Lecture 7-a
Departures from OLS Assumptions

Brooks (4th edition): Chapter 5

 1

Review: Model Selection Strategies

• Q: How do we propose and select a model (a DGP)?

• Potentially, we have a huge number of possible models with:
- Different functional form: *f(.)*, *g(.)*, *h(.)*, etc.
- Different explanatory variables: **X**, **Z**, **W**, dummy variables, **D**, etc.
Suppose, we have 4 different models to choose from:

Model 1 $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
Model 2 $\boldsymbol{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\xi}$
Model 3 $\boldsymbol{y} = (\mathbf{W}\boldsymbol{\gamma})^{\lambda} + \boldsymbol{\eta}$
Model 4 $\boldsymbol{y} = \exp(\mathbf{Z}\ \mathbf{D}\ \boldsymbol{\delta}) + \boldsymbol{\epsilon}$

• We want to select the best model, the one that is closest to the true and unobserved DGP. In practice, we aim for a "good" model.

Review: Model Selection Strategies – Methods

- There are several *model-selection methods*. We consider two:

- **Specific to General**. Start with a small "restricted model," do some testing & make model bigger model in the direction indicated by the tests (for example, add variable $\mathbf{x}_k$ when test reject H_0: $\beta_k = 0$).

Popular application: Stepwise Regression.

Main Problem: Not clear when to stop considering adding variables.

- **General to Specific**. Start with a big "general unrestricted model," do some testing & reduce model in the direction indicated by the tests (for example, eliminate variable $\mathbf{x}_k$ when test cannot reject H_0: $\beta_k = 0$).

Popular application: Best subset.

Main Problems: Mass significance & pre-testing (data mining).

Model Selection Strategies: Best Subset

- Begin with a big model, with k regressors:

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

The idea is to select the "best" subset of the k regressors in $\mathbf{X}$, where "best" is defined by the researcher, say MSE, Adjusted-R^2, etc.

- In theory, it requires 2^k regressions. It can take a while if k is big (k < 40 is no problem).

- Many tricks are used to reduce the number of regressions.

- In practice, we use best subset to reduce the number of models to consider. For example, from the regressions with one-variable, keep the best one-variable model, from the regression with two-variables, keep the best two-variable model, etc.

Model Selection Strategies: Best Subset

Example: We want to select a model for IBM excess returns, using the k=3 Fama-French factors: Market excess returns (Mkt_RF), SMB, & HML. We have 8 (=2^3) models and, thus, regressions:

1) Constant;
2) Mkt_RF (CAPM)
3) SMB
4) HML
5) Mkt_RF & SMB
6) Mkt_Rf & HML
7) SMB & HML
8) Mkt_RF, SMB, & HML (the 3-factor F-F Model).

• We select the model with the lower MSE. Or, we can carry two or three models of the best models to do *cross-validation*.

Model Selection Strategies: Best Subset

Example (continuation): We use library olsrr in R:

```
library(olsrr)
ff_step_data <- data.frame(Mkt_RF, SMB, HML)
fit_ibm_ff3_sb <- lm(ibm_x ~ ., data = ff_step_data)        # default p-value (penter) is 0.3
ols_step_best_subset(fit_ibm_ff3_sb, details = TRUE)        #  long final output

Model Index    Predictors
----------------------------
    1        Mkt_RF
    2        Mkt_RF SMB
    3        Mkt_RF SMB HML
----------------------------
```

Subsets Regression Summary

Model	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC	SBC	MSEP	FPE
1	0.3128	0.3116	0.308	8.3178	-1705.0204	-3424.8023	-1691.7998	2.1146	0.0035
2	0.3214	0.3192	0.3134	2.6125	-1710.7200	-3430.4398	-1693.0924	2.0913	0.0035
3	0.3221	0.3187	0.311	4.0000	-1709.3362	-3429.0366	-1687.3018	2.0927	0.0035

Model Selection Strategies: Best Subset

Example (continuation): Suppose we selected three model: CAPM (M1); Mkt_RF & SMB (M2); and the 3-factor F-F Model (M3).

Now, we use *K-fold cross-validation*, with $K = 5$. That is,

$$CV_5 = \frac{1}{5}\sum_{i=1}^{5} MSE_{(-i)}$$

CV_5 M1: 0.003542756

CV_5 M2: **0.003505873**

CV_5 M3: 0.003556918

Note: Models look very similar. Practitioners compute a SE for CV_K and use a one SE rule. If within one SE, keep simplest model (M1).

R Note: Use the same CV function, from Lecture 6-c, to compute CV.

Model Selection Strategies: Judgement Calls

• In the end, judgment must be used to select a model. In general, we consider various criteria:

- **The Economic Criterion** –are the estimated parameters plausible? (Economic Significance).

- **The First Order Statistical Criterion** –does the model provide a good fit (in-sample) with statistically significant parameter estimates?

- **The Second Order Statistical Criterion** –is the model generally free of misspecification problems – as evidenced in the diagnostic tests?

- **The Out of Sample Predictive Criterion** –does the model provide good out of sample predictions?

CLM: Review

• Recall the CLM Assumptions

(**A1**) DGP: $\boldsymbol{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$ 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: $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\boldsymbol{y}$

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

$\Rightarrow$ **b** unbiased and efficient (MVUE)

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

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). 2

CLM: Departures from the Assumptions

• So far, we have discussed some violations of CLM Assumptions:

(1) (**A1**) – OLS can easily deal with some non-linearities in the DGP.

$\Rightarrow$ as long as we have intrinsic linearity, **b** keeps its nice properties.

– Wald, F, & LM tests to check for misspecification

(2) (**A4**) – Perfect Multicollinearity means we need to change the model. Multicollinearity is a potential problem. In general, exogenous to the researcher. We need to be aware of this problem.

• Now, we examine assumptions (**A2**), (**A3**) and (**A5**). We change:

(i) $\mathbf{X}$ is stochastic. That is, it has a distribution.

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

(iii) $\boldsymbol{\varepsilon}|\mathbf{X}$ is not $N(\mathbf{0}, \sigma^2\mathbf{I}_T)$

3

CLM: Departures from (A2)

• The traditional derivation of the CLM assumes **X** as non-stochastic. In our derivation, however, we allowed **X** to be stochastic, but we conditioned on observing its realizations (an elegant trick, but not very realistic).

• Now, we allow a **stochastic X**. But, we need additional assumptions to get unbiasedness and consistency for the OLS **b**.

– We need independence between **X** & **ε**: $\{x_i, \varepsilon_i\}$ $i = 1, 2, ..., T$ is a sequence of independent observations.

– We require that **X** have finite means and variances. Similar requirement for **ε**, but we also require E[**ε**] = **0**.

Then, we get unbiasedness:

$$E[\mathbf{b}] = \boldsymbol{\beta} + E[(\mathbf{X'X})^{-1}\mathbf{X'}\,\boldsymbol{\varepsilon}] = \boldsymbol{\beta} + E[(\mathbf{X'X})^{-1}\mathbf{X'}]\,E[\boldsymbol{\varepsilon}] = \boldsymbol{\beta}$$

4

CLM: Departures from (A2)

• Technical Note: To get consistency (& asymptotic normality) for **b**, we need an additional (asymptotic) assumption regarding **X**:

$$\mathbf{X'X}/T \xrightarrow{p} \mathbf{Q} \qquad (\mathbf{Q}\text{: pd } (k\text{x}k) \text{ matrix of finite elements})$$

or $\quad \text{plim}\,(\mathbf{X'X}/T) = \mathbf{Q}$

- Q: Why do we need this assumption in terms of a ratio divided by T? Each element of **X'X** matrix is a sum of T numbers.

$$\boldsymbol{X'X} = \begin{bmatrix} \Sigma_{i=1}^{T} x_{i1}^2 & \Sigma_{i=1}^{T} x_{i1}x_{i2} & \dots & \Sigma_{i=1}^{T} x_{i1}x_{iK} \\ \Sigma_{i=1}^{T} x_{i2}x_{i1} & \Sigma_{i=1}^{T} x_{i2}^2 & \dots & \Sigma_{i=1}^{T} x_{i2}x_{iK} \\ \dots & \dots & \dots & \dots \\ \Sigma_{i=1}^{T} x_{iK}x_{i1} & \Sigma_{i=1}^{T} x_{iK}x_{i2} & \dots & \Sigma_{i=1}^{T} x_{iK}^2 \end{bmatrix}$$

As $T \to \infty$, these sums will become large. We divide by T so that the sums will not be too large.

5

CLM: Departures from (A2)

• We divide by T so that the sums will not be too large.

Note: This assumption is not a difficult one to make since the LLN suggests that the each component of $\mathbf{X'X}/T$ goes to the mean values of $\mathbf{X'X}$. We require that these values are finite.

– Implicitly, we assume that there is not too much dependence in **X**.

• If there is **dependence** between **X** & **ε**, OLS **b** is no longer unbiased or consistent. Easy to see the biased result: we cannot longer separate $E[(\mathbf{X'X})^{-1}\mathbf{X'}\,\boldsymbol{\varepsilon}]$ into a product of two expectations:

$$E[(\mathbf{X'X})^{-1}\mathbf{X'}\,\boldsymbol{\varepsilon}] \neq E[(\mathbf{X'X})^{-1}\mathbf{X'}]\, E[\boldsymbol{\varepsilon}]$$

Then,

$$E[\mathbf{b}] = \boldsymbol{\beta} + E[(\mathbf{X'X})^{-1}\mathbf{X'}\,\boldsymbol{\varepsilon}] \neq \boldsymbol{\beta}$$

5

CLM: Departures from (A2) – Endogeneity

• Then,

$$E[\mathbf{b}] = \boldsymbol{\beta} + E[(\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}] \neq \boldsymbol{\beta}$$

• Dependence between **X** & **ε** occurs when **X** is also an **endogenous variable**, like $\boldsymbol{y}$. This is common, especially in Corporate Finance.

Example: We study CEO compensation as function of Size of a firm, and Board composition. Board Composition & Size of a firm are endogenous –i.e., determined by the firm, dependent on CEO's decisions.

• Inconsistency is a fatal flaw in an estimator. In these situations, we use different estimation methods. The most popular is **Instrumental Variable (IV) estimation**.

6

CLM: Departures from (A2) – Asymptotics

• Now, we have a new set of assumptions in the CLM:

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

(**A2'**) $\mathbf{X}$ stochastic, but $E[\mathbf{X}'\,\boldsymbol{\varepsilon}] = 0$ and $E[\boldsymbol{\varepsilon}] = \mathbf{0}$.

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

(**A4'**) plim $(\mathbf{X'X}/T) = \mathbf{Q}$ (p.d. matrix with finite elements, rank=k)

• With these new assumptions and using properties of *probability limits* (plims) & the CLT, we can show the following asymptotic results:

1. $\mathbf{b}$ and s^2 are consistent.
2. $\sqrt{T}\,(\mathbf{b} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$ $\Rightarrow \mathbf{b} \xrightarrow{a} N(\boldsymbol{\beta}, (\sigma^2/T)\mathbf{Q}^{-1})$
3. *test-t* $\xrightarrow{d}$ N(0,1)

   F-tests & Wald tests $\xrightarrow{d} \chi^2_J$

15

CLM: Departures from (A5)

• Asymptotic results 2 and 3 state the asymptotic distribution of $\mathbf{b}$ and the *t*-, *F*- and Wald test. All derived from the new set of assumptions and the CLT. (**A5**) was not used.

• That is, we relax (**A5**), but, now, we require *large samples* ($T \to \infty$).

Note: In practice, we use the asymptotic distribution as an **approximation** to the finite sample –i.e., for any T– distribution. This is why we used the $\xrightarrow{a}$ notation in:

$$\mathbf{b} \xrightarrow{a} N(\boldsymbol{\beta}, (\sigma^2/T)\,\mathbf{Q}^{-1}) \qquad (\xrightarrow{a}: \text{"}\textit{approximation}\text{"})$$

We should be aware that this approximation may not be accurate in many situations.

16

CLM: Departures from (A5) – Remarks

• Two observations regarding relaxing (**A5**) $\boldsymbol{\varepsilon}|\mathbf{X} \sim$ *i.i.d.* $N(\mathbf{0}, \sigma^2\mathbf{I}_T)$:

– Throwing away the normality for $\boldsymbol{\varepsilon}|\mathbf{X}$ is not bad.
In many econometric situations, normality is not a good assumption (daily, weekly, or monthly stock returns do not follow a normal).

– Removing the *i.i.d.* assumption for $\boldsymbol{\varepsilon}|\mathbf{X}$ is also not bad.
In many econometric situations, identical distributions are not realistic, since different means and variances are common.

• Q: Do we need to throw away normality for $\boldsymbol{\varepsilon}|\mathbf{X}$? Not necessarily. We can test for normality on the residuals using a Jarque-Bera test.

Remark: Usually, for returns of financial assets normality is rejected, especially at the monthly, weekly, daily, and intra-daily frequencies.

17

CLM: Departures from (A5) – Remarks

• Q: Why are we interested in large sample properties, like consistency, when in practice we have finite samples?

A: As a first approximation, if we can show that an estimator has good large sample properties, then we may be optimistic about its finite sample properties.

For example, if an estimator is inconsistent, we know that for finite samples it will definitely be biased.

18

CLM: Departures from (A3)

• Now, we relax (**A3**). The CLM assumes that errors are uncorrelated and all are drawn from a distribution with the same variance, σ^2.

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

Instead, we will assume:

(**A3'**) $\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\Sigma}$ (sometimes written= $\sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega} \neq \mathbf{I}_T$)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

• Two Leading Cases:

– **Pure heteroscedasticity**: Model only the diagonal elements.

– **Pure cross/autocorrelation:** Model only the off-diagonal elements.

11

CLM: Departures from (A3) – Heteroscedasticity

• **Pure heteroscedasticity:**

$$\text{E}[\varepsilon_i\, \varepsilon_j\,|\mathbf{X}] = \sigma_{ij} = \sigma_i^2 \quad \text{if } i = j$$
$$= 0 \quad \text{if } i \neq j$$
$$\Rightarrow \text{Var}[\varepsilon_i\,|\mathbf{X}] = \sigma_i^2$$

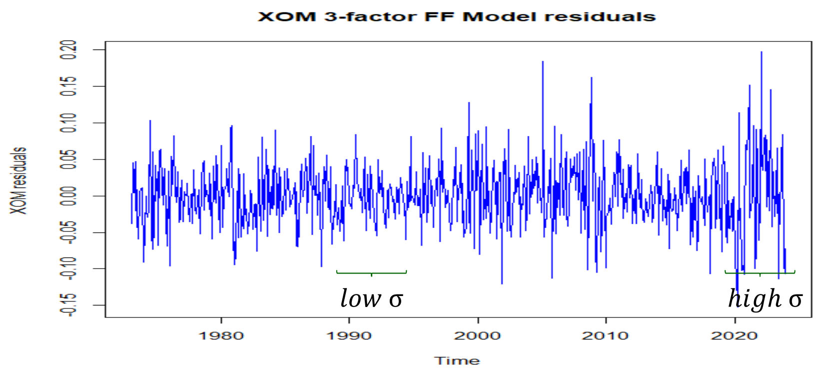
$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix}$$

• This type of variance-covariance structure is common in time series, where we observe the variance of the errors changing over time or subject to different regimes (bear/bull). Or in cross-sections, where we observe the variance of the errors change with the industry.

20

CLM: Departures from (A3) – Heteroscedasticity

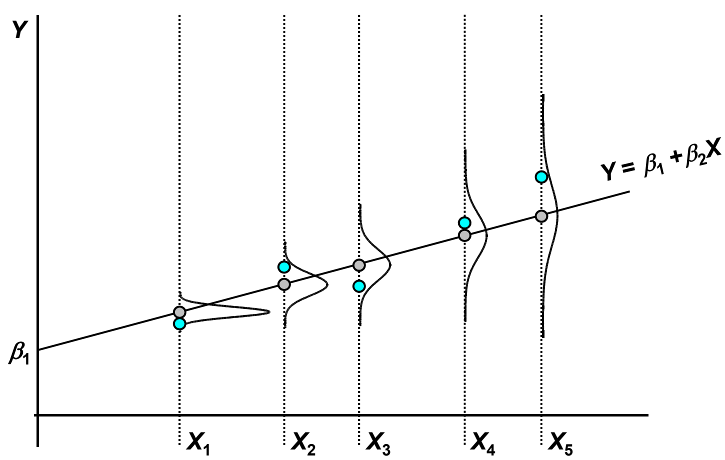
Example: We plot the residuals from the 3-factor FF model for XOM to see if there is evidence for heteroscedasticity. We observe a usual result: The variance of residuals from returns models changes over time: Periods with low variance & periods with high variance.



21

CLM: Departures from (A3) – Heteroscedasticity

• Relative to pure heteroscedasticity, LS gives each observation a weight of $1/T$. But, if the variances are not equal, then some observations (low variance ones) are more informative than others.



22

CLM: Departures from (A3) – Cross-correlation

• **Pure cross/auto-correlation:**

$$E[\varepsilon_i\, \varepsilon_j \,|\, \mathbf{X}] = \sigma_{ij} \qquad \text{if } i \neq j$$
$$= \sigma^2 \qquad \text{if } i = j$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma^2 \end{bmatrix}$$

• This type of variance-covariance structure is common in cross sections, where errors can show strong correlations, for example, when we model returns, the errors of two firms in the same industry can be subject to common (industry) shocks. Also common in time series, where we observe clustering of shocks over time.

23

CLM: Departures from (A3) – Heteroscedasticity

Example 1 (auto-correlation): We compute the interest differentials between quarterly interest rates in the U.S. and Japan, $y_t = i_{US,t} - i_{JAP,t}$. We want to check if there is evidence of autocorrelation. Below, we compute the correlations between y_t & y_{t-1} and between y_t & y_{t-2}: Cov(y_t, y_{t-1}) and Cov(y_t, y_{t-2}).

Cov(y_t, y_{t-1}) = **.872**

Cov(y_t, y_{t-2}) = **.719**

Conclusion (informal): Both covariances look very different from 0.

Example 2 (cross-correlation): We compute the correlation between the residuals of 3-factor FF model of XOM & SLB. It is equal to **.427**.

Conclusion (informal): The size of the correlation points to a moderate cross-correlation between both residuals.

24

CLM: Departures from (A3) – Cross-correlation

• Relative to pure cross/auto-correlation, LS is based on simple sums. In the case of autocorrelation, the information that one observation (today's) might provide about another (tomorrow's) is never used. In the case of cross-correlation, the information that i (XOM) might provide about j (SLB) is ignored.

Note: Heteroscedasticity and autocorrelation are different problems and generally occur with different types of data. But, the implications for OLS are the same.

25

CLM: Departures from (A3) – Implications

• OLS **b** is still *unbiased* and *consistent*. (Proofs do not rely on (**A3**).)

• OLS **b** still follows an *asymptotic normal distribution*. It is

– Easy to show for the pure heteroscedasticity case using a version of the CLT that assumes only independence

– More complicated derivation –i.e., with new assumptions– for the cross/auto-correlation case (there is dependence!).

• But, OLS **b** is **no longer** BLUE. There are more efficient estimators; estimators that take into account the heteroscedasticity in the data.

Note: We used (**A3**) to derive our test statistics. A revision is needed!

26

Finding Heteroscedasticity

• There are several theoretical reasons why the σ_i^2 may be related to some explanatory variables $z_1, \ldots, z_j$ and/or $z_1^2, \ldots, z_j^2$.

Examples:

1. Following the *error-learning models*, as people learn, their errors of behavior become smaller over time. Then, σ_i^2 is expected to decrease.6

2. As data collecting techniques improve, σ_i^2 is likely to decrease. Companies with sophisticated data processing techniques are likely to commit *fewer errors* in forecasting customer's orders.

3. As companies grow, companies expand and tend to be more diversified and, thus, safer. Hence, σ_i^2 is likely to decrease with size.

4. Companies with larger profits tend to have greater variability in their dividend/buyback policies than companies with lower profits.

27

Finding Heteroscedasticity

• Heteroscedasticity can also be the result of *outliers* (either very small or very large). The inclusion/exclusion of an outlier, especially if T is small, can affect the results of regressions.

• Violations of (**A1**) –*model is correctly specified*–, can produce heteroscedasticity, due to omitted variables from the model or incorrect functional form (e.g., linear vs log–linear models).

• *Skewness* in the distribution of one or more regressors included in the model can induce heteroscedasticity. Examples are economic variables such as income, wealth, and education.

28

Finding Heteroscedasticity

• Heteroscedasticity is usually modeled using one the following specifications:

– H1 : σ_t^2 is a function of past ε_t^2 and past σ_t^2 (ARCH models).

– H2 : σ_t^2 increases monotonically with one (or several) exogenous variable(s) $(z_1, \ldots, z_j)$.

– H3 : σ_t^2 decreases monotonically with *Size* (Market Cap).

– H4 : σ_t^2 is the same within p subsets of the data but differs across the subsets (*grouped heteroscedasticity*). This specification allows for structural breaks.

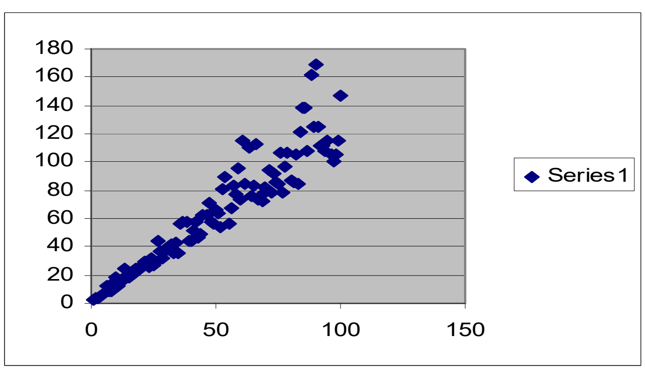
• These are the usual alternatives hypothesis (H_1) in the heteroscedasticity tests.

29

Finding Heteroscedasticity

• **Visual test**

In a plot of residuals against dependent variable or other variable will often produce a fan shape.



20

Testing for Heteroscedasticity

- Q: Why do we want to test for heteroscedasticity if **b** is unbiased?

A: OLS is no longer efficient. There is an estimator with lower asymptotic variance (the GLS/FGLS estimator).

- We want to test: H_0: $E(\varepsilon_i^2) = \sigma^2$ for all i.

- H_1 and the structure of the test depend on what we consider the drivers of σ_i^2 – i.e., in the previous examples: H1, H2, H3, H4, etc.

- The key is whether $E[\varepsilon_i^2] = \sigma_i^2$ is related to $\boldsymbol{X}$ and/or $\boldsymbol{X}^2$. Suppose we suspect a particular independent variable, say $\boldsymbol{x}_j$, is driving σ_i^2:

$$\sigma_i^2 = f(\boldsymbol{x}_j)$$

- Then, a simple test: Check the RSS for large values of $\boldsymbol{x}_j$, and the RSS for small values of $\boldsymbol{x}_j$. This is the Goldfeld-Quandt (GQ) test.

21

Testing for Heteroscedasticity: GQ Test

- GQ tests H_0: $\sigma_i^2 = \sigma^2$

  H_1: $\sigma_i^2 = f(\boldsymbol{x}_j)$

- Easy to compute:

– **Step 1**. Arrange the data from small to large values of the independent variable suspected of causing heteroscedasticity, $\boldsymbol{x}_j$.

– **Step 2**. Run two separate regressions, one for small values of $\boldsymbol{x}_j$ and one for large values of $\boldsymbol{x}_j$, omitting ***d*** middle observations (***d*** ≈ 20%). Get the RSS for each regression: RSS_1 for small values of $\boldsymbol{x}_j$ and RSS_2 for large $\boldsymbol{x}_j$'s.

– **Step 3**. Calculate the F ratio

$$GQ = \frac{RSS_2}{RSS_1}, \sim F_{df,df}, \text{ with } df = [(T - d) - 2(k+1)]/2 \quad (\textbf{A5} \text{ holds})$$

22

Testing for Heteroscedasticity: GQ Test

If (**A5**) does not hold, the F distribution becomes an approximation. Other tests may be preferred.

Note: When we suspect more than one variable is driving σ_i^2, the GQ test is not very useful.

- The GQ test is a popular test for structural breaks (two regimes) in variance. For these tests, we rewrite **Step 3** to allow for a different sample size in the sub-samples 1 and 2, since the breaking point does not have to be in the middle of the sample:

– **Step 3**. Calculate the F-test ratio

$GQ = [RSS_2/ (T_2 - k)]/[RSS_1/ (T_1 - k)]$

R Note: The package *lmtest* computes this test with function *gqtest*. It splits the sample in the middle. You need to specify ***d***.

Testing for Heteroscedasticity: GQ Test

Example: We test if the 3-factor FF model for IBM and GE returns shows heteroscedasticity with a GQ test, using *gqtest* in package *lmtest*.

- **IBM returns**

```
library(lmtest)
> gqtest(ibm_x ~ Mkt_RF + SMB + HML, fraction = .20)
        Goldfeld-Quandt test

data:  ibm_x ~ Mkt_RF + SMB + HML
GQ = 1.1006, df1 = 224, df2 = 223, p-value = 0.2371
alternative hypothesis: variance increases from segment 1 to 2
```

⇒ cannot reject H_0 at 5% level.

- **GE returns**

```
gqtest(ge_x ~ Mkt_RF + SMB + HML, fraction = .20)
        Goldfeld-Quandt test

data:  ge_x ~ Mkt_RF + SMB + HML
GQ = 2.744, df1 = 281, df2 = 281, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2
```

⇒ reject H_0 at 5% level.

Testing for Heteroscedasticity: LM Tests

• Popular heteroscedasticity LM tests:

- Breusch and Pagan (1979)'s LM test (BP).
- White (1980)'s general test.

• Both tests are based on OLS residuals, $\boldsymbol{e}$, and calculated under H_0 (No heteroscedasticity): σ^2. The squared residuals are used to estimate σ_i^2.

• The BP test is an LM test, derived under normality –i.e., (**A5**). It is a general tests designed to detect any linear forms of heteroscedasticity, driven by some variables, **z**. That is, the BP tests:

H_0: $\sigma_i^2 = \sigma^2$

H_1: $\sigma_i^2 = f(\mathbf{z}_i)$

24

Testing for Heteroscedasticity: LM Tests

• The White test is an asymptotic Wald-type test, normality is not needed. It allows for nonlinearities by using squares and cross-products of all the x's in the auxiliary regression –i.e., as the drivers of σ_i^2. That is, the White tests:

H_0: $\sigma_i^2 = \sigma^2$

H_1: $\sigma_i^2 = f(x_1^2, x_2^2, ..., x_J^2, x_1x_2, x_1x_3, x_2x_3, ...)$

24

Testing for Heteroscedasticity: BP Test

- The derivation of the BP test is complicated, it relies on the likelihood function, which is constructed under normality, and its first derivative, the score. However, the implementation of the BP test is simple, based on the squared OLS residuals, e_i^2.

- Calculation of the Breusch-Pagan test

- **Step 1**. Run OLS on DGP:

$\boldsymbol{y} = \mathbf{X}\,\beta + \varepsilon.$ –Keep e_i and compute σ_R^2 = RSS/T

- **Step 2**. (Auxiliary Regression). Run the regression of e_i^2/σ_R^2 on the m explanatory variables, **z**. In our example,

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

- **Step 3**. Keep the RSS from Step 2 regression. Let's call it RSS_e.

Calculate LM = (TSS - RSS_e)/2 $\xrightarrow{d} \chi_m^2$.

25

Testing for Heteroscedasticity: BP Test

- There is a version of the BP, which is robust to departures from normality: the "**studentized**" version of Koenker (1981). The BP test is asymptotically equivalent to a T*R^2 test, where R^2 is calculated from a regression of e_i^2/σ_R^2 on the variables **Z**. (Omitting σ_R^2, a constant, from the denominator is OK, since it does not affect R^2.)

- We have different **Steps 2** & **3**:

- **Step 2**. (Auxiliary Regression). Run the regression of $\boldsymbol{e_i^2}$ on the m explanatory variables, **z**. In our example,

$e_i^2 = \alpha_0 + \alpha_1\, z_{1,i} + ... + \alpha_m\, z_{m,i} + v_i$ –Keep R^2 (R_{e2}^2)

- **Step 3**. Using the R^2 from Step 2, R_{e2}^2, calculate

$$\text{LM} = T\, R_{e2}^2 \xrightarrow{d} \chi_m^2.$$

26

Testing for Heteroscedasticity: Example – IBM

Example: We suspect that squared Mkt_RF (x1) –a measure of the overall market's variance- drives heteroscedasticity. We do a studentized LM-BP test for IBM in the 3-factor FF model:

```
fit_ibm_ff3 <- lm (ibm_x ~ Mkt_RF + SMB + HML)   # Step 1 – OLS in DGP (3-factor FF model)
e_ibm <- fit_ibm_ff3$residuals                     # Step 1 – keep residuals
e2 <- e_ibm^2                                      # Step 1 – squared residuals
Mkt_RF_2 <- Mkt_RF^2
fit_ibm_BP <- lm (e2 ~ Mkt_RF_2)                   # Step 2 – Auxiliary regression
Re_2 <- summary(fit_ibm_BP)$r.squared              # Step 2 – keep R^2
LM_BP_test <- Re2 * T
> LM_BP_test                                       # Step 3 – Compute LM-BP test: R^2 * T
[1] 0.25038
> p_val <- 1 - pchisq(LM_BP_test, df = 1)          # p-value of LM_test
> p_val
[1] 0.6168019
```

LM-BP Test: **0.25028** $\Rightarrow$ cannot reject H_0 at 5% level ($\chi^2_{[1],.05} \approx$ **3.84**); with a *p-value* = **.6168**.

27

Testing for Heteroscedasticity: Example – IBM

Example (continuation): The *bptest* in the *lmtest* package performs a studentized LM-BP test for the same variables used in the model (Mkt, SMB and HML). For IBM in the 3-factor FF model:

```
> bptest(ibm_x ~ Mkt_RF + SMB + HML)     #bptest only allows to test H1:σ_i^2=f(x_i=model variables)

        studentized Breusch-Pagan test

data:  ibm_x ~ Mkt_RF + SMB + HML
BP = 4.1385, df = 3, p-value = 0.2469
```

LM-BP Test: **4.1385** $\Rightarrow$ cannot reject H_0 at 5% level ($\chi^2_{[3],.05} \approx$ **7.815**); with a *p-value* = **0.2469**.

Note: Heteroscedasticity in financial time series is very common. In general, it is driven by squared market returns or squared past errors.

28

Testing for Heteroscedasticity: Example – DIS

Example: We suspect that squared Market returns drive heteroscedasticity. We do an LM-BP (studentized) test for **Disney**:

```
lr_dis <- log(x_dis[-1]/x_dis[-T])                  # Log returns for DIS
dis_x <- lr_dis – RF                                # Disney excess returns
fit_dis_ff3 <- lm (dis_x ~ Mkt_RF + SMB + HML)      # Step 1 – OLS in DGP (3-factor FF model)
e_dis <- fit_dis_ff3 $residuals                     # Step 1 – keep residuals
e_dis2 <- e_dis^2                                   # Step 2 – squared residuals
fit_dis_BP <- lm (e_dis2 ~ Mkt_RF_2)                # Step 2 – Auxiliary regression
Re_2 <- summary(fit_dis_BP)$r.squared               # Step 2 – Keep R^2 from Auxiliary reg
LM_BP_test <- Re_2 * T                              # Step 3 – Compute LM Test: R^2 * T
> LM_BP_test
[1] 14.15224
p_val <- 1 - pchisq(LM_BP_test, df = 1)             # p-value of LM_test
> p_val
[1] 0.0001685967
```

LM-BP Test: **14.15** ⇒ reject H_0 at 5% level ($\chi^2_{[1],.05} \approx$ **3.84**); with a *p-value* = **.0001**.

41

Testing for Heteroscedasticity: Example – DIS

Example (continuation): We do the same test but with SMB squared for Disney:

```
e-dis2 <- e_dis^2
SMB_2 <- SMB^2
fit_dis_BP_2 <- lm (e_dis2 ~ SMB_2)
Re_2 <- summary(fit_dis_BP_2)$r.squared
LM_BP_test <- Re_2 * T
> LM_BP_test
[1] 7.564692
p_val <- 1 - pchisq(LM_BP_test, df = 1)  # p-value of LM_test
> p_val
[1] 0.005952284
```

LM-BP Test: **7.56** ⇒ reject H_0 at 5% level ($\chi^2_{[1],.05} \approx$**3.84**); with a *p-value*= **.006.**

42

Testing for Heteroscedasticity: Example – DIS

Example (continuation): If we do use the lmtest package, we get:

```
> bptest(dis_x ~ Mkt_RF + SMB + HML)

	studentized Breusch-Pagan test

data:  dis_x ~ Mkt_RF + SMB + HML
BP = 6.9935, df = 3, p-value = 0.07211
```

LM-BP Test: **6.99** ⇒ cannot reject H_0 at 5% level ($\chi^2_{3,.05} \approx$ **7.815**); with a p-value = **.07211**.

Note: In general, you need squared values when model heteroscedasticity in financial assets.

43

Testing for Heteroscedasticity: Example – GBP

Example: We suspect that squared interest rate differentials drive heteroscedasticity for residuals in encompassing (IFE + PPP) model for changes in the **USD/GBP**. We do an LM-BP (studentized) test:

```
y <- lr_usdgbp
fit_gbp <- lm (y ~ inf_dif + int_dif)
e_gbp <- fit_gbp$residuals
e_gbp2 <- e_gbp^2
int_dif_2 <- int_dif^2
fit_gbp_BP <- lm (e_gbp2 ~ int_dif_2)
Re_2 <- summary(fit_gbp_BP)$r.squared
LM_BP_test <- Re_2 * T
> LM_BP_test
[1] 21.11134
p_val <- 1 - pchisq(LM_BP_test, df = 1)                    # p-value of LM_test
>  p_val
[1] 4.333567e-06
```

LM-BP Test: **21.11134** ⇒ reject H_0 at 5% level (*p-value* < **.00001**).

32

Testing for Heteroscedasticity: White Test

- The White test derivation is also complicated, but, the usual calculation of the White test is a known one for us:

– **Step 1**. (Same as BP's Step 1). Run OLS on DGP:

$y = \mathbf{X}\,\beta + \varepsilon.$ Keep residuals, e_i.

– **Step 2**. (Auxiliary Regression). Regress e^2 on all the explanatory variables (x_j), their squares ($x_j{}^2$), & all their cross products $(x_j * x_i)$.

For example, when the model contains $k = 2$ explanatory variables, the test is based on:

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

Let m be the number of regressors in auxiliary regression (in the above example, $m = 5$). Keep R^2, say R_{e2}^2.

– **Step 3**. Compute the statistic: $\text{LM} = T\, R_{e2}^2 \xrightarrow{d} \chi_m^2$.

45

Testing for Heteroscedasticity: White Test

Example: White Test for 3 factor FF model for IBM returns, T=569:

```
e_ibm2 <- e_ibm^2
xx2 <- cbind(Mkt_RF_2, SMB_2, HML_2, Mkt_HML,Mkt_SMB, SMB_HML)    # Not including original variables is OK
fit_ibm_W <- lm(e_ibm2 ~ Mkt_RF + SMB + HML + xx2)
r2_e2 <- summary(fit_ibm_W)$r.squared          # Keep R^2 from Auxiliary regression
> r2_e2
[1] 0.0166492
lm_t <- T * r2_e2                              # Compute LM test: R² * T
> lm_t
[1] 10.93483
> qchisq(.95, df = df_lm)
[1] 12.59159
```

LM-White Test: **10.93** ⇒ cannot reject H_0 at 5% level ($\chi^2_{[6],.05}$≈**12.59**).

46

Testing for Heteroscedasticity: White Test

Example (continuation): Now, we do a White Test for the 3 factor F-F model for **DIS** and **GE** returns (T=569).

- For **DIS**, we get:

```
e_dis2 <- e_dis^2
fit_dis_W <- lm (e_dis2 ~ Mkt_RF + SMB + HML + xx2)
Re_2W <- summary(fit_dis_W)$r.squared
LM_W_test <- Re_2W * T
> LM_W_test
[1] 25.00148
```

⇒ reject H_0 at 5% level ($\chi^2_{[6],05}$≈**12.59**).

```
p_val <- 1 - pchisq(LM_W_test, df = df_lm)          # p-value of LM_test
> p_val
[1] 0.0003412389
```

- For **GE**, we get:

LM-White Test: **20.15** (*p-value* = **0.0026**) ⇒ reject H_0 at 5% level.

47

Testing for Heteroscedasticity: White Test

Example: We do a White Test for the residuals in the encompassing (IFE + PPP) model for changes in the USD/GBP (T=363):

```
e_gbp2 <- e_gbp^2
int_dif2 <- int_dif^2
inf_dif2 <- inf_dif^2
int_inf_dif <- int_dif*inf_dif
fit_gbp_W <- lm (e_gbp2 ~ int_dif + inf_dif + int_dif2 + inf_dif2+ int_inf_dif)
Re_2W <- summary(fit_gbp_W)$r.squared
LM_W_test <- Re_2W * T
p_val <- 1 - pchisq(LM_W_test, df = 3)          # p-value of LM_test

> LM_W_test
[1] 15.46692
> p_val
[1] 0.001458139
```

⇒ reject H_0 at 5% level

48

Testing for Heteroscedasticity: Remarks

• Drawbacks of the Breusch-Pagan test:

- It is sensitive to violations of the normality assumption. The studentized version of Koenker is more robust and, then, more used.

• Drawbacks of the White test

- If a model has several regressors, the test can consume a lot of df's.

- In cases where the White test statistic is statistically significant, heteroscedasticity may not necessarily be the cause, but model specification errors.

- It is general. It does not give us a clue about how 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.

49

Finding Auto-correlation

• There are several reasons why the ε_i may be related to ε_j. In general, we find autocorrelation (or serial correlation) in time series, where $\varepsilon_{i=t}$ is correlated to $\varepsilon_{j=t-l}$. Typical situation: it takes time to absorb a shock, then shocks show persistence over time.

• The shocks can also be correlated over the cross-section, causing cross-correlation. For example, if an unexpected new tax is imposed on the technology sector, all the companies in the sector are going to share this shock.

• Usually, we model autocorrelation using two models: autoregressive (AR) and moving averages (MA).

- In an AR model, the errors, ε_t, show a correlation over time.
- In an MA model, the errors, ε_t, are a function (similar to a weighted average) of previous errors, now denoted u_t's.

50

Finding Auto-correlation

Examples:

- First-order autoregressive autocorrelation: AR(1)

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + u_t$$

- p[th]-order autoregressive autocorrelation: AR(p)

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_p \varepsilon_{t-p} + u_t$$

- Third-order moving average autocorrelation: MA(3)

$$\varepsilon_t = u_t + \lambda_1 u_{t-1} + \lambda_2 u_{t-2} + \lambda_3 u_{t-3}$$

Note: The last example is described as third-order moving average autocorrelation, denoted MA(3), because it depends on the three previous innovations as well as the current one.

51

Finding Auto-correlation – Visual Check

• Plot data, usually residuals from a regression, to see if there is a pattern:

- **Positive autocorrelation**: A positive (negative) observation tends to be followed by a positive (negative) observation. We tend to see continuation in the series.

- **Negative autocorrelation**: A positive (negative) observation tends to be followed by a negative (positive) observation. We tend to see reversals.

- **No autocorrelation**: A positive (negative) observation has the same probability of being followed by a negative or positive (positive or negative) observation. We tend to see no pattern.

52

Finding Auto-correlation – Visual Check

Example: I simulate a y_t series, with N(0,1) u_t errors:

$$y_t = \rho_1 y_{t-1} + u_t$$

Three cases:

(1) **Positive autocorrelation**: $\rho_1 = .70$

(2) **Negative autocorrelation**: $\rho_1 = -.70$

(3) **No correlation**: $\rho_1 = 0$

- R code for simulation:

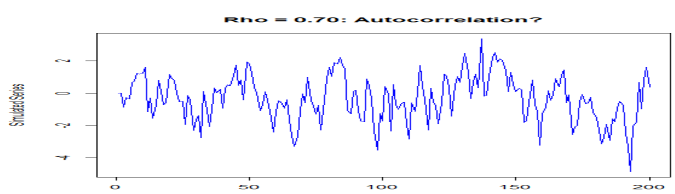
```
T_sim <- 200
u <- rnorm(200)                                # Draw T_sim normally distributed errors
y_sim <- matrix(0,T_sim,1)
rho <- .7                                      # Change to create different correlation patterns
a <- 2                                         # Time index for observations
while (a <= T_sim) {
        y_sim[a] = rho * y_sim[a-1] + u[a]     # y_sim simulated autocorrelated values
a <- a + 1
}
plot(y_sim, type="l", col="blue", ylab ="Simulated Series", xlab ="Time")
title("Visual Test: Autocorrelation?")
```

53

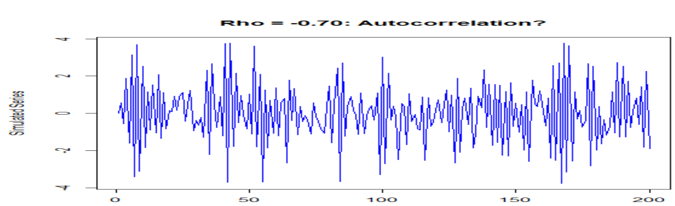
Finding Auto-correlation – Visual Check

Example (continuation):

(1) **Positive autocorrelation:** $\rho_1 = .70$



(2) **Negative autocorrelation:** $\rho_1 = -.70$

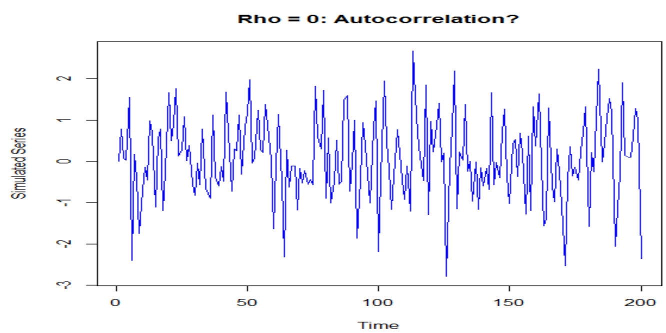


54

Finding Auto-correlation – Visual Check

Example (continuation):

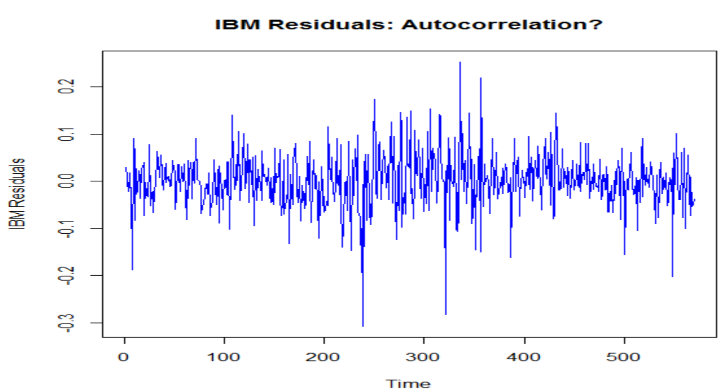
(3) **No autocorrelation**: $\rho_1 = 0$



55

Finding Auto-correlation – Visual Check: IBM

Example: Residual plot for the 3 factor F-F model for **IBM returns**:

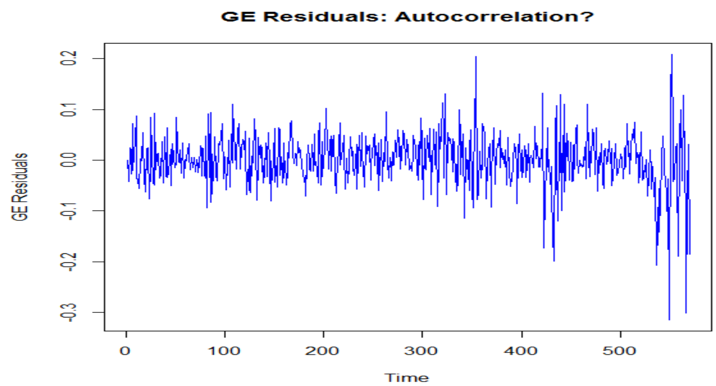


- It looks like a small ρ_1, but not very clear pattern from the graphs.

56

Finding Auto-correlation – Visual Check: GE

Example: Residual plot for the 3 factor F-F model for **GE returns**:

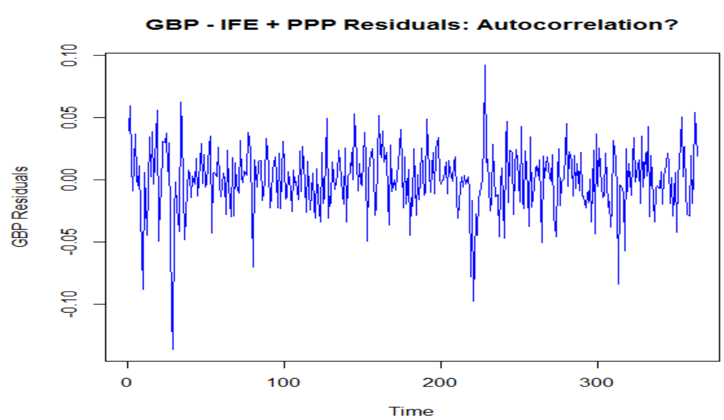


- It looks like a small ρ_1, but not very clear pattern from the graphs.

57

Finding Auto-correlation – Visual Check: GBP

Example: Residual plot for the encompassing model (IFE + PPP) for changes in the **USD/GBP**:



- Again, it looks like a small ρ_1, but not very clear pattern.

58

Testing for Autocorrelation: LM Test

• There are several autocorrelation tests. The AR(p) model to be tested is:

$$\varepsilon_t = \rho_1\, \varepsilon_{t-1} + \rho_2\, \varepsilon_{t-2} + \cdots + \rho_p\, \varepsilon_{t-p} + u_t$$

• Under the null hypothesis of no autocorrelation of order p, we have

H_0 (No autocorrelation): $\rho_1 = \ldots = \rho_p = 0$.

H_1 : At least one $\rho_i \neq 0$. $i = 1, 2, \ldots, p$

Under H_0, we can use OLS residuals, e_t.

• Breusch–Godfrey (1978) proposed an LM test for autocorrelations (BG test). It is very similar to the BP test.

59

Testing for Autocorrelation: LM Test

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

– **Step 1**. (Same as BP's Step 1). Run OLS on DGP:

$\mathbf{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$. - Keep residuals, e_t.

– **Step 2**. (Auxiliary Regression). Run the regression of e_t on all the explanatory variables, $\mathbf{X}$, and p lags of residuals, e_t:

$e_t = \boldsymbol{x}_t'\,\boldsymbol{\gamma} + \alpha_1\, e_{t-1} + \ldots + \alpha_p\, e_{t-p} + v_t$ - Keep R^2 (R_e^2)

– **Step 3**. Keep R_e^2. Then, calculate:

$\text{LM} = (T - p) * R_e^2 \xrightarrow{d} \chi_p^2$. $(T - p)$ = lost p observations

Note: In general, in **Step 2**, if we do not include $\boldsymbol{x}_t$, the LM test is not that different.

60

Testing for Autocorrelation: LM Test

Example: LM-AR Test for the 3 factor F-F model for **IBM returns** (p = **12 lags**):

```
e_ibm <- fit_ibm_ff3$residuals                    # OLS residuals
p_lag <- 12                                       # Select # of lags for test (set p)
e_lag <- matrix(0,T-p_lag,p_lag)                  # Matrix to collect lagged residuals
a <- 1
while (a<=p_lag) {                                # loop creates matrix (e_lag) with lagged e
  za <- e_ibm[a:(T-p_lag+a-1)]
   e_lag[,a] <- za
a <- a+1
}
Mkt_RF_p <- Mkt_RF[(p_lag+1):T]                   # Adjust for new sample size: T – p_lag
SMB_p <- SMB[(p_lag+1):T]
HML_p <- HML[(p_lag+1):T]
fit_ibm_ar <- lm(e_ibm[(p_lag+1):T] ~ e_lag + Mkt_RF_p + SMB_p + HML_p) # Aux R
r2_e1 <- summary(fit_ibm_ar)$r.squared            # get R² from Auxiliary Regression
```

61

Testing for Autocorrelation: LM Test

Example (continuation):

```
> r2_e1
[1] 0.0303721
> (T-p_lag)
[1] 557
lm_t <- (T-p_lag )* r2_e1                         # LM-test with p lags
> lm_t
[1] 16.91726
df <- ncol(e_lag)                                 # degrees of freedom for the LM Test
> 1-pchisq(lm_t,df)
[1] 0.1560063
```

LM-AR(**12**) Test: **16.91726** ⇒ cannot reject H_0 at 5% level (*p-value* > .05)

• If we run the test with p = **4 lags**, we get
LM-AR(4) Test: **2.9747** (*p-value* = 0.56) ⇒ cannot reject H_0 at 5% level

62

Testing for Autocorrelation: LM Test

Example (continuation):
The package *lmtest,* performs this test, *bgtest,* (and many others, used in this class, encompassing, jtest, waldtest, etc).

```
library(lmtest)
> bgtest(ibm_x ~ Mkt_RF + SMB + HML, order=12)

    Breusch-Godfrey test for serial correlation of order up to    12

data:  lr_ibm ~ Mkt_RF + SMB + HML
LM test = 16.259, df = 12, p-value = 0.1797
```

(minor difference with the previous test, likely due to multiplication by *T*. Results do not change much)

Note: If you do not include in the Auxiliary Regression the original regressors (Mkt_RF, SMB, HML) the test do not change much. You get LM-AR(**12**) Test: **16.83253** ⇒ very similar. Not entirely correct, but it works well.

63

Testing for Autocorrelation: LM Test

Example (continuation):
Autocorrelation is very common. If I run the test for Disney, CNP, or GE, instead, we get significant test results.

- For **DIS**:

```
lr_dis <- log(x_dis[-1]/x_dis[-T])
dis_x <- lr_dis – RF


> bgtest(dis_x ~ Mkt_RF + SMB + HML, order=4)
    Breusch-Godfrey test for serial correlation of order up to  4

data:  dis_x ~ Mkt_RF + SMB + HML
LM test = 8.6382, df = 4, p-value = 0.07081
```

⇒ cannot reject H_0 at 5% level (*p-value* >.05)

```
> bgtest(dis_x ~ Mkt_RF + SMB + HML, order=12)
    Breusch-Godfrey test for serial correlation of order up to  12

data:  dis_x ~ Mkt_RF + SMB + HML
LM test = 30.068, df = 12, p-value = 0.002728
```

⇒ reject H_0 at 5% level (*p-value* < .05)

64

Testing for Autocorrelation: LM Test

Example (continuation):
LM tests for autocorrelation (with 4 or 12 lags) for **GE** & **CNP** show significant test results:

```
lr_ge <- log(x_ge[-1]/x_ge[-T]); ge_x <- lr_ge – RF
lr_cnp <- log(x_cnp[-1]/x_cnp[-T]); cnp_x <- lr_cnp – RF
```

• For **GE**:

```
> bgtest(ge_x ~ Mkt_RF + SMB + HML, order=4)
      Breusch-Godfrey test for serial correlation of order up to  4

data:  ge_x ~ Mkt_RF + SMB + HML
LM test = 28.257, df = 4, p-value = 0.005073
```

⇒ cannot reject H_0 at 5% level (*p-value* >.05)

• For **CNP**:

```
> bgtest(cnp_x ~ Mkt_RF + SMB + HML, order=12)
      Breusch-Godfrey test for serial correlation of order up to  12

data:  cnp_x ~ Mkt_RF + SMB + HML
LM test = 31.718, df = 12, p-value = 0.00153
```

⇒ reject H_0 at 5% level (*p-value* < .05)

65

Testing for Autocorrelation: LM Test

• Q: How many lags are needed in the test?
A: Enough to make sure there is no auto-correlation left in the residuals.

There are some popular rule of thumbs: for daily data, 5 or 20 lags; for weekly, 4 or 12 lags; for monthly data, 12 lags; for quarterly data, 4 lags.

66

Testing for Autocorrelation: Durbin-Watson

• The Durbin-Watson (1950) (DW) test for AR(1) autocorrelation: H_0: $\rho_1 = 0$ against H_1: $\rho_1 \neq 0$. Based on simple correlations of $\boldsymbol{e}$.

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

• It is easy to show that when $T \rightarrow \infty$, $d \approx 2(1 - \rho_1)$.

• ρ_1 is estimated by the sample correlation r.

• Under H_0, ρ_1=0. Then, d should be distributed randomly around 2.

• Small values (close to 0) or Big values (close to 4) of d lead to rejection of H_0. The distribution depends on **X**. Since there are better tests, in practice, the DW is used "visually:" Is d close to 2?

The R function *dwtest* from the *lmtest* package produces also a *p-value*.

67

Testing for Autocorrelation: DW Test

Example: DW Test for the 3 factor F-F model for IBM returns

```
RSS <- sum(e_ibm^2)                                   # RSS
DW <- sum((e_ibm[1:(T-1)] - e_ibm[2:T])^2)/RSS        # DW stat
> DW
[1] 2.042728        ⇒ DW statistic ≈ 2  ⇒ No evidence for autocorrelation of order 1.
> 2 * (1 - cor(e_ibm[1:(T-1)], e_ibm[2:T]))           # approximate DW stat
[1] 2.048281
```

• Similar finding for Disney returns:

```
> DW
        [,1]
[1,] 2.1609          ⇒ DW statistic ≈ 2  ⇒  But, DIS suffers from autocorrelation!
```

⇒ This is why DW are not that informative. They only test for AR(1) in residuals.

Note: The package *lmtest* performs this test too, *dwtest*:

```
> dwtest(fit_ibm_ff3)
DW = 2.0427, p-value = 0.7087
```

68

Testing for Autocorrelation: DW Test

Example: DW Test for the residuals of the encompassing model (IFE + PPP) for changes in **USD/GBP**:

```
e_gbp <- fit_gbp$residuals
> dwtest(fit_gbp)

        Durbin-Watson test

data:  fit_gbp
DW = 1.8588, p-value = 0.08037
alternative hypothesis: true autocorrelation is greater than 0
```

⇒ not significant at 5% level.

69