_t=\rho u_{t-1}+\epsilon_t$, $\epsilon_t\sim IID(0,\sigma_\epsilon^2)$, $|\rho|<1$, $\sigma_u^2=\sigma_\epsilon^2/(1-\rho^2)$, $\mathrm{Cov}(u_t,u_{t-1})=\rho\sigma_u^2$
autocov mtrx of AR(1): $\mathbf{\Omega}(\rho)=\sigma_\epsilon^2\backslash1-\rho^2\times$ mtrx with 1 diag and ρ^i incr. away from diag.; this mtrx is $\mathbf{\Delta}(\rho)$
MA(1): $u_t=\epsilon_t+\alpha_1\epsilon_{t-1}$, $\epsilon_t\sim IID(0,\sigma_\epsilon^2)$
$\sigma_u^2=(1+\alpha_1^2)\sigma_\epsilon^2$, $\mathrm{Cov}(u_t,u_{t-1})=\alpha_1\sigma_\epsilon^2$
Cov mtrx: $\mathbf{\Omega}(\alpha_1)=\sigma_\epsilon^2\mathbf{\Delta}(\alpha_1)$, $\mathbf{\Delta}(\alpha_1)$ w/ $(1+\alpha_1)^2$ diag and α_1 1 from diag, 0 else.
$y_t=\mathbf{X}_t(\tilde{\beta}+\mathbf{b}_\beta)+(y_{t-1}-\mathbf{X}_{t-1}\tilde{\beta})\rho+\epsilon_t\Rightarrow\mathbf{y}=\mathbf{X}\beta+b_\rho\tilde{\mathbf{u}}_{t-1}+\epsilon$, asymp. t/F test $b_\rho=0$
t stat pivotal since scale invar. and depends on res and $\mathbf{X}$ only $\Rightarrow$ exact Monte Carlo. For AR(1), t. AR(p): F. If lagged dep. vars, generate 1 row at a time. If not CNLM, resample.
Hetero-robust: Wald w/ HCCME, Wild
$\mathbf{\Psi}(p)$: $(1-\rho^2)^{1/2}$ in first, 1 in diag, $-\rho$ one above diag, ρ estim by $b\rho$ FWL regress
If explan. var not all exogenous, don't use GLS. Iterated GLS: OLS, get ρ, feasible GLS, update res, repeat. Underspec. looks like serial corr.
Random effects mdl: $E(v_i|\mathbf{X})=0\Rightarrow E(u_{it}|\mathbf{X})=0$, but u_{it} not IID. $\Sigma=\sigma_\epsilon^2\mathbf{I}_T+\sigma_v^2\iota\iota^\top$, on m block diag of Ω
Regress $\mathbf{P_D}\mathbf{y}=\mathbf{P_D}\mathbf{X}\beta$ + res (BG est), var is $\sigma_v^2+\sigma_\epsilon^2\backslash T$. Divide SSR by $m-k$, minus $\hat{\sigma}_\epsilon^2/T$ to get $\hat{\sigma}_v^2$. $\hat{\beta}=(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{M_D}\mathbf{X}\hat{\beta}_{FE}+(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P_D}\mathbf{X}\hat{\beta}_{BG}$
$(\mathbf{I}-\lambda\mathbf{P_D})\mathbf{y}=(\mathbf{I}-\mathbf{P_D})\mathbf{X}\beta$ + res, $\lambda=1-(T\sigma_v^2/\sigma_\epsilon^2+1)^{-1/2}$. If unbal, $T=T_i$.