Lecture 1
Review II

CLM: Review

• CLM Assumptions

(**A1**) DGP: $\mathbf{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$

(**A3**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\,\mathbf{I}_T$

(**A4**) $\mathbf{X}$ has full column rank – rank($\mathbf{X}$)=k-, where $T \geq k$.

• OLS estimation: $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\mathbf{y}$

$Var[\mathbf{b}|\mathbf{X}] = \sigma^2\,(\mathbf{X'X})^{-1}$

=> **b** unbiased and efficient (MVUE)

• If (**A5**) $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$ => $\mathbf{b}|\mathbf{X} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X'}\,\mathbf{X})^{-1})$

Now, **b** is also the MLE (consistency, efficiency, invariance, etc). (**A5**) gives us *finite sample* results for **b** (and for tests: *t-test*, *F-test*, Wald tests).

CLM: Review - Relaxing the Assumptions

• Relaxing the CLM Assumptions:

(1) (**A1**) – Lecture 5. Now, some non-linearities in the DGP are OK.

=> as long as we have intrinsic linearity, **b** keeps its nice properties.

(2) (**A4**) and (**A5**) – Lecture 7. Now, **X** stochastic: $\{x_i, \varepsilon_i\}$ i=1, 2, ..., T is a sequence of independent observations. **X** has finite means and variances. Similar requirement for $\boldsymbol{\varepsilon}$, but we also require E[$\boldsymbol{\varepsilon}$]=**0.**

Two new assumptions:

(**A2'**) plim $(\mathbf{X'\varepsilon}/T) = \mathbf{0}$.

(**A4'**) plim $(\mathbf{X'X}/T)=\mathbf{Q}$.

=> Only asymptotic results for **b** (consistency, asymptotic normality). Tests only have large sample distributions. Boostrapping or simulations may give us better finite sample behavior.

CLM: Review - Relaxing the Assumptions

(3) (**A2'**) – Lecture 8. A new estimation: IVE/2SLS. Find l variables **Z** such that

(1) plim$(\mathbf{Z'X}/T) \neq \mathbf{0}$ (*relevant* condition)

(2) plim$(\mathbf{Z'\varepsilon}/T) = \mathbf{0}$ (*valid* condition –or *exogeneity*)

$$b_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$$
$$b_{IV} = (Z'X)^{-1}Z'y$$

=> We only get asymptotic results for $\mathbf{b}_{2SLS}$ (consistency, asymptotic normality). Tests only have asymptotic distributions. Small sample behavior may be bad. Problem: Finding **Z**.

(4) (**A1**) again! – Lecture 9. Any functional form is allowed. General estimation framework: M-estimation, with only asymptotic results. A special case: NLLS. Numerical optimization needed.

Generalized Regression Model

• Back to the CLM Assumptions:

(**A1**) DGP: $\mathbf{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$

(**A3**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\,\mathbf{I}_T$

(**A4**) $\mathbf{X}$ has full column rank – rank($\mathbf{X}$)=k-, where $T \geq k$.

• Relax (**A3**) =>(**A3'**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$ where $\boldsymbol{\Omega} \neq \mathbf{I}_T$

• Generalized regression model (GRM): Variances differ across observations and non-zero correlation across observations.

• Implication: Under (**A3'**) $Var[\mathbf{b}|\mathbf{X}] \neq \sigma^2\,(\mathbf{X}'\mathbf{X})^{-1}$.

=> True variance of **b**: $Var_T[\mathbf{b}|\mathbf{X}] = \sigma^2\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}$

Generalized Regression Model

• (**A3'**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$.

• Leading Cases:

- Pure heteroscedasticity: $E[\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j|\mathbf{X}] = \sigma_{ij} = \sigma_i^2$ if $i=j$
  $= 0$ if $i \neq j$
- Pure autocorrelation: $E[\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j|\mathbf{X}] = \sigma_{ij}$ if $i \neq j$
  $= \sigma^2$ if $i=j$

• Heteroscedasticity and autocorrelation are different problems. They generally occur with different types of data. But, have similar implications for OLS.

Generalized Regression Model: OLS Properties

(1) *Unbiasedness*. From (**A2**), the OLS estimator **b** is still *unbiased*.

(2) Consistency? Assume (**A2'**) instead. To get *consistency*, we need

plim $(\mathbf{X}'\mathbf{X}/T) = \mathbf{Q}_{XX}$ ($\mathbf{Q}_{XX}$ a pd matrix of finite elements).

plim $(\mathbf{X}'\mathbf{\Omega}\mathbf{X}/T) = \mathbf{Q}_{X\Omega X}$, a finite matrix.

Then, the true variance of **b**,

$$\text{Var}_T[\mathbf{b}|\mathbf{X}] = (\sigma^2/T)(\mathbf{X}'\mathbf{X}/T)^{-1}(\mathbf{X}'\mathbf{\Omega}\mathbf{X}/T)(\mathbf{X}'\mathbf{X}/T)^{-1}$$

converges to 0 as T grows.

(3) *Asymptotic normality*? For OLS we apply the CLT to $\mathbf{X}'\boldsymbol{\varepsilon}/\sqrt{T}$.
- Easy to do for heteroscedastic data. Apply the Lindeberg-Feller (assuming only independence) version of the CLT.
- Difficult for autocorrelated data, since $\mathbf{X}'\boldsymbol{\varepsilon}/\sqrt{T}$ is not longer an independent sum. We need more assumptions.

GR Model: Robust Covariance Matrix

• $\mathbf{\Omega}$ is unknown. It has Tx$(T+1)/2$ elements to estimate. Too many! We need a model for $\mathbf{\Omega}$.

• But, models for $\mathbf{\Omega}$ may be incorrect. The *robust* estimation of the covariance matrix is robust to misspecifications of (**A3'**).

• We need to estimate $\text{Var}_T[\mathbf{b}|\mathbf{X}] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

• Important distinction:
- $\sigma^2\mathbf{\Omega}$, a (TxT) matrix => difficult to estimate with T observations.
- $\sigma^2\mathbf{X}'\mathbf{\Omega}\mathbf{X} = \sigma^2\Sigma_i\Sigma_j\omega_{ij}\,\mathbf{x}_i\,\mathbf{x}_j'$, a ($kxk$) matrix => easier!

GR Model: Robust Covariance Matrix

• We estimate $\sigma^2 \mathbf{X'\Omega X} = \sigma^2 \Sigma_i \Sigma_j \omega_{ij}\, \mathbf{x}_i\, \mathbf{x}_j'$, a ($kxk$) matrix. We estimate ($k$x($k$+1))/2 elements.

• This distinction is very important in modern applied econometrics:
– The White estimator
– The Newey-West estimator

• The White estimator assumes heteroscedasticity only. $\mathbf{\Omega}$ is a diagonal matrix. We only need to estimate $\sigma_i^2 = \sigma^2\, \omega_i$
=> $\mathbf{Q}^* = (1/T)\, \sigma^2 \mathbf{X'\Omega X} = (1/T)\, \sigma^2 \Sigma_i \omega_i\, \mathbf{x}_i\, \mathbf{x}_i' = (1/T)\, \Sigma_i\, \sigma_i^2\, \mathbf{x}_i\, \mathbf{x}_i'$

Covariance Matrix: The White Estimator

• Since only heteroscedasticity is assumed, we need to estimate:

$$\mathbf{Q}^* = (1/T)\, \sigma^2 \mathbf{X'\Omega X} = (1/T)\, \Sigma_i\, \sigma_i^2\, \mathbf{x}_i\, \mathbf{x}_i'$$

• Since $\mathbf{b}$ is a consistent estimator of $\boldsymbol{\beta}$, $\mathbf{e}$, are consistent estimators of $\boldsymbol{\varepsilon}$. Then we use e_i^2 to estimate σ_i^2.
=> Estimate $(1/T)\, \sigma^2 \mathbf{X'\Omega X}$ with $\mathbf{S_0} = (1/T)\, \Sigma_i\, e_i^2\, \mathbf{x}_i\, \mathbf{x}_i'$.

Note: The estimator is also called the *sandwich estimator,* the *White estimator*. It is also known as *Eiker-Huber-White estimator*.

• The White estimator allows us to make inferences using the OLS estimator $\mathbf{b}$ in situations where heteroscedasticity is suspected, but we do not know enough its functional form or nature.

The White Estimator: Some Remarks

(1) The White estimator is consistent, it may not perform well in finite samples. A good small sample adjustment:

$$(\mathbf{X'X})^{-1} [\Sigma_i e_i^2/(1-h_{ii})^2 \mathbf{x}_i \mathbf{x}_i'] (\mathbf{X'X})^{-1}$$

where $h_{ii} = \mathbf{x}_i(\mathbf{X'X})^{-1} \mathbf{x}_i'$

(2) The White estimator is biased (show it!). Biased corrections are popular –see above & Wu (1986).

(3) In large samples, we can overcome the problem of biased standard errors. The t-tests and F-tests are asymptotically valid.

(4) The OLS estimator remains inefficient. But inferences are asymptotically correct.

(5) The HC standard errors can be larger or smaller than the OLS ones. It can make a difference to the tests.

(6) It is included in all the packaged software programs.

Newey-West Estimator

- Now, we also have autocorrelation. We need to estimate

$$\mathbf{Q}^* = (1/T)\, \sigma^2 \mathbf{X}'\boldsymbol{\Omega}\mathbf{X} = (1/T)\, \Sigma_i\Sigma_j\, \sigma_{ij}\, \mathbf{x}_i\, \mathbf{x}_j'$$

• We would like to produce a HAC (Heteroscedasticity and Autocorrelation Consistent) estimator.

• Again, use residuals to estimate covariances. That is, we use e_ie_j to estimate σ_{ij} => natural estimator of $\mathbf{Q}^*$: $(1/T)\, \Sigma_i\Sigma_j\, e_ie_j\, \mathbf{x}_i\, \mathbf{x}_j'$

Problem: This sum has T^2 terms. Difficult to get convergence.

Solution: Cut short the sum. Usually, use weights in the sum that imply that the process becomes less autocorrelated as time goes by.

Newey-West Estimator

• We want to estimate $\mathbf{Q}^* = (1/T)\, \Sigma_i \Sigma_j\, \boldsymbol{\sigma}_{ij}\, \mathbf{x}_i\, \mathbf{x}_j'$
We use $e_i e_j$ to estimate $\boldsymbol{\sigma}_{ij}$.

• <u>Practical problem</u>: $\mathbf{Q}^*$ needs to be pd. Based on a quadratic form, Newey-West (1987) produce a consistent pd estimator of $\mathbf{Q}^*$.

• Two components for the HAC estimator:
(1) Start with Heteroscedasticity Component:
$\mathbf{S_0} = (1/T)\, \Sigma_i\, e_i^2\, \mathbf{x}_i\, \mathbf{x}_i'$ --the White estimator.
(2) Add the Autocorrelation Component
$\mathbf{S} = \mathbf{S}_0 + (1/T)\, \Sigma_l\, w_L(l)\, \Sigma_{j=l+1,\ldots,T}\, (\mathbf{x}_{i-l}\, e_{t-l}\, e_t\, \mathbf{x}_j' + \mathbf{x}_i\, e_t\, e_{t-l}\, \mathbf{x}_{j-l}')$
where
$w_L(l) = 1 - |l|/L+1$ --This is the *Bartlett kernel* or *window*.
Then,
Est. Var[$\mathbf{b}$] = $(1/T)\, (\mathbf{X'X}/T)^{-1}\, \mathbf{S}\, (\mathbf{X'X}/T)^{-1}$ --NW's HAC est.

Newey-West Estimator

• Other kernels - $w_L(l)$- besides the Bartlett have been proposed in the HAC estimator literature to estimate $\mathbf{Q}^*$:
$\mathbf{Q}^* = \mathbf{S}_0 + (1/T)\, \Sigma_l\, w_L(l)\, \Sigma_{j=l+1,\ldots,T}\, (\mathbf{x}_{i-l}\, e_{t-l}\, e_t\, \mathbf{x}_j' + \mathbf{x}_i\, e_t\, e_{t-l}\, \mathbf{x}_{j-l}')$

- Parzen kernel –Gallant (1987).
- Quadratic spectral kernel –Andrews (1991):
$w_L(l) = 25/(12\pi^2 l^2)[\sin(6\,\pi\, l/5)/(6\,\pi\, l) - \cos(6\,\pi\, l/5)]$
- Daniel kernel –Ng and Perron (1996):
$w_L(l) = \sin(\pi\, l)/(\pi\, l)$

• The quardratic spectral kernel has the lowest (8.6% better than Parzen's) asymptotic MSE. The Bartlett kernel is the least efficient.

• There are estimators of $\mathbf{Q}^*$ that are not consistent, but with better small sample properties. See Kiefer, Vogelsang and Bunzel (2000).

Newey-West Estimator

• A key assumption in establishing consistency is that $L \to \infty$ as $T \to \infty$, but $L/T \to 0$.

• In practice, the fraction L/T is never equal to 0, but approaches some positive fraction.

• Kiefer and Vogelsang (2005) derive the limiting distribution of $\mathbf{Q}^*$ under the assumption that the fraction L/T stays constant as $T \to \infty$. Under this assumption, the NW estimator is no longer consistent.

• Thus, traditional t- and F-tests no longer converge in distribution to Normal and χ^2 RVs, but they do converge in distribution to RVs that do not depend on the unknown value of $\mathbf{\Omega}$. Tests are still possible.

Generalized Least Squares (GLS)

• Assumptions (**A1**), (**A2**), (**A3'**) & (**A4**) hold. That is,

(**A1**) DGP: $\mathbf{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$

(**A3'**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\,\mathbf{\Omega}$ (recall $\mathbf{\Omega}$ is symmetric => $\mathbf{T'T} = \mathbf{\Omega}$)

(**A4**) $\mathbf{X}$ has full column rank –i.e., rank($\mathbf{X}$)=k--, where $T \geq k$.

• GLS: Transform the linear model in (**A1**) using $\mathbf{P} = \mathbf{\Omega}^{-1/2}$.

$$\mathbf{Py} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon} \quad \text{or}$$

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*.$$

$$E[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}|\mathbf{X}^*] = \sigma^2\mathbf{I}_T$$

=> OLS in the transformed model satisfies G-M theorem.

$$=> \quad \mathbf{b}_{GLS} = \mathbf{b}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$$

Note I: $\mathbf{b}_{GLS} \neq \mathbf{b}$. $\mathbf{b}_{GLS}$ is BLUE by construction, $\mathbf{b}$ is not.

Note II: Both unbiased and consistent. Should not be that different.

Generalized Least Squares (GLS)

• Relax (**A2**) and (**A4**), then only asymptotic properties for GLS:
- Consistency - "well behaved data"
- Asymptotic distribution under usual assumptions
(easy for heteroscedasticity, complicated for autocorrelation)
- Wald tests and *F*-tests with usual asymptotic χ^2 distributions.

Consistency – Autocorrelation case

$$\frac{X'\Omega^{-1}X}{T} = \frac{1}{T}\sum_{j=1}^{T}\sum_{i=1}^{T}\frac{1}{\omega_{ij}}x_i x_j' = \frac{\sigma_0}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\rho_{t-s}$$

• If the $\{X_t\}$ were uncorrelated –i.e., ρ_k=0-, then Var[$\mathbf{b}_{GLS}|\mathbf{X}$] → 0.

• We need to impose restrictions on the dependence among the X_t's. Usually, we require that the autocorrelation, ρ_k, gets weaker as *t-s* grows (and the double sum becomes finite).

Asymptotic Normality – Autocorrelation case

For the autocorrelation case

$$\frac{1}{n}\mathbf{X}'\mathbf{\Omega}^{-1}\boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{\Omega}^{\mathbf{ij}}\,\mathbf{x}_i\mathbf{x}_j\varepsilon_i\varepsilon_i$$

Does the double sum converge? Uncertain. Requires elements of $\mathbf{\Omega}^{-1}$ to become small as the distance between i and j increases. (Has to resemble the heteroscedasticity case.)

• The dependence is usually broken by assuming $\{\mathbf{x}_i\ e_t\}$ form a *mixing* sequence. The intuition behind mixing is simple; but, the formal details and its application to the CLT can get complicated.

• *Intuition*: $\{Z_t\}$ is a mixing sequence if any two groups of terms of the sequence that are far apart from each other are approximately independent --and the further apart, the closer to being independent.

Brief Detour: Time Series

• With autocorrelated data, we get dependent observations. Recall,

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

• The independence assumption (**A2'**) is violated. The LLN and the CLT cannot be easily applied, in this context. We need new tools and definitions:

- S*tationarity*. It imposes conditions on the moments/distribution (time invariant.)

- *Ergodicity*. The ergodic theorem give us a counterpart to the LLN for dependent RVs.

• We can also rely on the *martingale CLT*.

GLS: General Remarks

- GLS is great (BLUE) if we know **Ω**. Very rare case.
- It needs the specification of **Ω** -i.e., the functional form of autocorrelation and heteroscedasticity.
- If the specification is bad => estimates are biased.
- In general, GLS is used for larger samples, because more parameters need to be estimated.
- Feasible GLS is not BLUE (unlike GLS); but, it is consistent and asymptotically more efficient than OLS.
- We use GLS for inference and/or efficiency. OLS is still unbiased and consistent.
- OLS and GLS estimates will be different due to sampling error. But, if they are very different, then it is likely that some other CLM assumption is violated –likely, (**A2'**).

Testing for Heteroscedasticity

• Usual strategy when heteroscedasticity is suspected: Use OLS along the White estimator. This will give us consistent inferences.

• Q: Why do we want to test for heteroscedasticity?
A: OLS is no longer efficient. The estimator with lower asymptotic variance: GLS/FGLS estimator.

• We want to test: H_0: $E(\varepsilon^2 | x_1, x_2, \ldots, x_k) = E(\varepsilon^2) = \sigma^2$

• Key: Whether $E[\varepsilon^2] = \sigma^2\omega_i$ is related to $\mathbf{x}$ and/or x_i^2.

• Popular LM tests: White and BP

Testing for Heteroscedasticity

• Simple strategy: Use OLS residuals to estimate disturbances and look for relationships between e_i^2 and x_i and/or x_i^2.

• Suppose that the relationship between ε^2 and **X** is linear:

$$\boldsymbol{\varepsilon}^2 = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{v}$$

Then, we test: H_0: $\boldsymbol{\alpha} = 0$ against H_1: $\boldsymbol{\alpha} \neq 0$.

• Base the test on how the squared OLS residuals **e** correlate with **X**.

• These LM tests are asymptotically equivalent to a $T\,R^2$ test, where R^2 is calculated from a regression of e_i^2 on the variables that are suspected to cause heteroscedasticity.

Testing for Heteroscedasticity: BP Test

• Usual calculation of the Breusch-Pagan test

- Step 1. From your model for the mean, get OLS residuals, **e**.
- Step 2. (Auxiliary Regression). Run the regression of e_i^2 on all the explanatory variables, **z**. In our example,

$$e_i^2 = \alpha_0 + z_{i,1}'\,\alpha_1 + ... + z_{i,m}'\,\alpha_m + v_i$$

- Step 3. Keep the R^2 from this regression. Let's call it R_{e2}^2 Calculate either

(a) $F = (R_{e2}^2/m)/[(1-R_{e2}^2)/(T-(m+1)]$, which follows a $F_{m,(T-(m+1)}$

or

(b) $LM = T\,R_{e2}^2$, which χ^2_m.

• Koenker's (1981) studentized LM test is a good variation. It is robust to departures from normality:

$$\text{LM-S} = (2\,\sigma_R^4)\ \text{LM-BP}/[\Sigma\ (\varepsilon_i^2-\sigma_R^2)^2\ /T] \xrightarrow{d} \chi^2_m$$

Testing for Heteroscedasticity: White Test

- Usual calculation of the White test

- Step 1. From your model for the mean, get OLS residuals, **e**.

- Step 2. (Auxiliary Regression). Regress e^2 on all the explanatory variables (X_j), their squares (X_j^2), and all their cross products. For example, when the model contains $k = 2$ explanatory variables, the test is based on:

$$e_i^2 = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i}^2 + \beta_4 x_{2,i}^2 + \beta_5 x_1 x_{2,i} + v_i$$

Let m be the number of regressors in auxiliary regression. Keep R^2, say R_{e2}^2.

- Step 3. Compute the statistic

$$LM = T\, R_{e2}^2, \text{ which follows a } \chi^2_m.$$

Testing for Heteroscedasticity: Remarks

• Drawbacks of the Breusch-Pagan test:

- Sensitive to violations of the normality assumption.

- Three other popular LM tests: the Glejser test; the Harvey-Godfrey test, and the Park test, are also sensitive to such violations.

• Drawbacks of the White test

- With several regressors, the test can consume a lot of df's.

- Too general, it does not give us a precise way to model heteroscedasticity to do FGLS. The BP test points us in a direction.

- In simulations, it does not perform well relative to others, especially, for time-varying heteroscedasticity, typical of financial time series.

- The White test does not depend on normality; but the Koenker's test seems to have more power

Testing for Autocorrelation: LM tests

• Several autocorrelation tests. Under the null hypothesis of no autocorrelation of order p, we have H_0: $\rho_1 = ... = \rho_p = 0$.

Under H_0, we can use OLS residuals.

• Breusch–Godfrey (1978) LM test. Similar to the BP test:

- Step 1. (Auxiliary Regression). Run the regression of e_i on all the explanatory variables, **z**. In our example,

$$e_t = \mathbf{X}_t' \boldsymbol{\beta} + \alpha_1 e_{t-1} + ... + \alpha_p e_{t-p} + v_t$$

- Step 2. Keep the R^2 from this regression. Let's call it R_e^2. Then,

$$LM = T R_e^2, \text{ which } \chi^2_p.$$

Testing for Autocorrelation: Portmanteu tests

• Box-Pierce (1970) test.

It test H_0: $\rho_1 = ... = \rho_p = 0$ using the sample correlation r_j:

$$r_j = \Sigma_{t=1,...T-j} e_t e_{t-j} / \Sigma_{t=1,...T} e_t^2$$

Then, under H_0

$$Q = T \Sigma_{j=1,...,p} r_j^2 \xrightarrow{d} \chi^2_p$$

• Ljung-Box (1978) test.

A variation of the Box-Pierce test. It has a small sample correction.

$$LB = T (T-2) \Sigma_{j=1,...,p} r_j^2 / (T-j)$$

• The LB statistic is widely used. But, the Breusch–Godfrey (1978) LM test conditions on **X**. Thus, it is more powerful.

Building the Model

- Old (pre-LSE school) view: A feature of the data
  - "Account" for heteroscedasticity/autocorrelation in the data.
  - Different models, different estimators

- Contemporary view: Why is there heteroscedasticity and or autocorrelation?
  - What is missing from the model?
  - Build in appropriate dynamic structures
  - Autocorrelation should be "built out" of the model
  - Use robust procedures (White or Newey-West) instead of elaborate models specifically for the autocorrelation.