Business School, Brunel University
MSc. EC5501/5509 Modelling Financial Decisions and Markets/Introduction to Quantitative Methods
Prof. Menelaos Karanasos (Room SS269, Tel. 01895265284)

Lecture Notes 6

1. Diagnostic (Misspecification) Tests: Testing the Assumptions and Consequences of their Failure

Having completed a discussion of the classical normal linear regression model, it seems natural to review each of the model's assumptions in turn. It is important to test the validity of the assumptions which underlie our model because when one or more assumptions fail our inference might be misleading. The properties of the least squares estimator depend on the assumptions of the CLRM. The derivation of the $t-$ and $F-tests$ also depends on these assumptions. Much of econometrics is about testing whether these assumptions hold. If they do not, then the model should be respecified and perhaps estimated by a different method, depending on the exact nature of the misspecification.

The various diagnostic tests will be presented in the context of the classical multiple linear regression model:

$$\begin{aligned} y_t &= \beta' x_t + \varepsilon_t, \ t = 1, ..., T \\ \text{or } y_t &= \beta_1 + \beta_{2t} x_{2t} + \beta_{3t} x_{3t} + ... + \beta_{kt} x_{kt} + \varepsilon_t. \end{aligned} \tag{1}$$

1.1. No Serial Correlation

When the error terms from different time periods (or cross-section observations) are correlated, we say that the error term is *serially correlated.* Serial correlation occurs in time-series studies when the errors associated with observations in a given time period carry over into future time periods. For example, if we are predicting the growth of stock dividends, an overestimate in one year is likely to lead to overestimates in succeeding years.

Serial correlation (also called *autocorrelation*) in the residuals means that they contain information, which should itself be modelled. First order serial correlation arises when the residuals in one time period are correlated directly with the residuals in the ensuing time period; second-order serial correlation refers to correlation between residuals two periods apart, third-order serial correlation refers to correlation between residuals three periods apart, and so on. (With annual data the most common form of serial correlation is first order, with quarterly data there is usually fourth order serial correlation.) A pattern where successive residuals tend to have the same sign indicates positive first order serial correlation, whereas a pattern where successive residuals tend to have opposite signs indicates negative first order serial correlation. (Observe that we need to use the words "tend to" because not all the residuals can have the same sign - they must sum to zero.)

The assumption of no serial correlation can be expressed as

$$E\left(\varepsilon_t \varepsilon_s\right) = 0, t \neq s, \text{ or } V\left(y_t/x_t\right) \equiv E\left(\varepsilon\varepsilon'\right) = \sigma^2 I.$$

On the other hand, the presence of serial correlation can be expressed as

$$E\left(\varepsilon_t \varepsilon_s\right) \neq 0, t \neq s, \text{ or } V\left(y_t/x_t\right) \equiv E\left(\varepsilon\varepsilon'\right) = \sigma^2 \Omega,$$

where the off-diagonal elements of Ω are not all equal to zero (the diagonal elements are all equal to one).

1.1.1. Testing for first order serial correlation: the Durbin-Watson statistic (DW)

The DW statistic is a test for the detection of first order serial correlation in the residuals. It is printed out by every econometric package, and is computed as:

$$DW = \frac{\sum\limits_{t=2}^{T} \left(\widehat{\varepsilon}_t - \widehat{\varepsilon}_{t-1}\right)^2}{\sum\limits_{t=2}^{T} \widehat{\varepsilon}_{t-1}^2} \cong 2\left(1 - \widehat{\rho}\right),$$

where $\widehat{\rho}$ is the first order autocorrelation coefficient of the residuals, i.e. it is the coefficient in the regression: $\widehat{\varepsilon}_t = \rho \widehat{\varepsilon}_{t-1} + u_t$, $-1 < \rho < 1$, $u_t \sim iid\left(0, \sigma_u^2\right)$. The null hypothesis is no serial correlation $\left(H_0 : \ \rho = 0\right)$. The DW statistic will lie in the 0 to 4 range, with a value near 2 indicating no first order serial correlation. The tabulated values of the DW statistic have one peculiar feature. For any significance level, sample size and number of regressors, two values of the statistic are tabulated: these are usually referred to as d_L (lower value) and d_U (upper value). Positive serial correlation is associated with DW values below 2 (in this case the alternative hypothesis is $H_1 : \rho > 0$). The decision rule is as follows:

$$\begin{array}{cccc} \text{If:} & \underbrace{0 < DW < d_L}; & \underbrace{d_L < DW < d_U}; & \underbrace{d_U < DW < 2} \\ \text{then:} & \text{Reject } H_0; & \text{No Conclusion}; & \text{Accept } H_0. \end{array}$$

Negative serial correlation is associated with DW values above 2 (i.e. the alternative hypothesis is $H_1 : \rho < 0$). In this case we subtract the value of the DW from 4 and proceed as follows:

$$\begin{array}{cccc} \text{If:} & \underbrace{0 < 4 - DW < d_L}; & \underbrace{d_L < 4 - DW < d_U}; & \underbrace{d_U < 4 - DW < 2} \\ \text{then:} & \text{Reject } H_0; & \text{No Conclusion}; & \text{Accept } H_0. \end{array}$$

Note: if $R^2 > DW$, then we are probably dealing with a *spurious regression.* That means that despite the apparent statistical significance of the explanatory variables there is no underlying relationship between the dependent and explanatory variables. Time series can appear highly correlated because of common trends and not because one affects the other (spurious correlation).

In the presence of a lagged endogenous variable in the model the DW statistic becomes unreliable. Instead we can use **Durbin's h statistic**:

$$h = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{T}{1 - Ts_l^2}},$$

where s_l^2 is the estimated variance of the coefficient of the lagged dependent variable. Since Durbin has shown that the h statistic is approximately normally distributed, the test for first order serial correlation can be done directly by using the normal distribution table.

To conclude, the drawbacks of the DW statistic are: (i) it can only test for first order serial correlation in the residuals, (ii) it is not reliable in the presence of lagged endogenous variables, (iii) no decision can be reached when it lies in the d_L to d_U range. As a result, it is better to adopt the testing procedure given below.

1.1.2. Testing for first and higher order serial correlation

We test by running the following (auxiliary) regression:

$$\begin{aligned}
\widehat{\varepsilon}_t &= \gamma_1 + \gamma_{2t}x_{2t} + ... + \gamma_{kt}x_{kt} + \sum_{i=1}^{N} \rho_i \widehat{\varepsilon}_{t-i} + u_t, \qquad (2) \\
H_0 &: \rho_i = 0,\ i = 1, ..., N \\
H_1 &: \rho_i \neq 0,
\end{aligned}$$

where $\widehat{\varepsilon}_t$ are the residuals of eq.(1). It is not difficult to see that model (2) is the unrestricted model, while model (1) is the restricted one. We can test the above hypotheses individually, using $t - tests$, or jointly using an $F - test$:

$$F - test = \frac{\left(\sum_t \widehat{\varepsilon}_t^2 - \sum_t \widehat{u}_t^2\right)/N}{\left(\sum_t \widehat{u}_t^2\right)/\left(T - k - N\right)} \sim F\left(N, T - k - N\right).$$

Alternatively, we can test the joint significance of the ρ's by using the asymptotic **Lagrange-multiplier test** $(LM-test)$. *The operational version of the test* is carried out by obtaining the product of the number of observations (T) and the coefficient of determination (R^2) of the auxiliary regression (2):

$$LM-test = TR^2 \sim \chi^2 (N).$$

Decision rule: if the test statistic is smaller than its critical value we cannot reject the null.

1.1.3. What are the effects of serial correlation?

Exactly what effect serial correlation has on the properties of the OLS estimator depends on why it arises:

- If the residuals are serially correlated because of serial correlation in the disturbances (the unsystematic part), then the OLS estimator $\left(\widehat{\beta}\right)$ remains unbiased and consistent but ceases to have minimum variance. In particular, OLS produces biased estimates of the standard errors of the coefficients; this renders hypothesis testing unreliable. In this case consistent estimators of the standard errors can be obtained by appropriately transforming the variables and then estimate the transformed model with OLS (this estimation procedure is called *feasible or estimated Generalized Least Squares* (GLS)); note that Microfit provides consistent standard errors on request.

- The most usual cause of serial correlation in the residuals is the omission of relevant variables as regressors (in other words, the cause of residual serial correlation lies in the systematic part of the regression). In this case the coefficient estimates themselves will be biased and inconsistent. The most obvious candidates for the omitted variables which produce serial correlation are lagged values of the regressors already included and of the dependent variable itself. Therefore, the appropriate way to deal with the problem of serial correlation is to respecify the model by including these lagged variables.

1.2. Homoscedasticity

There are occasions in econometric modeling when the assumption of constant variance, or *homoscedasticity*, will be unreasonable. For example, consider a cross section study of family income and expenditures. It seems plausible to expect that low-income individuals would spend at a rather steady rate, while the spending patterns of high-income families would be relatively volatile. This suggests that in a model where expenditures are the dependent variable, error variances associated with high-income families would be greater than their low-income counterparts. In other words, *heteroscedasticity* is present in the model.

The assumption of homoscedasticity can be expressed as

$$E\left(\varepsilon_t^2\right) = \sigma^2, \text{ for all } t, \text{ or } V\left(y_t/x_t\right) \equiv E\left(\varepsilon\varepsilon'\right) = \sigma^2 I.$$

On the other hand, heteroscedasticity can be expressed as

$$V\left(y_t/x_t\right) \equiv E\left(\varepsilon\varepsilon'\right) = \sigma^2\Omega,$$

where the diagonal elements of Ω are not all equal to one (the off-diagonal elements are all equal to zero).

1.2.1. Testing for heteroscedasticity

There are various types of tests depending on the nature of heteroscedasticity. In what follows we are going to examine two of the most commonly used tests.

- The **Reset-type test** involves the estimation of the following (auxiliary) regression:

$$\begin{aligned} \widehat{\varepsilon}_t^2 &= \gamma_0 + \gamma_1 \widehat{y}_t^2 + u_t, \qquad (3a) \\ H_0 &: \gamma_1 = 0, \\ H_1 &: \gamma_1 \neq 0, \end{aligned}$$

  i.e. we regress the squared residuals of model (1) on a constant and on the squared fitted values of model (1). Under the assumption of homoscedasticity, the slope coefficient of eq. (3a) is zero. We can test the statistical significance of γ_1 by using a $t - test$, or an $F - test$ (in this case the $F - test \sim F(1, T-2)$ and is given by the square of the $t - test$), or an $LM - test$ (in this case it follows a $\chi^2(1)$ distribution).

- *(This is optional)* The (autoregressive conditional heteroscedasticity test) **ARCH-test** involves the estimation of the following (auxiliary) regression:

$$\begin{aligned}\widehat{\varepsilon}_t^2 &= \gamma_0 + \gamma_1 \widehat{\varepsilon}_{t-1}^2 + \gamma_2 \widehat{\varepsilon}_{t-2}^2 + \dots + \gamma_p \widehat{\varepsilon}_{t-p}^2 + u_t \qquad (3b)\\ &= \gamma_0 + \sum_{i=1}^{p} \gamma_i \widehat{\varepsilon}_{t-i}^2 + u_t,\\ H_0 &: \gamma_i = 0,\ i = 1, 2, \dots, p\\ H_1 &: \gamma_i \neq 0.\end{aligned}$$

  We can use individual $t - tests$ or we can test the joint significance of the γ_i's with an $F - test$ and/or an $LM - test$:

$$\begin{aligned}F - test &= \frac{\left(\sum_t \widehat{v}_t^2 - \sum_t \widehat{u}_t^2\right)/p}{\left(\sum_t \widehat{u}_t^2\right)/(T-1-p)} \sim F(p, T-1-p),\\ LM - test &= TR^2 \sim \chi^2(p),\end{aligned}$$

  where $\widehat{v}_t$ is the residual series obtained from a regression of $\widehat{\varepsilon}_t$ on a constant.

1.2.2. What are the effects of heteroscedasticity?

When the residuals are heteroscedastic the OLS estimator $\left(\widehat{\beta}\right)$ remains unbiased and consistent but ceases to have minimum variance. In particular, OLS produces biased estimates of the standard errors of the coefficients; this renders hypothesis testing unreliable. In this case consistent estimators of the standard errors can be obtained by appropriately transforming the variables and then estimate the transformed model with OLS (this estimation procedure is called *feasible or estimated Generalized Least Squares* (GLS)); note that Microfit provides consistent standard errors on request.

1.3. Linearity

The assumption of linearity can be expressed as follows:

$$E(y_t/x_t) = \beta_1 + \beta_{2t} x_{2t} + \beta_{3t} x_{3t} + \dots + \beta_{kt} x_{kt}.$$

This specification is not as limiting as it might seem, because the linear regression model can be applied to a more general class of equations that are *inherently linear*. Inherently linear models can be expressed in a form that is linear in the parameters by a transformation of the variables. Inherently nonlinear models, on the other hand, cannot be transformed to the linear form. The non-linearities of interest here are the ones which cannot be accommodated into a linear conditional mean after transformation.

- One of the most common ways to test the linearity assumption is to use the **Reset-type test.** This testing procedure involves the estimation of the following (auxiliary) regression:

$$\begin{aligned} \widehat{\varepsilon}_t &= \gamma_1 + \gamma_{2t}x_{2t} + ... + \gamma_{kt}x_{kt} + \delta\widehat{y}_t^2 + u_t, \qquad (4) \\ H_0 &: \delta = 0, \\ H_1 &: \delta \neq 0. \end{aligned}$$

  It is not difficult to show that (4) is the unrestricted model, whereas model (1) is the restricted one. We can now test the statistical significance of δ by using the $t-test$ $\left(= \frac{\text{estimate of } \delta}{\text{standard error of } \widehat{\delta}}\right)$, or the $F-$ and/or $LM-tests$:

$$\begin{aligned} F-test &= \frac{\left(\sum_t \widehat{\varepsilon}_t^2 - \sum_t \widehat{u}_t^2\right)/1}{\left(\sum_t \widehat{u}_t^2\right)/(T-k-1)} \sim F\left(1, T-k-1\right), \\ LM-test &= TR^2 \sim \chi^2\left(1\right). \end{aligned}$$

- The failure of linearity has major consequences for our estimation. In particular, when linearity does not hold the OLS estimators are biased and inconsistent. In other words estimation and testing results are invalid and we need to respecify our model.

1.4. Normality

The assumption of normality can be expressed as follows:

$$\varepsilon_t \sim N\left(0, \sigma^2\right), \text{ or } \left(y_t/x_t\right) \sim N\left(\beta' x_t, \sigma^2\right).$$

If the assumption of normality does not hold, then the OLS estimator $\left(\widehat{\beta}\right)$ remains the Best Linear Unbiased Estimator (BLUE), i.e. it has the minimum variance among all linear unbiased estimators. It remains consistent, but is not the maximum likelihood estimator which can only be defined if a particular distribution is specified for y_t. However, without normality one cannot use the standard formulae for the t and F distributions to perform statistical tests. Fortunately, the central-limit theorem provides a rational for using standard statistical tests as approximately correct for reasonably large sample sizes.

Before we proceed with the normality tests let us specify the null hypothesis. The null is that the skewness (α_3) and kurtosis (α_4) coefficients of the conditional distribution of y_t (or, equivalently, of the distribution of ε_t) are 0 and 3, respectively:

$$\begin{aligned} H_0 : \ & \alpha_3 = 0, \quad (\text{if } \alpha_3 < 0 \text{ then } f\left(y_t/x_t\right) \text{ is skewed to the left}) \\ & \alpha_4 = 3, \quad (\text{if } \alpha_4 > 3 \text{ then } f\left(y_t/x_t\right) \text{ is leptokurtic}) \end{aligned}$$

(This is optional) The above assumptions can be tested jointly using the Jarque-Bera test (JB) which follows asymptotically a chi-square distribution:

$$\begin{aligned} JB - test &= \left[\frac{T}{6}\widehat{\alpha}_3^2 + \frac{T}{24}\left(\widehat{\alpha}_4 - 3\right)^2\right] \sim \chi^2\left(2\right), \\ \text{where } \widehat{\alpha}_3 &= \left[\left(\frac{1}{T}\sum_t \widehat{\varepsilon}_t^3\right) / \left(\frac{1}{T}\sum_t \widehat{\varepsilon}_t^2\right)^{3/2}\right], \\ \text{and } \widehat{\alpha}_4 &= \left[\left(\frac{1}{T}\sum_t \widehat{\varepsilon}_t^4\right) / \left(\frac{1}{T}\sum_t \widehat{\varepsilon}_t^2\right)^{2}\right]. \end{aligned}$$

Note that the $JB - test$ is sensitive to outliers. The above assumptions can also be tested individually, using the asymptotic distributions of $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$:

$$\begin{aligned} H_0 &: \alpha_3 = 0, \\ test - statistic &= \sqrt{\frac{T}{6}}\widehat{\alpha}_3 \sim N\left(0,1\right). \\ H_0 &: \alpha_4 = 3, \\ tets - statistic &= \sqrt{\frac{T}{24}}\left(\widehat{\alpha}_4 - 3\right) \sim N\left(0,1\right). \end{aligned}$$

1.5. No Perfect Multicollinearity: $Rank\,(X) = k$

If there is an exact linear relationship among the right-hand side variables of our model, then we say that we have the problem of **perfect multicollinearity**: $rank\,(X) < k$, and so $(X'X)$ is not invertible, and as a consequence **estimation of the model is not feasible.**

Multicollinearity arises when two or more variables (or combinations of variables) are highly (but not perfectly) correlated with each other. In this case the estimated coefficients $\left(\widehat{\beta}\right)$ remain unbiased but their standard errors get very large. Generally, worrying about multicollinearity does more damage than multicollinearity itself.

1.6. Parameter Time Invariance

Without loss of generality we are going to present and test the above assumption in the context of the following bivariate linear model:

$$y_t = \beta x_t + \varepsilon_t, \ t = 1, 2, ..., T. \tag{5}$$

Under the null hypothesis the parameters β and σ^2 remain constant throughout the sample period. Under the alternative hypothesis β and σ^2 change between two specified time periods:

$$\begin{aligned} y_t &= \alpha x_t + u_{1t}, \ u_{1t} \sim iid\left(0, \sigma_1^2\right), \ t = 1, 2, ..., T1; & \text{(6a)} \\ y_t &= \gamma x_t + u_{2t}, \ u_{2t} \sim iid\left(0, \sigma_2^2\right), \ t = T1+1, T1+2, ..., T. & \text{(6b)} \end{aligned}$$

It is easy to see that the number of observations of the first subsample is $T1$, whereas the number of observations of the second subsample is $T2 = T - T1$. Below we present three tests depending on whether β, σ^2, or both may change.

- Testing **Variance Equality** $(H_0: \ \sigma_1^2 = \sigma_2^2)$:

$$\left(\frac{s_1^2}{s_2^2}\right) \sim F\left(T1 - k, T2 - k\right),$$

  where $s_1^2 = \frac{\sum \widehat{u}_{1t}^2}{T1-k}$, $s_2^2 = \frac{\sum \widehat{u}_{2t}^2}{T2-k}$, and k is the number of coefficients we estimate in eq. (5) (note that in this case $k = 1$). The larger variance should be used as the numerator. This is the *Goldfeld-Quandt Variance Ratio test,* for heteroscedasticity of a very specific type; it is sensitive to the failure of the normality assumption.

- Testing **Coefficient Equality** $(H_0: \ \alpha = \gamma)$ conditional on variance equality:

$$F - test = \frac{\left(\sum_{t=1}^{T} \widehat{\varepsilon}_t^2 - \sum_{t=1}^{T1} \widehat{u}_{1t}^2 - \sum_{t=T1+1}^{T} \widehat{u}_{2t}^2\right)/k}{\left(\sum_{t=1}^{T1} \widehat{u}_{1t}^2 + \sum_{t=T1+1}^{T} \widehat{u}_{2t}^2\right)/\left(T - 2k\right)} \sim F\left(k, T - 2k\right).$$

  Note that in the context of model (5) $k = 1$. The unrestricted model comprises of eq. (6a) and (6b), while the restricted model is given by eq. (5). The above is *Chow's test* for coefficient equality and it can be used when $T1 > k$ and $T2 > k$.

- Testing **Predictive Failure** (i.e. whether the model of the first subsample predicts the second subsample):

$$F - test = \frac{\left(\sum_{t=1}^{T} \widehat{\varepsilon}_t^2 - \sum_{t=1}^{T1} \widehat{u}_{1t}^2 \right) / T2}{\left(\sum_{t=1}^{T1} \widehat{u}_{1t}^2 \right) / (T1 - k)} \sim F\left(T2, T1 - k\right).$$

The above test is particularly useful when the observations of the 2nd subsample do not allow us to estimate the model. In this case eq. (5) is the restricted model, whereas eq. (6a) is the unrestricted one.

- Observe that the all three tests above assume that you know the point at which the parameters change, i.e. you know when the *structural break* occurs. If you do not, then Microfit provides the **CUSUM** and **CUSUMSQ** plots to test for structural stability. If the plots cross the two lines denoting the 95% confidence interval bounds, they indicate that there has been a structural shift.