Regression Assumptions

Assumption	Formulas	Description	Violation Consequences
Linearity (in parameters)	$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ik}\beta_k$ where $x_{i1} = 1, \forall i$ and $E[y\|x] = \beta_1 f_1(.) + \beta_2 f_2(.) + ... + \beta_k f_k(.)$	Allows for the parameters to be independent from each other and not interacting	Least Squares estimation technique cannot be used on such a model
Full Rank	$det(\mathbf{X}) \neq 0$	No exact linear relationship between any of the independent variables (the regressors)	Multi-colinearity will result and the estimation will fail without full rank (linear independence)
Independent ϵ	$E[\epsilon_i\|\epsilon_1, \epsilon_2, ..., \epsilon_n] = 0$	Each draw of ϵ is random and independent from other ϵ	Problematic for time-series; can generate spurious regressions because contributions from the dynamic interaction of disturbances are attributed to the regressors
Exogeneity of Regressors	$E[\epsilon\|\mathbf{X}] = \mathbf{0} \implies E[\epsilon_i] = 0$	Zero Conditional Mean; all of which implies $E[\epsilon_i\|x_i] = 0$	Model is then not properly specified and there would be an omitted variable bias
Homoscedasticity and Non-autocorrelation	$Var[\epsilon_i\|\mathbf{X}] = \sigma^2$ and $\text{Cov}[\epsilon_i, \epsilon_j] = 0$ when $i \neq j$	Variance of ϵ are not correlated with $\mathbf{X}$ and each ϵ realisation is uncorrelated	Violation of homoscedasticity does not bias coefficient estimates (though it becomes inefficient) but makes LS standard errors invalid; autocorrelation should make the LS estimation technique invalid under no adjustments
Data Generation		$\mathbf{X}$ may be fixed or random	Fixed $\mathbf{X}$ implies that the probability distributions of the ϵ can be imagined as unconditional and random $\mathbf{X}$ implies the need for conditioning; this affects how we view the sampling distributions and tests.
Normality of ϵ	$\epsilon\|\mathbf{X} \sim N(0, \sigma^2 I)$	ϵ is distributed normally	Test statisics and confidence intervals are constructed based on this assumption

Relaxation of Assumptions

Issue	Formula	Impacts on Regression	Solutions or Rectification
Multi-colinearity	$Var(\mathbf{b}_k\|\mathbf{X}) = \frac{\sigma^2}{(1-R_k^2)\mathbf{S}_{kk}}$ where R_k^2 is the R^2 of the regression of variable x_k on the other explanatory variables, $x_i, \forall i \neq k$	Raise Standard errors; signs and magnitudes of coefficients may be unexpected and estimates sensitive to specific data points or presence of certain other explanatory variables	Transform variables, drop colinear variables or change sampling methods
Hetero-scedasticity	$E(\varepsilon'\varepsilon\|\mathbf{X}) = Var(\varepsilon\|\mathbf{X}) = \sigma^2\Omega = \Sigma$	LSE remains unbiased: $E(\mathbf{b}\|\mathbf{X}) = E(\beta\|\mathbf{X}) + E[(\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon\|\mathbf{X}] = \beta$	Estimate $Q = \frac{1}{n}\sum \sigma^2 \mathbf{x_i x_i}'$ with $S_o = \frac{1}{n}\sum_i^n e_i^2 \mathbf{x_i x_i'}$ and White's Estimated Asymptotic $Var(b) = n(\mathbf{X'X})^{-1}S_o(\mathbf{X'X})^{-1}$
	where Ω is no longer the identity matrix; note that in the way it is written, there is still a common scalar constant	Standard Errors Invalid: $E(\varepsilon'\varepsilon\|\mathbf{X}) = \sigma^2\Omega$ so $Var(\mathbf{b}\|\mathbf{X}) = \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'}\Omega\mathbf{X}(\mathbf{X'X})^{-1}$ so $s^2(\mathbf{X'X})^{-1}$ no longer an estimator since s^2 is a biased estimator for σ^2 and $(\mathbf{X'X})^{-1}$ is now the wrong matrix	GLS Estimator: $\hat{\beta} = (\mathbf{X'\Omega^{-1}X})^{-1}\mathbf{X'\Omega^{-1}y} = [\sum_i^n \frac{x_i x_i'}{\omega_i}]^{-1}\sum_i^n \frac{x_i y_i}{\omega_i}$; essentially dividing across by ω_i; with $\hat{\sigma^2} = \frac{1}{n-K}\sum_i^n \frac{y_i - \mathbf{x_i'}\hat{\beta}}{\omega_i}$
Endogeneity	$E(\varepsilon_i\|\mathbf{x_i}) = \eta_i, \eta_i \neq 0$ & $E(x_i\varepsilon_i) = \gamma, \gamma \neq 0 \implies plim(\frac{\mathbf{X}'\varepsilon}{n}) = \gamma$	All coefficient estimates of Least Squares becomes **biased** and **inconsistent**; $E(\mathbf{b^{LS}}\|\mathbf{X}) = \beta + (\mathbf{X'X})^{-1}\mathbf{X'}\eta$ & $plim(\mathbf{b^{LS}}) = \beta + [plim(\frac{\mathbf{X'X}}{n})]^{-1}[plim(\frac{\mathbf{X}'\varepsilon}{n})] = \beta + \mathbf{Q_{XX}^{-1}}\gamma$	Use Instrumental Variable; could arise from: 1. Measurement Errors 2. Simultaneous Equations Bias (Reverse Causality) 3. Omitted Variable Bias 4. Lagged Dependent Variable

Least Square Estimator Properties

Types	Formula	Remarks
Finite Sample	$\mathbf{b} = \beta + (\mathbf{X'X})^{-1}X'\varepsilon$ $Var(\mathbf{b}\|\mathbf{X}) = \sigma^2(\mathbf{X'X})^{-1}$	Linear in ε, allowing expectation to be taken such that $E(\mathbf{b}) = \beta$ Variance relies on the homoscedasticity & nonautocorrelation assumption that allows $E(\varepsilon\varepsilon'\|\mathbf{X}) = \sigma^2\mathbf{I}$
Probability Limits Properties	Slutsky Theorem Other plim 'rules' Matrices plims Limiting Distributions	If X_n is random and $plim(X_n) = \theta$ [meaning $\lim_{x\to\infty} P(\|X_n - \theta\| \geq \varepsilon) = 0$] then for $g(.)$ a C^1 function, $plim(g(x)) = g(plim(x))$ For any x_n and y_n such that $plim(x_n) = \theta$ and $plim(y_n) = \mu$, 1. $plim(x_n \pm y_n) = \theta \pm \mu$ 2. $plim(x_n \times y_n) = \theta \times \mu$ 3. $plim(\frac{x_n}{y_n}) = \frac{\theta}{\mu}$ (if $\mu \neq 0$) 4. $plim(g(x_n, y_n)) = g(\theta, \mu)$ (assuming $g(.)$ is C^1 with continuous partials, etc) If $\mathbf{A}_n$ and $\mathbf{B}_n$ such that $plim(\mathbf{A}_n) = \mathbf{A}$ and $plim(\mathbf{B}_n) = \mathbf{B}$ (element by element), then 1. $plim(\mathbf{A}_n^{-1}) = [plim(\mathbf{A}_n)]^{-1} = \mathbf{A}^{-1}$ 2. $plim(\mathbf{A}_n\mathbf{B}_n) = [plim(\mathbf{A}_n)][plim(\mathbf{B}_n)] = \mathbf{AB}$ If $x_n \xrightarrow{d} x$ & $g(x_n, \theta) \xrightarrow{d} g$ then if $plim(y_n) = \theta$, then $g(x_n, y_n) \xrightarrow{d} g$
Asymptotic Properties	Consistency Limiting Distribution Asymptotic Efficiency	$\mathbf{b} \to \beta$, converges to a constant with zero variance $\mathbf{b} \xrightarrow{a} N(\beta, \frac{\sigma^2}{n}\mathbf{Q^{-1}})$, converges to an approximate distribution, applying Slutsky Theorem & Delta method Limiting variance: $Var(\mathbf{b}) \to \frac{\sigma^2}{n}\mathbf{Q^{-1}}$, converging at rate of $\frac{1}{\sqrt{n}}$ where the estimator for the asymptotic variance is $\frac{s^2}{n}(\frac{X'X}{n})^{-1} = s^2(\mathbf{X'X})^{-1}$ Derived only from the asymptotically normal distributed estimators with the delta method of deriving asymptotic variances. Defined as having a covariance matrix than that of no other asymptotically normal distributed estimators

Functional Form

Model	Dependent Variable	Independent Variable	Interpretation of β_2	Purpose
Linear	Y	X	$\Delta Y = \beta_2 \Delta X$	Marginal Effect, $\beta_2 = \frac{\Delta Y}{\Delta X}$
Log-Linear	$\ln Y$	$\ln X$	$\%\Delta Y = \beta_2 \%\Delta X$	Elasticity, $\beta_2 = \frac{\%\Delta Y}{\%\Delta X} = \frac{\partial Y}{\partial X}\frac{X}{Y}$
Semi-Log	$\ln Y$	X	$\%\Delta Y = 100\beta_2\Delta X$ Exact $\%\Delta Y = 100(e^{b_2} - 1)\Delta X$	Marginal Effect on $\%\Delta Y$
Linear-Log	Y	$\ln X$	$\Delta Y = \frac{\beta_2}{100}\%\Delta X$	Marginal Effect of $\%\Delta X$
Quadratic	Y	X & X^2	$\Delta Y = (\beta_2 + \beta_3 X)\Delta X$	Marginal Effects of X changes at every X

*Note on Semilog model: $\hat{y} = e^{\hat{\log y}}$ systematically underestimates the expected value of y and the consistent estimator for y is $\hat{y} = e^{\frac{s^2}{2}} e^{\hat{\log y}}$ if error terms are known to be normally distributed. If they are not, then we have to estimate $\hat{\alpha}$ in $\hat{y} = \hat{\alpha}e^{\hat{\log y}}$. A method to do this is to create a variable $\hat{m} = e^{\log y}$ then regress y on $\hat{m}$ without a constant (model is $y = \hat{\alpha}\hat{m}$); the coefficient on $\hat{m}$ is the estimator.

Frisch-Waugh-Lowell Theorem

For regression of y on $[i\ X]$, $b_2 = [\mathbf{X}'(\mathbf{I} - \mathbf{i}(\mathbf{i'i})^{-1}\mathbf{i}')\mathbf{X}]^{-1}[\mathbf{X}'(\mathbf{i}(\mathbf{i'i})^{-1}\mathbf{i}')\mathbf{y}]$; defining $\mathbf{M}_0 = \mathbf{I} - \mathbf{i}(\mathbf{i'i})^{-1}\mathbf{i}'$, we know $\mathbf{M}_0$ is the mean-deviation maker so $\hat{b} = [\mathbf{X}'\mathbf{M}_0'\mathbf{M}_0\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{M}_0'\mathbf{M}_0\mathbf{y}] = [\mathbf{X}'\mathbf{M}_0\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{M}_0\mathbf{y}] = \mathbf{b}$

Likewise for y on $[X_1\ X_2]$, $b_2 = [\mathbf{X}_2'(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_1]^{-1}[\mathbf{X}_2'(\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{y}]$; defining $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$, we know bfM_1 is the residual maker of y on X_1 so we have $\mathbf{a}_2 = [\mathbf{X}_2'\mathbf{M}_1'\mathbf{M}_1\mathbf{X}_2]^{-1}[\mathbf{X}_2'\mathbf{M}_1'\mathbf{M}_1\mathbf{y}] = [\mathbf{X}'\mathbf{M}_0\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{M}_0\mathbf{y}] = \mathbf{b}_2$ where $\mathbf{a}_2$ is the coefficient vector of the regression of residuals of y on $\mathbf{X}_1$ on that of $\mathbf{X}_2$ on $\mathbf{X}_1$.

Tests & Test Statistics

Tests	Purpose	Test-statistic
Chow Test	Whether to run regressions as pooled or separate	$F_{(g-1)k,(n-gk)} = \frac{(SSE_r - SSE_u)/(g-1)k}{SSE_u/(n-gk)}$ where $(g-1)k$ is the number of restrictions since we have gk instead of k parameters now. Chow Test assumes $SSE_r = SSE_p$ & $SSE_u = SSE_1 + ... + SSE_g$
Single Restrictions	Test if $\mathbf{R}\beta - \mathbf{q}$ is sufficiently close to 0 (Wald Distance Measure)	$\mathbf{t_k} = \frac{\mathbf{b_k} - \beta_k^0}{\sqrt{\sigma^2 \mathbf{S^{kk}}}}$ distributed t-distribution dF $n-1$ equivalent to the square root of the F-ratio
Multiple Restrictions (Wald Criterion)	Testing whether $m = \mathbf{R}\beta - \mathbf{q}$ is merely sampling error Alternative Method Simplified Method	$W = \mathbf{m'}(Var[\mathbf{m}\vert\mathbf{X}])\mathbf{m} = (\mathbf{Rb} - \mathbf{q})'[\sigma^2 \mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}](\mathbf{Rb} - \mathbf{q}) \sim \chi^2_J$ but σ^2 unknown so we use $s^2 = \frac{\mathbf{e'e}}{n-K}$ as estimate converting the statistic into a F-statistic: $F_{J,n-K} = \frac{(\mathbf{Rb}-\mathbf{q})'[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}]^{-1}(\mathbf{Rb}-\mathbf{q})]/\mathbf{J}}{\mathbf{e'e}/(\mathbf{n}-\mathbf{K})}$ $F_{J,n-K} = \frac{(\mathbf{e_r'e_r} - \mathbf{e'e})/\mathbf{J}}{\mathbf{e'e}/(\mathbf{n}-\mathbf{K})}$ where $\mathbf{e_r}$ is the sum of sq residuals from the restricted regression $F_{J,n-K} = \frac{(R^2 - R_r^2)/J}{(1-R^2)/(n-K)}$
Instrumental Variables (Hausman) Wu-Hausman Test	Testing for systematic differences across coefficient estimates from OLS & IV Testing if residuals from 1st stage is significant when included in structural model	$H = (\mathbf{b^{IV}} - \mathbf{b^{LS}})[\hat{Var}(\mathbf{b^{IV}}) - \hat{Var}(\mathbf{b^{LS}})](\mathbf{b^{IV}} - \mathbf{b^{LS}})$; where $\hat{Var}(\mathbf{b^{IV}}) = \frac{\hat{Var}(\mathbf{b^{LS}})}{\mathbf{r^2_{ZX}}}$. $H \sim \chi^2_k$ where k is the number of regressors Saving the residuals of the first stage regression and then running $y = \mathbf{X}\beta + \mathbf{X}^*\gamma + \varepsilon^*$ where X^* is the n by K* matrix of residuals from the first stage.
Over-identification	Check if excess instruments are invalid such that $Cov(Z, \varepsilon) \neq 0$ Modified Breusch-Pagan Version	On STATA: use 'ivregress 2SLS' then use 'estat overid'; χ^2_k-distribution where k is the number of instruments minus number of endogenous variables. Use IV regression, save the residuals and then regress saved residuals on the exogenous variables and instruments and then compute nR^2 which is a χ^2-distribution degree of freedom equals to above.
Hetero-scedasticity	White Test - Checking if the residuals of regression are heteroscedasti (in general); makes no assumption about the type of heteroscedasticity	On STATA: use 'estat imtest, white'. Regress the square of the residuals of the regression on the squared regressors and the cross products of the regressors (auxillary regression). Calculate test statistic nR^2 which is a χ^2-distribution with degree of freedom equals to number of regressors in this auxillary regression. $H_0 = \sigma_i^2 = \sigma^2$ and $H_0 = \sigma_i^2 \neq \sigma^2$
'Known' Hetero-scedasticity	Breusch-Pagan Test - assumes some specific sort of heteroscedasticity	On STATA: use 'estat hettest'. Stata takes the square of the residuals and regress them on the predicted values from the original regression ($\hat{\mathbf{y}}$), then compute nR^2 from this auxillary regression. Has a χ^2-distribution with degrees of freedom equals to the rank of the 'Z' matrix (which is the matrix of constant and the variables related to the variance of disturbance) minus 1 (due to considering the constant term). STATA is assuming X is Z in this case since $\hat{y}$ encapsulate all the information in X. Has the same H_0 but a more specific H_1

General Concepts

Concepts	Details
Weak Stationarity	Conditions: 1. Mean Reversion; Finite mean & constant 2. Finite Variance & constant (time-invariant) unconditional variance $Var(X_t) = Var(X_{t-1})$ 3. Covariance between terms only depends on the distance between observations ($\|t-s\|$) and not on the time itself (value of t or s) In a random walk of $y_t = y_{t-1} + u_t \Rightarrow y_t = y_0 + \sum_i^t = 0u_i$, 1. Taking expectations we see that $E(y_t) = y_0$, so first condition holds; we assume $y_0 = 0$ 2. Variance will be $Var(y_t) = \rho^2 Var(y_{t-1}) + Var(u_t) = \rho^2 E(y_{t-1}^2) + \sum^t \sigma_u^2 = \rho^2 E(y_{t-1}^2) + t\sigma_u^2$, which means it is dependent on time, t 3. Covariance of y_t and y_{t-h} will also depend on t: $Cov(y_t, y_{t-h}) = E(y_t y_{t-h}) = E[\sum_{s=1}^t y_s(\sum_{k=1}^{t-h} y_k)] = \sum_{k=1}^{t-h} E(y_k^2) = \sum_{k=1}^{t-h} \sigma_y^2 = (t-h)\sigma_y^2$, for all $h > 0$
Auto-correlation Function (ACF)	Obtained by taking autovariance λ_k by variance λ_0 $\rho_k = \frac{\lambda_k}{\lambda_0}, \rho_k \in [-1, 1]$ Gives a sequence of $Corr(y_t, y_{t-1})$, $Corr(y_t, y_{t-2})$, $Corr(y_t, y_{t-3})$,... **without** holding the effects of intermediate lags constant. Allows identification of the MA(q) process because the function value abruptly drops to zero at one lag past q, the order of the process
Partial ACF	Gives the ACF except it holds the effects of intermediate lags constant.
Augmented Dickey-Fuller (ADF) Test vs Phillips -Perron (PP) Test	ADF accommodates some forms of serial correlation by adding the p lagged values of the first difference of the variable tested. $y_t = \gamma y_{t-1} + \alpha_1 \Delta y_{t-1} + \sum_{i=2}^p \alpha_i \Delta y_{t-i}$ or $\Delta y_t = \theta y_{t-1} + \alpha_1 \Delta y_{t-1} + \sum_{i=2}^p \alpha_i \Delta y_{t-i}$ DF Test-statistic: $\tau = \frac{\hat{\gamma}-1}{SE(\hat{\gamma})}$ or $\tau = \frac{\hat{\theta}}{SE(\hat{\theta})}$ - in both cases we need to select the value of p and test against the χ_p^2 distribution. In the case of ADF, when we select too few lags, then we get a biased estimate (because we leave autocorrelation in the disturbance) but reduce the power of the unit root test when we select too many lags. PP overcomes the problem using test-statistics that are modified (Z_t stats) that does not depend on the number of lags. PP also doesn't assume (unlike ADF) that ε_t are conditionally heteroscedastic; yet ADF would be more powerful than PP if p is correctly specified. Also, the PP test requires a "bandwidth" parameter selection (as part of the construction of the Newey-West covariance estimator) that creates finite sample problems analogous to those associated the lag length selection issue in applying the ADF test.

Discrete Choice Models

Models	Assumptions	Description	Application/Problems
Linear Probability Model (OLS)	$E(y\|x) = P(y = 1\|x) = F(x'\beta) \Rightarrow P(y = 0\|x) = 1 - F(x'\beta)$ $\hat{y}_i = P(y_i = 1\|x)$	Unbiased β as long as exogeneity, linearity and non-multicolinearity assumption holds; Conditional expectation of y and the 'predicted' level of y have an intitutive interpretation as the probability that the outcome variable equals 1 given X	Disturbances are heteroscedastic; probabilities may not be linear in X; predicted probabilities could exceed 1 or be negative; therefore use FGLS or robust SE to correct for the hetereoscedasticity for statistical inference
Linear Probability Model (FGLS)	$Var(\varepsilon\|X) = x'\beta(1-x'\beta)$ due to contraints on the value of the dependent variable	Regress the standard LPM, obtain $\hat{y} = x'\beta$; drop all $\hat{y}_i > 1$ observations and define $h_i = \sqrt{\hat{y}_i(1-\hat{y}_i)}$ and run FGLS: $\frac{y_i}{h_i} = \frac{1}{h_i} + \sum_i (\frac{x_i}{h_i})'\beta_i$ (note: constant $= \frac{1}{h_i}$)	Overcomes the heterogeneity problem and should give non-negative probabilities which do not exceed 1
Probit Model	Assumes a standard normal cumulative distribution function $F(z) = \int_{-\infty}^{z} f(z)dz$	$z = x'\beta$ & $P(y = 1\|x) = F(z)$ so since $F(.)$ is the cdf of a probability function, we have $0 < F(z) < 1$. In fact, as $x'\beta \to +\infty$, $F(x'\beta) = 1$ and $x'\beta \to -\infty$, $F(x'\beta) = 0$. z is a continous latent variable that is positive when $y = 1$ and negative when $y = 0$	Standard Normal pdf: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x'\beta)^2}{2}}$; Marginal Effects calculated using $\frac{\partial E(y\|x)}{\partial x} = f(x'\beta)\beta$
Logit Model	Assumes a logistic cumulative distribution function $F(z) = \frac{e^z}{1+e^z}$		Marginal Effects are not constant across x-values

Stationary vs Nonstationary Models

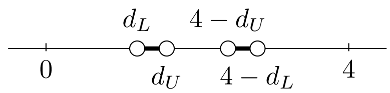
Models	Assumptions	Description	Application/Implications
Stationary AR(1)	$X_t = \alpha + \rho X_{t-1} + \varepsilon_t$ $\alpha \neq 0$, $\|\rho\| < 1$ & $\beta \neq 0$	Stationary; I(0) with expectations a constant	$E(X_t) = \mu = \frac{\alpha}{1-\rho}$
Deterministic Trend	$X_t = \alpha + \beta t + \varepsilon_t$ $\alpha \neq 0$ & $\beta \neq 0$	Stationary; I(0) with expectations non-zero	Non- autocorrelated time series fluctuating around a nonzero value, $E(X_t) = \alpha + \beta t$
Deterministic Trend with Stationary Component	$X_t = \alpha + \rho X_{t-1} + \beta t + \varepsilon_t$ $\alpha \neq 0$, $\|\rho\| < 1$ & $\beta \neq 0$	Stationary; I(0) with expectations non-zero	$E(X_t) = \mu + \delta t = [\frac{\alpha}{1-\rho} - \frac{\rho\beta}{(1-\rho)^2}] + \frac{\beta}{1-\rho}t$; Autocorrelated time series fluctuating around a nonzero, rising/falling value
Random Walk	$X_t = \rho X_{t-1} + \varepsilon_t$ $\rho = 1$	Non-stationary; I(1) with expectations zero	
Random Walk with Drift	$X_t = \alpha + \rho X_{t-1} + \varepsilon_t$ $\alpha \neq 0$ & $\rho = 1$	Non-stationary; I(1) with expectations non-zero	
Random Walk with Drift & Deterministic Trend	$X_t = \alpha + \rho X_{t-1} + \beta t + \varepsilon_t$ $\alpha \neq 0$, $\rho = 1$ & $\beta \neq 0$	Non-stationary; I(1) with expectations non-zero	
Bivariate Cointegration	Y_t & X_t have a long-run stationary relationship $\Rightarrow u_t = Y_t - \beta X_t$ should be I(0)	Residuals represents the long-run relationship	Testing the residuals of a regression of 2 non-stationary variables to see if it is stationary
Multi-variate Cointegration	Johansen approach assumes each independent variable has a LR relation with the dependent variable	For y_t, x_t, z_t, we assume LR1(y_t, x_t) & LR2(y_t, z_t)	SR Models: $\Delta y_t = \text{lagged}(\Delta y_t, \Delta x_t, \Delta z_t) - \lambda_{11}\text{LR1}_{t-1} - \lambda_{12}\text{LR2}_{t-1} + \varepsilon_{1t}$ $\Delta x_t = \text{lagged}(\Delta y_t, \Delta x_t, \Delta z_t) - \lambda_{21}\text{LR1}_{t-1} - \lambda_{22}\text{LR2}_{t-1} + \varepsilon_{2t}$ $\Delta z_t = \text{lagged}(\Delta y_t, \Delta x_t, \Delta z_t) - \lambda_{31}\text{LR1}_{t-1} - \lambda_{32}\text{LR2}_{t-1} + \varepsilon_{3t}$
Error Correction Models	Models $u_t = y_t - \alpha - \beta x_t$ as the LR_{xy} relation to be included in the SR model	(1) Estimate LR-relationship and save residuals; (2) Incorporate residuals into SR model	Note that the critical values for testing residuals are different from the usual Dickey-Fuller ones

Time Series Models

Models	Assumptions	Description	Application/Implications
General Auto-correlated Time Series	$Corr(\varepsilon_t, \varepsilon_s\|X) \neq 0$ for $t \neq s \Rightarrow$ OLS estimator inefficient and have wrong standard errors though still consistent (& unbiased)	$Cov(\varepsilon_t, \varepsilon_{t-s}\|X) = Cov(\varepsilon_t, \varepsilon_{t+s}\|X) = \sigma^2\Omega_{t,t-s} = \gamma_s$ Defining $Var(\varepsilon_t\|X) = \sigma^2\Omega_{t,t} = \gamma_0$, we obtain the autocovariance matrix $\Gamma = E(\varepsilon\varepsilon'\|X) = \sigma^2\Omega = \gamma_0\mathbf{R}$ where $\mathbf{R}$ is the autocorrelation matrix with components ρ_{ts}	Can be caused by: 1. Misspecification of the functional form of the model 2. Omitted variables 3. Measurement error of the data
GLS under Auto-correlation	GLS Estimator $\hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y)$	Est.Var$[\hat{\beta}] = \hat{\sigma}^2_\varepsilon[X'\Omega^{-1}X]^{-1}$ & $\hat{\sigma}^2_\varepsilon = \frac{(y-X\hat{\beta})'\Omega^{-1}(y-X\hat{\beta})}{T}$ & Variance of transformed disturbance u_t is $Var(\varepsilon_t - \rho\varepsilon_{t-1}) = \sigma^2_u)$	In AR(1), data is transformed such that every observation is $y_t - y_{t-1}$ and $x_t - x_{t-1}$ except the first obs (see FGLS)
FGLS	Prais- Wintsten & Cochrane- Orcutt Methods	PW assumes the first observation is $\sqrt{1-\rho^2}y_1$ & $\sqrt{1-\rho^2}\mathbf{x_1}$ while CO omits the first observation Both transforms data such that $y^i_* = y_t - \hat{\rho}y_{t-1}$ for each entry on the transformed vector $\mathbf{y}_*$ after the first observation	In each cycle, estimation uses $\hat{\rho} = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t-1}^T e_t^2}$ for the data-transformation
Newey-West Auto-correlations consistent variance estimator	Assumes that terms in the matrix with subscript pairs $\|t-s\|$ are progressively less correlated as distance between them grows	$\hat{Q}_* = \mathbf{S_0} + \frac{1}{T}\sum_{l=1}^L \sum_{t=l+1}^T w_l e_t e_{t-l}(\mathbf{x_t x'_{t-1}} + \mathbf{x_{t-1}x'_t})$ $\mathbf{S_0} = \frac{1}{T}\sum_{t=1}^T e_t^2 \mathbf{x_t x'_t}$ $w_l = 1 - \frac{l}{L+1}$, usual practice is to set $L \approx T^{1/4}$	Provides an improvement over the White Heteroscedasticity-consistent estimator by accounting for the convergence to a positive definite matrix
AR(1) Process; or any other lag length[1]	Having the disturbance follow $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ and we know $\sigma^2\Omega$	AR(1) can be written as MA(∞) with $\lambda = \rho$. For $Var(\varepsilon_t)$ to stay finite, we need to restrict $\|\rho\| < 1$ so that the data-generation is stationary $\Rightarrow Var(\varepsilon_t) = Var(\varepsilon_s) = \frac{\sigma_u^2}{1-\rho^2}$ ρ is the correlation coefficient between ε_t and ε_{t-1}	$\sigma^2\Omega = \gamma_0\mathbf{R} = \left(\frac{\sigma_u^2}{1-\rho^2}\right) \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$; note also that Trace$[\Omega]$ = n
MA(1) Process	$\varepsilon_t = u_t - \lambda u_{t-1}$	Memory in this process is only 1 period with $\gamma_0 = \sigma_u^2(1+\lambda^2)$ and $\gamma_1 = -\lambda\sigma_u^2$ but $\gamma_s = 0$	
ARCH Models (q lags of ε_t)	$y_t = x'_t\beta + \varepsilon_t$ $\varepsilon_t = u_t\sqrt{\alpha_0 + \alpha_1\varepsilon^2_{t-1}}$	$E(\varepsilon_t) = 0$ & $Var(\varepsilon_t)$ constant but $\sigma^2 = Var(\varepsilon_t\|\varepsilon_{t-1}) = 1 * (\alpha_0 + \alpha_1\varepsilon^2_{t-1})$ - heteroscedastic with clustering of large deviations If stationary, $Var(\varepsilon_t) = \frac{\alpha_0}{1-\alpha_1}, \|\alpha_1\| < 1$	LS is still BLUE; but more efficient nonlinear estimator exists: $\ln L = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_t^T \ln\alpha_0 + \alpha_1\varepsilon^2_{t-1} - \frac{1}{2}\sum_t^T \frac{\varepsilon_t^2}{\alpha_0+\alpha_1\varepsilon^2_{t-1}}$
ARCH-M Models	$y_t = x'_t\beta + \delta\sigma_t^2 + \varepsilon_t$ $\varepsilon_t\|\Psi_t \sim N(0, \sigma_t^2)$ $\sigma_t^2 = ARCH(q)$	Mean & Variance parameters no longer uncorrelated Hessian for ARCH-M not block diagonal If the ARCH part of the model is misspecified then β & δ estimates will not be consistent.	All parameters must be estimated simultaneously Under certain assumptions δ is the coefficient of relative risk aversion.
GARCH Models	$y_t = x'_t\beta + \varepsilon_t$ $\varepsilon_t\|\Psi_t \sim N(0, \sigma_t^2)$ Conditional variance evolves over time	Inclusion of conditional variance into conditional variance itself rather than regression model; Conditional variance of $Var(\varepsilon_t\|\Psi_t)$ = ARMA(p,q) $\Rightarrow \sigma_t^2 = \alpha_0 + \sum_q \alpha_q\varepsilon^2_{t-q} + \sum_p \delta_p\sigma^2_{t-p}$	Can be written as $Var(\varepsilon_t\|\Psi_t) = \gamma' z_t$ where $\gamma = [\alpha_0, \alpha_1, \cdots, \delta_1, \cdots]$ and $z_t = [1, \varepsilon^2_{t-1}, \cdots, \sigma^2_{t-1}, \cdots]$ Stationary condition: Higher moments of disturbances normal and finite

[1] Number of lags based on regressions that minimizes some Information Criteria (Akaike or Schwarz-Bayesian); $IC(p) = \ln\left(\frac{\mathbf{e'e}}{T-p_{max}-K^*}\right) + (p+K^*)\left(\frac{A^*}{T-p_{max}-K^*}\right)$ where K^* is the # parameters & A^* is 2 for Akaike and $\ln(T - p_max - K^*)$ for Bayesian & p_{max} is the largest lag considered

Discrete Choice & Time Series Hypothesis Testing

Test	Hypotheses	Test Statistic	Test Remarks
Probit Likelihood Ratio Test	H_0 = Additional variables jointly-insignificant H_1 = Added variables are jointly significant	$LR = -2(\ln L_R - \ln L_U) \sim \chi^2_q$ where R refers to the restricted regression (without the variables in question) and U refers to the unrestricted regression (including the variables);	McFadden's Pseudo $R^2 = 1 - \frac{\ln L_U}{\ln L_R}$ where L_R assumes all slope coefficients are zero
Durbin-Watson Test [Results can be inconclusive (dark areas)]	H_0 : $\rho = 0$ in AR(1) for disturbance term H_1 : $\rho \neq 0$ so there is autocorrelation OR H_1 : $\rho > 0$ so there is positive autocorrelation	$\mathbf{d} = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = 2\left(1 - \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T-1} e_t^2}\right) - \frac{e_t^2 - e_{t-1}^2}{\sum_{t=1}^{T-1} e_t^2} \Rightarrow \mathbf{d} \approx 2(1-r)$ where $r = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T-1} e_t^2}$ 	Critical values d_L & d_U from DW-tables. **Cannot** be used when there are lagged dependent variables in the specification
Durbin h-test (only for 1 lagged dependent)	H_0: No Autocorrelation H_1: Presence of autocorrelation	$h = r\sqrt{T/(1 - Ts_c^2)}$ where s_c^2 is the estimated variance of the least squares regression coefficient on the lagged variable (y_{t-1})	Cannot be computed if $s_c^2 > \frac{1}{T}$
Durbin Test with P lags	H_0: No Autocorrelation H_1: Presence of autocorrelation	F_D = F-stat of the joint-significance of P lags of the residuals of a regression of e_t on all x, y_{t-1}, y_{t-2}, $\cdots$, y_{t-R}, e_{t-1}, e_{t-2}, $\cdots$, e_{t-P}	Reject H_0 if $F_D > F_{P,T-K-P}$ where K is # explanatory variables
Godfrey-Breusch LM Test	H_0: No autocorrelation in the disturbance against the alternative hypothesis that there disturbance has either AR(p) or MA(p) process H_1: ε_t = AR(p) or ε_t = MA(p)	Run the same auxillary regression as Durbin Test with P lags to obtain the R^2, LM Stat $= TR^2$ where T is the sample size; follows χ^2_P if H_0 is true. The test controls for intervening effects of independent variables	In matrix form, $LM = T\left(\frac{e'X_0(X_0'X_0)^{-1}X_0 e}{e'e}\right) = TR_0^2$
Q-Statistic	H_0: No Autocorrelation H_1: Disturbance not well-behaved (less powerful than LM Test when H_0 false since it doesn't condition on $\mathbf{x_t}$)	Box-Pierce Version: $Q = T\sum_{j=1}^{P} r_j^2$ Ljung Modified Version: $Q' = T(T+2)\sum_{j=1}^{P}\frac{r_j^2}{T-j}$; both follow χ^2_P if H_0 is true	If H_0 keeps getting rejected then we have to correct for the autocorrelation
Testing ARCH	H_0: ARCH(0) or $\alpha_q = 0\ \forall q$ H_1: ARCH(q)	Save residuals e_t & then regress $e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \cdots + \alpha_q e_{t-q}^2$; LM Stat $= TR^2 \sim \chi^2_q$ where R^2 is from the auxillary regression	F-test of joint significance can be used with F-stat $\sim F_{q,T-q-1}$
Testing Unit Roots	Usual manual tests: $X_t = \alpha + \rho X_{t-1} + \varepsilon_t$ $H_0 : \rho = 1$ $H_1 : \|\rho\| < 1$ STATA method: $\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t$ $(\theta = \rho - 1)$ $H_0 : \theta = 0$ $H_1 : \theta < 0$	Running a regular OLS regression but applying special critical values based on Dickey-Fuller's Monte Carlo results. Can be tested with 3 variations: (1) No intercept or trend; (2) With intercept no trend; (3) With both intercept & trend STATA options includes 'noconstant', 'drift' (constant no trend), 'trend' (both)	STATA command is 'dfuller *varlist*, *options* regress lags(p)';
Johansen Test for Cointegration	H_0: $r = 0$ H_1: $r \neq 0$, where r is the rank of the matrix (of the SR models)	Test runs to check if the null hypothsis is rejected and stops once it is not rejected	STATA command is: 'vecrank *varlist*, lags(p)'