Assumptions of OLS: **MR1:** $y = X\beta + e$. **MR2:** $E[e|X] = 0 \rightarrow Cov(e, X) = 0$ **MR3:** $E[ee'|X] = \sigma^2 I_T$ (Homoscedasticity and No Autocorrelation). **MR4:** Non-stochastic Matrix X. **MR5:** Rank(x) = k $\subset$ T. **MR6:** $f(e|X) \sim N(0, \sigma^2)$(Optional). (**MR1-MR5:** Gauss-Markov Theorem: smallest variance of linear & unbiased estimators i.e. BLUE). **Var (e_t) or (σ^2)** $= E[e_t - E(e_t)]^2 = \Sigma\, \hat{e}_t^2$ (or $\hat{e}'\hat{e}$)/T **(biased) Or** $= \Sigma\, \hat{e}_t^2$ (or $\hat{e}'\hat{e}$)/(T-k) **(unbiased) Or** = SSE/(T-k). **Var(b)** $= E[(b-\beta)(b-\beta)'] = \sigma^2(XX')^{-1}$.; **(b- β)** $= (XX')^{-1}X'e$; e = error term. **If** $e \sim N(0, \sigma^2 I)$ then, $b \sim N[\beta, \sigma^2(XX')^{-1}]$. **Idempotent Matrix:** $[I - X\,X'(XX')^{-1}] = M$ (symmetric and Idempotent); MM = M and MX = 0.

Use of T-dist. instead of N-dist. for hypothesis test: **Derivation of t-stat:** we can transform normal random variable (b_k) into standard normal variable Z, $Z = (b_k - \beta_k)/\sqrt{var(b_k)} \sim N[0,1]$; k = 1, 2, …K. **When** we replace σ^2 by estimator $\hat{\sigma}^2$, then, $t = (b_k - \beta_k)/\sqrt{\widehat{var}(b_k)} \sim t_{(t-k)} \rightarrow t = (b_k - \beta_k)/se(b_k) \sim t_{(t-k)}$. **Chi-square Dist.** = $(SNV)^2$. If $Z_1, Z_2 \ldots Z_m$ are m independent $N(0,1)$ RVs, then $V = (Z_1)^2 + (Z_1)^2 + \ldots + (Z_m)^2 \sim \chi^2_{(m)}$. $\rightarrow V \sim \chi^2_{(m)}$. $V = \Sigma\, \hat{e}_t^2/\sigma^2 = (T-k)\hat{\sigma}^2/\sigma^2$; with T-k df. (All T residuals $\hat{e} = y - Xb$ depends upon least square estimators b_k & T-k of the least squares residuals are independent. $V = (T-k)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(T-k)}$. **If** $Z \sim N[0,1]$ and $V \sim \chi^2_{(m)}$, and Z and V are independent, then $t = Z/\sqrt{(V/m)} \sim t_{(m)}$. The t-dist. shape is determined by df parameter m i.e. $t_{(m)}$. $\mathbf{t} = Z/\sqrt{[V/(T-k)]} = [(b_k - \beta_k)/\sqrt{(\sigma^2 c_{kk})}]/\sqrt{[\{(T-k)\,\hat{\sigma}^2/\sigma^2\}/(T-k)]} = (b_k - \beta_k)/se(b_k) \sim t_{(t-k)}$; c_{kk} is the k_{th} element of cofactor matrix. $\rightarrow$ **T-stat** $= (b_k - \beta_k)/se(b_k)$; $se(b_k) = \sqrt{var(b_k)}$.

Standardized normal RV?: Normal Distribution: $X \sim N(\beta, \sigma^2)$. **Prob fn:** $f(x) = \exp[-(x-\beta)^2/(2\sigma^2)] * [1/\sqrt{(2\pi\sigma^2)}]$ Probability of normal random variables given by area of pdf and hard to compute $\rightarrow$ Standardize NRV! **If** $X \sim N(\beta, \sigma^2)$, $\rightarrow Z = (x-\beta)/\sigma$; $Z \sim N(0, 1)$ then, $f(z) = \exp[-Z^2/2]/\sqrt{(2\pi)}$.

Compute probabilities for normal random variable:

$P[X \geq a] = P[\{(x-\beta)/\sigma\} \geq \{(a-\beta)/\sigma\}] \rightarrow P[Z \geq \{(a-\beta)/\sigma\}]$.

Confidence Interval $= \beta_k \pm t_{cric} * se(b_k)$; $t_{0.05/2} = 1.96$.

CI for Prediction (forecast) Error: $y = X\beta + e$; $(e_{t*1}) \sim N(0_{T*1}, \sigma^2 I_T)$ & $y_o = X'_o\beta + e_o \sim N(0_{(T+1)*1}, \sigma^2 I_{T+1})$. Let, point forecast $\tilde{y}_o = X'_o b$ is BLU Forecast of y_o. Then, $\tilde{y}_o$ is normal, and var of $(\tilde{y}_o - y_o) = X'_o b - X'_o\beta + e_o = X'_o(b-\beta) - e_o = \sigma^2 X'_o(XX')^{-1}X_o + \sigma^2 = \sigma^2[X'_o(XX')^{-1}X_o + 1]$. **1.** $Z = (\tilde{y}_o - y_o)/$ var of $(\tilde{y}_o - y_o)$. **2.** $V = (T-k)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(T-k)}$. **3.** $t = Z/\sqrt{[V/(T-k)]}$ (solve) **4. Construct CI for** $\mathbf{y_o = \tilde{y}_o \pm t_{\alpha/2(T-k)}\sqrt{[\hat{\sigma}^2[X'_o(XX')^{-1}X_o + 1]]}}$.

Coefficient of Determination (R^2)

$y_t - \bar{y} = b_2(x_t - \bar{x}) = (\hat{e}_t - \bar{\hat{e}}) = b_2(x_t - \bar{x}) + \hat{e}_t = \tilde{y}_t - \bar{y} + \hat{e}_t$

Propty: $R^2 = 1 \rightarrow SSE = 0$; $R^2 = 1 \rightarrow SSR = 0$; $1 \geq R^2 \geq 0$.

SST = SSR + SSE $\rightarrow \Sigma(y_t - \bar{y})^2 = (b_2)^2\,\Sigma(x_t - \bar{x})^2 + \Sigma(\hat{e}_t)^2$

$\Sigma(y_t - \bar{y})^2$ **(SST)** $= \Sigma(\tilde{y}_t - \bar{y})^2$ **(SSR)** $+ \Sigma(\hat{e}_t)^2$ **(SSE)**.

SSR $= \sqrt{[SSE/(T-K)]}$

$\mathbf{R^2} = SSR/SST = \Sigma(\tilde{y}_t - \bar{y})^2/\Sigma(y_t - \bar{y})^2$ **Or**

$\mathbf{R^2} = 1 - SSE/SST = 1 - \Sigma(\hat{e}_t)^2/\Sigma(y_t - \bar{y})^2$

Centered $\mathbf{R^2} = \Sigma(\tilde{y}_t - \bar{y})^2/\Sigma(y_t - \bar{y})^2$

Uncentered $\mathbf{R^2}$: $\Sigma_t(\tilde{y}_t)^2/\Sigma_t(y_t)^2$

Adj $\mathbf{R^2} = 1 - [SSE/(T-K)]/[SST/(T-1)]$

Adj $\mathbf{R^2} = 1 - [\hat{\sigma}^2/\{\Sigma(y_t - \bar{y})^2/(T-1)\}]$ **Or**

Adj $\mathbf{R^2} = 1 - [(1-R^2)(T-1)/(T-k)]$

AIC (Akaike Info. Criterion) $= \ln(\hat{e}\,\hat{e}'/T) + (2k/T)$

SC (Schwarz Criterion) $= \ln(\hat{e}\,\hat{e}'/T) + (k/T)\ln T$

F-value $= [\hat{e}_r\,\hat{e}'_r/s]/[\hat{e}\,\hat{e}'/(T-k)]$

F-value $= [(SSE_R - SSE_U)/S]/[SSE_U/(T-k)] \sim F_{(s, T-k)}$.

F-value $= [SSR/(k-1)]/[SSE(T-k)] \sim F_{(k-1, T-k)}$.

F-value $= [R^2/(k-1)]/[(1-R^2)/(T-k)] \sim F_{(k-1, T-k)}$.

Find St. Err. (& R^2s) when XX', X'y, $\Sigma(y_t)^2$ are given: **St. Er.: 1.** N = A_{11} of XX'. **2.** SSE = $\hat{e}\,\hat{e}'$ [= $\Sigma(y_t)^2$] - X'y (= b' hatt)* XX'. **3.** $\hat{\sigma}^2 = I\,\hat{e}\,\hat{e}'/(T-k) = SSE/(T-K)$ **4.** $Var(b) = \hat{\sigma}^2(X'X)^{-1}$ (first, find inverse matrix of XX'). **5.** The diagonals $\rightarrow$ co**var**iance of b_0, b_1, b_2. **6.** St Er = $\sqrt{\text{Diagonals}}$. **R^2 & Adj-R^2: 1.** Find SSE (above). **2.** $SST = \Sigma(y_t - \bar{y})^2 \rightarrow \Sigma(y_t)^2 + 2y_t * \bar{y} + \bar{y}^2 * N$; [$\bar{y} = \Sigma(y_t)/N$] [$\Sigma(y_t) = A_{11}$ of X'y]. **3.** Use R^2s fmla.

Seasonal Adjustments (Deseasonalizing): 1. Regress y on D: $y = Db + y^s \rightarrow y^s = y - Db$ **2.** Regress y on D and Trend: $y = Db + Pc$ (trend term) $+ e \rightarrow y^T = y - Db$. T = # of years.

Chow Test: Use two sub-sample or dummy variables: Two Methods. First: Construct restricted model (main effects only) and unrestricted model (main and interaction effects), obtain SSE_R, SSE_R, and do F-test. Second: **1.** Divide full sample into two sub-sample **2.** Estimate model with each sub-sample to get SSE_1 & SSE_2; $SSE_U = SSE_1 + SSE_2$ **3.** Estimate restricted mode (as above) with full sample to get SSE_R **4.** Do f-test. $F = [(SSE_R - SSE_U)/S]/[SSEU/(T-2k)] \sim F_{(s, T1 + T2 - 2s)}$.; S = # of parameters.

Multi-collinearity & Model Specification Error (Presence/Absence, Severity & Nature): The data from uncontrolled experiment makes economic variables move together in systematic way. These variables are *collinear* variables and problem is *collinearity* or *multi-collinearity*. **Consequences: 1.** Exact linear relationship between explanatory variables$\rightarrow$ least squares estimator is not defined. **2.** Variance, Standard Error and Covariance large $\rightarrow$ hypothesis test from t & F-tests are likely NOT reject H_o despite high R^2 & F-value (Type II Error).

3. Estimates may be sensitive to addition or deletion of a few observations (but accurate forecast is still possible). **4.** Affects efficiency but not predictability and still remain unbiased.

Identification: 1. Sample correlation coefficient **2.** Determinants of XX' **3.** Auxiliary Regression (test for each IVs) **4.** Variance Infatuation Factor (VIF). **Remedy:** 1. Obtain more info 2) Add structure to the problem by introducing non-sample information (prior) **3.** Drop problematic variable and find proxy variable. **4.** Ridge regression (increase diagonal element of XX' matrix leaving off diagonal elements unchanged but biased, and variance are small. Problem in interpretation of estimates).

Specification Error: 1. Choice of functional form: Variables can be transformed in any convenient way as long as OLS assumptions met and functional form fits data. (Parameter in linear form). **A.** Linear Model: $y = \beta_1 + \beta_2 x + e$: β_2 (Slope/Marginal Effect) $\beta_2(x/y)$ (Elasticity) **B.** Reciprocal Model: $y = \beta_1 + \beta_2(1/x) + e$: $\boldsymbol{\delta}y/\boldsymbol{\delta}x = -\beta_2(1/x^2)$ (slope) & $(\boldsymbol{\delta}y/\boldsymbol{\delta}x)*(x/y) = -\beta_2(1/xy)$ (Elasticity). **C.** Log-log Model: $Log(y) = \beta_1 + \beta_2 log(x) + e$: $\boldsymbol{\delta}y/\boldsymbol{\delta}x = \beta_2(y/x)$ (slope) & $(\boldsymbol{\delta}y/\boldsymbol{\delta}x)*(x/y) = \beta_2$ (Elasticity) **D.** Log-linear (exponential) Model: $log(y) = \beta_1 + \beta_2 x + e$: $\boldsymbol{\delta}y/\boldsymbol{\delta}x = \beta_2 y$ (slope) & $(\boldsymbol{\delta}y/\boldsymbol{\delta}x)*(x/y) = \beta_2 x$ (Elasticity) **E.** Linear-log (semi-log) model: $y = \beta_1 + \beta_2 log(x) + e$ $\beta_2(1/x)$ (Slope) & $\beta_2(1/y)$ (Elasticity) **2.** Choice of regressors/IVs: **A. Test of Model Misspecification:** Kolmogorov's nonlinearity test: Include all variables and their interactions in regression and test H_o: interactions are zero using F-test. H_a: at least one of them are different. **B.** Regression Specification Error Test (Ramsey RESET Test) **3.** Meet assumptions of OLS (MR1-MR6) or not.

Consequences of Omitted Variables: 1. OLS estimators for β's and σ^2 (variance) will be biased unless each omitted variable is uncorrelated with the included variables (omitting intercept, OLS estimators for estimated variables are biased unless variables have zero means).Variance of OLS estimator is small and biased upward (low efficiency) but estimated variance might not necessarily be small. **Consequences of including Irrelevant Variables:** OLS estimators and variance are unbiased. Variance of OLS estimators and estimated variance are larger than those from true model.

Scaling the Data: Scaling data does not affect the underlying relationship (i.e. elasticity) but the interpretation of coefficient estimates and some summary measures are affected. **Scaling x:** the standard error of regression coefficient changes with same multiplicative factor as the coefficient. Thus their ratio, t-stat does not change. **Scaling y:** the OLS residuals will also be scaled due to scaling in error term and affect the standard error of the regression coefficients but does not affect t-stat and R^2. **Scaling x & y by same factor:** Intercept, residuals and Interpretations of parameters changes.

Heteroscedasticity: (violation of MR3): $y = X\beta + e$; $e \sim N(0, \sigma^2\Omega_T)$ Regression model consistently and accurately predicts the lower values of the DV but highly inconsistent and inaccurate when it predicts high values or vice versa. **Consequences:** The variance of standard error is not equal; least square is still unbiased and consistent but inefficient (variance unequal, inaccurate predictions); the standard error for least square estimates is incorrect, the hypothesis test and confidence intervals based on standard errors are incorrect; increase type I error.

Homoscedasticity of ERROR TERMS determines whether a regression model is consistent and accurate across all the value of DV and its prediction. **Some Additional Notes: The unbiased estimator E(b)=β and unbiasedness** is small sample property. It has nothing to do with large sample. **The R^2 and Adj R^2** only measures linear relationship between dependent and independent variables. **T-distribution or F-distribution**: to test single hypothesis test. **F-distribution:** to test multiple hypothesis.

Testing Heteroscedasticity: **1.** Residual Plots: Pattern = Heteroscedasticity. 2) Goldfeld-Quandt (GQ) Test: Split data ~ equally, run separate regression and estimate variance for both regression and divide GQ = $(\sigma_1^2/\sigma_2^2) > 1$. If GQ > $F_{((T1-K),\ (T2-K))}$, Reject H_o. **2.** The Lagrange Multiplier Test: Run regression model, compute variance regression, and compute residual, Run auxiliary regression: a. Breusch-Pagan test b. Glesjer Test c. Haravey-Godfrey test, compute TS: **LM** = T*R^2 (R^2 from variance equation) H_o: parameter estimates are equal. Reject H_o if LM > $\chi^2_{(p-1)}$. **3.** White Test: Run the regression, compute residuals, regress the squared residuals on squares and interactions of variables, compute TS: LM = T*R^2 (R^2 from variance equation) H_o: parameter estimates are equal. Reject H_o if LM > $\chi^2_{(p-1)}$. Problem: large number of variables gives problem due to df, reduce power of test resulting homoscedasticity. **4. Asymptotic test statistics:** **a.** Likelihood Ratio (LR) Test LR = max L(Θ) rest / max L(Θ) unrest; LR TS = -2lnλ = 2[ln(max L(Θ) rest)-ln(max L(Θ) unrest)]→$\chi^2_{(q = \#\text{ of hypo.})}$ OR LR = T(ln(ee)$_{rest}$-ln(ee)$_{unrest}$)→ $\chi^2_{(q = \#\text{ of hypo.})}$ **b.** Wald Test W = T(SSE$_R$-SSE$_U$)/SSE$_U$→ $\chi^2_{(q = \#\text{ of hypo.})}$ **c.** Lagrange Multiplier Test LM = T* R^2 (from regression of e_{rest} on X) = $T_{(=\#\text{ of obs.})}$*(SSE$_R$-SSE$_U$)/SSE$_U$→ $\chi^2_{(q = \#\text{ of hypo.})}$.

W ≥ LR ≥ LM. When to use: W→ small sample, linear hypothesis and linear model, LR or LM→ nonlinear model, nonlinear hypothesis. Eg. Multiplicative heteroscedasticity test → LR or LM.

Estimating under Heteroscedasticity: **1.** White variance covariance matrix (HCC) **2.** GLS when V is known: find hetero-variable, scale by $\sqrt{(X_t)}$ including intercept, run OLS with transformed variables. **3.** GLS when unknown V (FGLS): Estimate V & run GLS. Var(e_t*) = var $(e/\sqrt{x_t})$ = $1/x_t \text{var}(e_t)$ = $\sigma^2 * 1/x\ \sigma^2 = \sigma^2$.

Autocorrelation: (violation of MR4) AR Nature:

More formal description of the AR problem for estimation

Covariance:

$$\gamma_s = E(e_t, e_{t-s}), \quad s = 0, \pm 1, \pm 2, \ldots; E(e_t) = 0$$
$$\gamma_0 = E(e_t^2) = \sigma_e^2$$

$$\text{var}(e) = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{bmatrix} = \sigma_e^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{bmatrix} = \sigma_e^2 \Omega$$

Correlation coefficient:

$$\rho_s = \frac{\text{cov}(e_t e_{t+s})}{\sqrt{\text{var}(e_t)\,\text{var}(e_{t+s})}} = \frac{\gamma_s}{\gamma_0}$$

ρ_s = cov($e_t e_{t+s}$)/√((var(e_t) var(e_{t+s})). Some form of AR Structure: **a.** Ist order autoregressive process AR(1): $e_t = \lambda e_{t-1} + u_t$; $|\lambda| < 1$, $u_t \sim$ iid $(0, \sigma_u^2)$. **b.** AR(2): $e_t = \lambda e_{t-1} + \lambda e_{t-2} + u_t$; $|\lambda| < 1$, $u_t \sim$ iid $(0, \sigma_u^2)$. **c.** AR(P): $e_t = \lambda e_{t-1} + \lambda e_{t-2} + \ldots + \lambda_p e_{t-p}\ u_t$; $|\lambda| < 1$, $u_t \sim$ iid $(0, \sigma_u^2)$. **d.** Moving Average (MA) Process: Ist order MA: MA(1) $e_t = \theta u_{t-1} + u_t$; $u_t \sim$ iid $(0, \sigma_u^2)$. **Consequences:** **1.** Least square estimator is still linear and unbiased and consistent but not best/smallest. **2.** Standard errors from OLS are wrong and confidence interval, hypothesis test are wrong. **Proof (right):**

Proof

Consider the model,

$$y_t = \beta x_t + e_t; \quad e_t = \rho e_{t-1} + u_t; \quad e \sim (0, \sigma_e^2 \Omega_\rho) \text{ and } u \sim (0, \sigma_u^2 I)$$

- $\mathbf{b} = \beta + (\mathbf{X'X})^{-1}\mathbf{X'}e \quad \Rightarrow E(\mathbf{b}) = \beta + (\mathbf{X'X})^{-1}\mathbf{X'}E(e) = \beta$
- $\text{VAR}[\mathbf{b}] = \text{E}[(\mathbf{b} - \beta)(\mathbf{b} - \beta)']$
  $= \text{E}[((\mathbf{X'X})^{-1}\mathbf{X'}e)((\mathbf{X'X})^{-1}\mathbf{X'}e)']$
  $= \text{E}[(\mathbf{X'X})^{-1}\mathbf{X'}ee'\mathbf{X}(\mathbf{X'X})^{-1}]$
  $= (\mathbf{X'X})^{-1}\mathbf{X'}\text{E}[ee']\mathbf{X}(\mathbf{X'X})^{-1}].$

$$\text{VAR[b]} = (X'X)^{-1}X'(\sigma_e^2\Omega_\rho)X(X'X)^{-1}$$
$$= \sigma_e^2 (X'X)^{-1}(X'\Omega_\rho X)(X'X)^{-1}$$

GLS Procedure: **1.** When Ω (ρ) is Known:

Consider the model,

$$y_t = \beta x_t + e_t; \quad e_t = \rho e_{t-1} + u_t; \quad e \sim (0, \sigma_e^2 \Omega_\rho) \text{ and } u \sim (0, \sigma_u^2 I)$$

$$\text{var}(e) = \sigma_e^2 \Omega_\rho = \frac{\sigma_u^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix} = \sigma_u^2 \Omega$$

- $b_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$
- With non-singular matrix P, $\Omega^{-1} = P'P$

$$b_{GLS} = (X'P'PX)^{-1}X'P'Py = [(PX)'PX]^{-1}(PX)'(Py)$$

- $y_* = X_*\beta + e_*; \quad y_* = Py, \quad X_* = PX, \quad e_* = Pe$
  $\text{var}(e_*) = E(Pee'P') = \sigma^2 P\Omega P'$
  $= \sigma^2 PP^{-1}P'^{-1}P' = \sigma^2 I; \quad \Omega = P^{-1}P'^{-1}$
- $b_{GLS} = (X_*'X_*)^{-1}X_*'y_*$
  $\text{var}(b_{GLS}) = \sigma^2(X_*'X_*)^{-1}$
- Both b_{ML} and b_{GLS} can be used for hypothesis tests because the estimates were based on asymptotically or finite normal assumption for disturbance terms

How does the matrix "P" look like?

With non-singular matrix P, $\Omega^{-1} = P'P$

$$b_{GLS} = (X'P'PX)^{-1}X'P'Py = [(PX)'PX]^{-1}(PX)'(Py)$$

$$\text{var}(e) = \sigma_e^2 \Omega_\rho = \frac{\sigma_u^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix} = \sigma_u^2 \Omega;$$

$$\Omega = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix}; \quad \Omega^{-1} = \begin{bmatrix} 1 & -\rho & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 \\ 0 & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & -\rho & 1 \end{bmatrix} = P'P$$

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 & 0 \\ 0 & -\rho & 1 & 0 & 0 \\ 0 & 0 & -\rho & 1 & 0 \\ 0 & 0 & 0 & -\rho & 1 \end{bmatrix} \quad \Leftarrow \quad \Omega^{-1} = P'P \quad \text{"Choleski factorization"}$$

$$b = (X'P'PX)^{-1}X'P'Py = [(PX)'PX]^{-1}(PX)'(Py)$$

The model with transformed data $\quad y_t^* = \alpha + \beta x_t^* + e_t^*$

$$\begin{bmatrix} \sqrt{1-\rho^2}y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} = \alpha \begin{bmatrix} \sqrt{1-\rho^2} \\ 1-\rho \\ 1-\rho \\ \vdots \\ 1-\rho \end{bmatrix} + \begin{bmatrix} \sqrt{1-\rho^2}x_1 \\ x_2 - \rho x_1 \\ x_3 - \rho x_2 \\ \vdots \\ x_T - \rho x_{T-1} \end{bmatrix} + \begin{bmatrix} \sqrt{1-\rho^2}e_1 \\ e_2 - \rho e_1 \\ e_3 - \rho e_2 \\ \vdots \\ e_T - \rho e_{T-1} \end{bmatrix}$$

Ignoring the first obs. produces "Cochrane Orcutt method"

$$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + e_t - \rho e_{t-1}; \quad u_t = e_t - \rho e_{t-1}.$$

Recovering first observation (Prais-Winsten): Dropping the first observation and applying OLS is not the BLUE method (limitation of Co. & Or.): lost efficiency (variance of error associated with First obs. is not equal to that of others).

Transforming the First Observation:

Transform the First Observation

See matrix P!!

The first observation in the regression model is

$$y_1 = \beta_1 + x_1\beta_2 + e_1 \quad \text{with} \quad \text{var}(e_1) = \sigma_e^2 = \sigma_v^2/(1-\rho^2) \qquad (8)$$

- To yield an error variance of σ_v^2, multiply $\sqrt{1-\rho^2}$

$$\sqrt{1-\rho^2}y_1 = \sqrt{1-\rho^2}\beta_1 + \sqrt{1-\rho^2}x_1\beta_2 + \sqrt{1-\rho^2}e_1 \qquad (9)$$

or

$$y_1^* = x_{11}^*\beta_1 + x_{12}^*\beta_2 + e_1^*$$

where

$$y_1^* = \sqrt{1-\rho^2}y_1 \quad x_{11}^* = \sqrt{1-\rho^2}$$
$$x_{12}^* = \sqrt{1-\rho^2}x_1 \quad e_1^* = \sqrt{1-\rho^2}e_1$$

Providing ρ is known, we can find the best linear unbiased estimator for β_1 and β_2 by applying least squares to the transformed model

$$y_t^* = \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + v_t$$

where the transformed variables are defined by

$$y_1^* = \sqrt{1-\rho^2}y_1 \qquad x_{11}^* = \sqrt{1-\rho^2} \qquad x_{12}^* = \sqrt{1-\rho^2}x_1$$

for the first observation, and

$$y_t^* = y_t - \rho y_{t-1} \qquad x_{t1}^* = 1-\rho \qquad x_{t2}^* = x_t - \rho x_{t-1}$$

for the remaining $t = 2, 3, \ldots, T$ observations.

2. GLS Procedure when Ω (ρ) is Unknown: FGLS:

FGLS procedure for AR problem

1. Estimate ρ:

$$\hat{e}_t = \rho\hat{e}_{t-1} + \hat{v}_t \qquad \hat{\rho} = \frac{\sum_{t=2}^{T}\hat{e}_t\hat{e}_{t-1}}{\sum_{t=2}^{T}\hat{e}_{t-1}^2}$$

2. Transform the variables:

$$y_1^* = \sqrt{1-\hat{\rho}^2}y_1, \quad x_{11}^* = \sqrt{1-\hat{\rho}^2}, \quad x_{12}^* = \sqrt{1-\hat{\rho}^2}x_1$$

$$y_t^* = y_t - \hat{\rho}y_{t-1}, \quad x_{1t}^* = 1-\hat{\rho}, \quad x_{2t}^* = x_{2t} - \hat{\rho}x_{2t-1}; \quad t = 2,3,\ldots,T$$

3. Run LS with the transformed data

MLE procedure for AR problem

$$\ln L = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\ln|V| - \frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)$$

$$V = \sigma_v^2\Omega = \frac{\sigma_v^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix}$$

1. Estimate ρ and σ_v^2 : $\hat{e}_t = \rho\hat{e}_{t-1} + \hat{v}_t$
2. Run MLE
3. Repeat 1 and 2 iteratively until estimates converge

Note: The OLS estimator of ρ in step 1 has good statistical properties if sample size (T) is large.

Test Autocorrelation: 1. The Durbin Watson Test:

The Durbin-Watson test is by far the most important one for detecting AR(1) errors.

- Consider again the linear regression model

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

where $e_t = \rho e_{t-1} + v_t$

- Assuming the v_t are independent random errors with distribution $N(0, \sigma_v^2)$ => DW.
- For a null hypothesis of no autocorrelation: $H_0: \rho = 0$.
- For an alternative hypothesis: $H_1: \rho > 0$ or $H_1: \rho < 0$ or $H_1: \rho \neq 0$.

- In most empirical applications in economics, we choose $H_1: \rho > 0$ because positive autocorrelation is the most likely form that autocorrelation will take.

$$H_0: \rho = 0 \quad \text{against} \quad H_1: \rho > 0$$

- The DW statistic is

$$d = \frac{\sum_{t=2}^{T}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{T}\hat{e}_t^2} \qquad \hat{\rho} = \frac{\sum_{t=2}^{T}\hat{e}_t\hat{e}_{t-1}}{\sum_{t=2}^{T}\hat{e}_{t-1}^2}$$

where the $\hat{e}_t$ are the least squares residuals $\hat{e}_t = y_t - b_1 - b_2 x_t$.

- Why can't we test for $\hat{\rho}$ rather than using DW statistic d?

Relationship between d and $\hat{\rho}$

$$d = \frac{\sum_{t=2}^{T}\hat{e}_t^2 + \sum_{t=2}^{T}\hat{e}_{t-1}^2 - 2\sum_{t=2}^{T}\hat{e}_t\hat{e}_{t-1}}{\sum_{t=1}^{T}\hat{e}_t^2} = \frac{\sum_{t=2}^{T}\hat{e}_t^2}{\sum_{t=1}^{T}\hat{e}_t^2} + \frac{\sum_{t=2}^{T}\hat{e}_{t-1}^2}{\sum_{t=1}^{T}\hat{e}_t^2} - 2\frac{\sum_{t=2}^{T}\hat{e}_t\hat{e}_{t-1}}{\sum_{t=1}^{T}\hat{e}_t^2} \approx 1 + 1 - 2\hat{\rho}$$

The Durbin-Watson statistic: $d \approx 2(1-\hat{\rho})$

- If $\hat{\rho} = 0$, then the Durbin-Watson statistic $d \approx 2$: the model errors are not autocorrelated.
- If $\hat{\rho} = 1$ then $d \approx 0$: the model errors are correlated, and $\rho > 0$.
- If $\hat{\rho} = -1$ then $d \approx 4$: the model errors are correlated, and $\rho < 0$.

- In most empirical applications in economics, we choose $H_1: \rho > 0$ because positive autocorrelation is the most likely form that autocorrelation will take.

$$H_0: \rho = 0 \quad \text{against} \quad H_1: \rho > 0$$

- The DW statistic is

$$d = \frac{\sum_{t=2}^{T}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{T}\hat{e}_t^2} \qquad \hat{\rho} = \frac{\sum_{t=2}^{T}\hat{e}_t\hat{e}_{t-1}}{\sum_{t=2}^{T}\hat{e}_{t-1}^2}$$

where the $\hat{e}_t$ are the least squares residuals $\hat{e}_t = y_t - b_1 - b_2 x_t$.

- Why can't we test for $\hat{\rho}$ rather than using DW statistic d?

How close d to be zero can be considered as "AR problem"?

$H_1: \rho > 0$ | $H_1: \rho < 0$

Reject $\rho = 0$	Inconclusive	Not Reject $\rho = 0$	Inconclusive	Reject $\rho = 0$

$d = 0$, $\rho = 1$ — d_L — d_U — 2, $\rho = 0$ — $4-d_U$ — $4-d_L$ — 4, $\rho = -1$

When d < 2
Test $H_0: \rho = 0$ against $H_1: \rho > 0$: Reject if $d \le d_L$. Cannot reject if $d \ge d_U$

When d > 2
Test $H_0: \rho = 0$ against $H_1: \rho < 0$: Reject if $d \ge 4-d_L$. Cannot reject if $d \le 4-d_U$

What if d is in inconclusive region?

Breusch-Godfrey Test - A Lagrange Multiplier Test

(1) $y_t = \beta_1 + \beta_2 x_t + e_t$;
(2) $e_t = \beta_3 e_{t-1} + u_t$; $\quad u \sim iidN(0, \sigma_u^2 I), |\beta_3| < 1$

Rewriting (1) gives $e_{t-1} = y_{t-1} - \beta_1 - \beta_2 x_{t-1}$

Substituting the new equation back to (1) and (2) gives

$$y_t = \beta_1 + \beta_2 x_t + \beta_3(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + u_t$$
$$= \beta_1(1-\beta_3) + \beta_2 x_t + \beta_3 y_{t-1} - \beta_2\beta_3 x_{t-1} + u_t$$

When $\beta_3 = 0$, this is reduced to (1) with the Gaussian noise => LM test

6. Prediction With AR(1) Errors

For the problem of forecasting or predicting a future observation y_0 that we assume

$$y_0 = \beta_1 + x_0\beta_2 + e_0$$

where x_0 is a given future value of an explanatory variable and e_0 is a future error term.

Two things to remember when predicting with AR (1) errors:

1. When the errors are autocorrelated, the generalized least squares estimators, denoted by $\hat{\beta}_1$ and $\hat{\beta}_2$, are more precise than their least squares counterparts b_1 and b_2.
2. When e_0 is correlated with past errors, need to consider future error terms to improve forecast.

- When we are predicting one period into the future, the model with an AR(1) error can be written as

$$y_{T+1} = \beta_1 + \beta_2 x_{T+1} + e_{T+1}$$
$$= \beta_1 + \beta_2 x_{T+1} + \rho e_T + v_{T+1}$$

where we have used $e_{T+1} = \rho e_T + v_{T+1}$

Using GLS estimators, $\hat{y}_{T+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+1} + \hat{\rho}\tilde{e}_T$

$$\hat{y}_{T+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+1} + \hat{\rho}\tilde{e}_T$$

Where do we obtain ρ and e_T?

1. For ρ we can use the estimator $\hat{\rho}$ used for GLS procedure.
2. To estimate e_T we use the GLS residual, defined as

$$\tilde{e}_T = y_T - \hat{\beta}_1 - \hat{\beta}_2 x_T$$

What about predicting more than one period into the future?

For h periods ahead, $\hat{y}_{T+h} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+h} + \hat{\rho}^h\tilde{e}_T$

Assuming $|\hat{\rho}| < 1$, the influence of the term $\hat{\rho}^h\tilde{e}_T$ diminishes the further we go into the future (the larger h becomes).

An example with sweet corn data

$$\hat{\beta}_1 = 6.1641 \quad \hat{\beta}_2 = 1.0066 \quad \hat{\rho} = 0.342$$

and

$$\tilde{e}_T = y_T - \hat{\beta}_1 - \hat{\beta}_2 x_T = \ln(A_T) - \hat{\beta}_1 - \hat{\beta}_2\ln(P_T) = 5.4596 - 6.1641 - 1.0066(-0.9374) = 0.239$$

To predict y_{T+1} and y_{T+2} for a sweet corn price of 0.4,

$$\hat{y}_{T+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+1} + \hat{\rho}\tilde{e}_T = 6.1641 + 1.0066\ln(0.4) + (0.342)(0.239) = 5.3235$$

$$\hat{y}_{T+2} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+1} + \hat{\rho}^2\tilde{e}_T = 6.1641 + 1.0066\ln(0.4) + (0.342)^2(0.239) = 5.2697$$

__**Simultaneous Equation Models:**__ considers econometric models for data that are jointly determined by two or more economic relations. In each models there are two or more dependent variables and have set of equations. Eg. **Supply and Demand Model** (takes two (supply: Q = $\beta_1 P + e_s$) and demand (Q = $\alpha_1 P + \alpha_2 X + e_d$) equations to describe equilibrium. P & Q are endogenous as their values are determined within their system (we created). Income (X) is created outside system (Exogenous variable). $E(e_d) = 0$, $E(e_s) = 0$, $Var(e_d) = \sigma_d^2$, $Var(e_s) = \sigma_s^2$ Cov (e_{dt}, e_{dk}) = 0, Cov (e_{st}, e_{sk}) = 0, Cov (e_{dt}, e_{st}) ≠ 0.

Makes OLS estimator biased and inconsistent. Because X is not correlated with V_1 and V_2, (zero mean, constant variance and zero covariance) the LSE is BLUE to estimate π_1 & π_2. P is still correlated with e_s and e_d. So, LSE in is biased and inconsistent due to correlation between endogenous variables on the right hand side of the equation and the random error.

Reduced Form Equations

The two previous demand and supply model, **structural equations model**, can be solved to express the endogenous variables P and Q as functions of the exogenous variable X: called the reformulation of the model to the **reduced form model** => Solving for P and Q with the equilibrium condition, $Q_d = Q_s$ leads to:

$$P = \frac{\alpha_2}{(\beta_1 - \alpha_1)}X + \frac{e_d - e_s}{(\beta_1 - \alpha_1)} = \pi_1 X + v_1$$

$$Q = \beta_1 P + e_s = \beta_1\left[\frac{\alpha_2}{(\beta_1 - \alpha_1)}X + \frac{e_d - e_s}{(\beta_1 - \alpha_1)}\right] + e_s = \frac{\beta_1\alpha_2}{(\beta_1 - \alpha_1)}X + \frac{\beta_1 e_d - \alpha_1 e_s}{(\beta_1 - \alpha_1)} = \pi_2 X + v_2$$

__**Identification: Necessary condition:**__ In a system of M equations, which jointly determine the values of M endogenous variables, at least M-1 variables must be absent from an equation for estimation of its parameter to be possible. Then equation is identified. If fewer than M-1 variables are omitted, it is unidentified. If more than M-1 variables are omitted, it is over identified (preferable). Check identification problem and address endogenous explanatory variable problem.

The Two Stage Least Square (2LS) Estimation:

After equations are identified, there are two steps: **1.** least square estimation of the reduced form equation for P and the calculation of its predicted value *P*-hatt **2.** Least square estimation of the structural equation in which the right-hand-side endogenous variable *P* is replaced by its predicted value *P*-hatt. **Properties of two-stage least squares estimators:** Biased but consistent. For large sample, 2SLS estimators normally distributed.

Reduced Form Equations: specify each endogenous variable as a function of all exogenous variables. Shift variables should be statistically significant to identify the equations. Two stage least squares estimators perform very poorly if the shift variables are not strongly significant.

Instrumental Variables (IV) Estimators /IV/ 2SLS:

$X_k = \gamma_1 + \gamma_2 X_2 + \ldots + \gamma_{k-1} X_{k-1} + \theta_1 z_1$

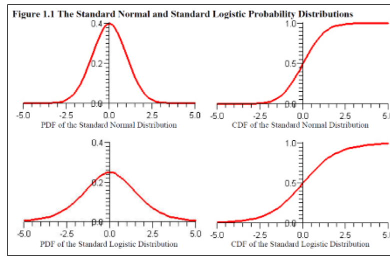
Stage I: Regress endogenous variable on all exogenous variables and instrumental variables. The strength of the instrumental variable is the strength of its relationship with X_k. Required: at least one instrumental variable should be strong.

Stage II: Replace X_k with the predicted X_k to run the regression as $y = \beta 1 + \beta_2 X_2 + \ldots + \beta_k X_{k\text{-pred}} + e^*$

Necessary conditions for IV estimation: If G good explanatory variables, B bad explanatory variables and L lucky instrumental variables, then L ≥ B. if L = B, just enough variables to do IV estimation and parameters can be consistently estimated. If L > B, more instrumental variables, and over-identified.

Discrete and Limited Dependent Variable Models:

Figure 1.1 The Standard Normal and Standard Logistic Probability Distributions



- The difference between Logistic and Probit models lies in this assumption about the distribution of the errors
- Logit

$$\ln\left(\frac{p_i}{(1-p_i)}\right) = \sum_{k=0}^{k=n}\beta_k x_{ik}$$

- *Standard logistic* distribution of errors
- Probit

$$\Phi^{-1}(p_i) = \sum_{k=0}^{k=n}\beta_k x_{ik}$$

- *Normal* distribution of errors

The shape of the logistic and normal probability functions are different and MLE of β_1 & β_2 will be different. Marginal probabilities and predicted probabilities differs very little in most cases.

Note: If a model has **unobservable dependent variable**, that the model is called index model. This is not regression model.

The distinguishing feature of the **multinomial logit** model is that there is a single explanatory variable that describes the individual, not the alternatives facing the individual. Such

- For a single explanatory factor, x_i, in the multinomial specification, the choice probabilities of individual i choosing alternative j =1 , 2, 3 are:

$$p_{i1} = \frac{1}{1 + \exp(\beta_{12} + \beta_{22}x_i) + \exp(\beta_{13} + \beta_{23}x_i)}, j = 1$$

$$p_{i2} = \frac{\exp(\beta_{12} + \beta_{22}x_i)}{1 + \exp(\beta_{12} + \beta_{22}x_i) + \exp(\beta_{13} + \beta_{23}x_i)}, j = 2$$

$$p_{i3} = \frac{\exp(\beta_{13} + \beta_{23}x_i)}{1 + \exp(\beta_{12} + \beta_{22}x_i) + \exp(\beta_{13} + \beta_{23}x_i)}, j = 3$$

- The parameters for the 1st alternative are set to zero to make the probabilities sum to one. Setting $\beta_{11} = \beta_{21} = 0; \exp(\beta_{11} + \beta_{21}x_i) = 1$

The probability that $y = 1$ is:

$$p = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x)}} = \frac{\exp(\beta_1 + \beta_2 x)}{1 + \exp(\beta_1 + \beta_2 x)}$$

The probability that $y = 0$ is:

$$1 - p = \frac{1}{1 + \exp(\beta_1 + \beta_2 x)}$$

variable are called individual specific, which is contrasted with alternative specific (varies from individual to individual and choices) in the **conditional logit model**.