Note 2 to Computer class: Standard mis-specification tests*

Ragnar Nymoen

September 2, 2013

1 Why mis-specification testing of econometric models?

As econometricians we must relate to the fact that the "data generating process", DGP, which has produced the data, is not the same as the econometric model that we have specified and estimated. The situation is very different from the case where the data is from a laboratory experiment. Then the DGP can in principle be regarded as known (anything else can be seen a result of "bad experimental design") since the experiment has been devised and supervised by the researchers themselves. The situation that an experimental researcher is in can be thought of as follows:

$$\underset{\text{result}}{Y_i} = \underset{\text{input}}{g(X_i)} + \underset{\text{shock}}{v_i}. \tag{1}$$

The variable Y_i is the result of the experiment, while the X_i is the imputed input variable which is decided by the researcher. $g(X_i)$ is a deterministic function. The variable v_i is a shock which leads to some separate variation in Y_i for the chosen X_i. The aim of the experiment is to find the effect that X has as a causal variable on Y. If the $g(X_i)$-function is linear, this causal relationship can be investigated with the use of OLS estimation.

In economics, the use of experimental data is increasing, but still the bulk of applied econometric analysis makes use of non-experimental data. Non-experimental economic data is usually collected for other purposes than research and the data reflect the real-life decisions made by a vast number of heterogenous agents. Hence, the starting point of an econometric modelling project is usually fundamentally different from the statistical anaysis of experimental data. In order to maintain (1) as a "model of econometrics" for this kind of data, we have to invoke the *axiom of correct specification*, meaning that we know the DGP before the analysis is made.

If we want to avoid the axiom of correct specification then, instead of (1), we need to write

$$\underset{\text{observed}}{Y_i} = \underset{\text{explained}}{f(X_i)} + \underset{\text{remainder}}{\varepsilon_i} \tag{2}$$

where Y_i are the observations of the dependent variable which we seek to explain by the use of economic theory and our knowledge of the subject matter. Our explanation is given by the function $f(X_i)$ which in the regression case can be characteristic

*This note is a translated extract from Chapter 8 in Bårdsen and Nymoen (2011). It also draws on Hendry (1995, Ch 1) and Hendry and Nielsen (2007, Ch 11).

precisely as the conditional expectation function. The non-experimental Y_i is not determined or caused by $f(X_i)$, it is determined by a DGP that is unknown for us, and all variation in Y_i that we do not account for, must therefore "end up" in the remainder ε_i. Unlike (1), where v_i represents free and independent variation to Y_i, ε_i in (2) is an *implied* variable which gets its properties from the DGP and the explanation, in effect from the model $f(X_i)$. Hence in econometrics, we should write:

$$\varepsilon_i = Y_i - f(X_i) \tag{3}$$

to describe that whatever we do on the right hand side of (3) by way of changing the specification of $f(X_i)$ or by changing the measurement of Y_i, the left-hand side is derived as a result.

This analysis poses at least two important questions. The first is related to causation: Although we can make $f(X_i)$ precise, as a conditional expectation function, we cannot claim that X_i is a causal variable. Again this is different from the experimental case. However, as we shall see, we can often combine economic theory and further econometric analysis of the *system* that contains Y_i and X_i as endogenous variable, to reach a meaningful interpretation of the joint evidence in favour of one-way causality, or two-way causality. Recently, there has also been a surge in micro data sets based on large registers, which has opened up a new approach to causal modelling based on *natural experiments* and "difference in differences estimators".[1]

The second major issue, is potential model mis-specification and how to discover mis-specification if it occurs. Residual mis-specification, in particular, is defined relatively to the *classical regression model* . Hence we say that there is residual mis-specification if the residual from the model behaves significantly differently from what we would expect to see if the true disturbances of the model adhered to the classical assumptions about homoscedasticity, non-autocorrelation, or no cross-sectional dependence.

Clearly, if the axiom of correct specification holds, we would see little evidence of residual mis-specification. However, even the smallest experience of applied econometrics will show that mis-specification frequently happens. As we know from elementary econometrics, the consequences for mis-specification for the properties of estimators and tests are sometimes not very serious. For example, non-normality alone only entails that there are problems with knowing the exact distribution of for example the $t-$statistic in small sample sizes. There are ways around this problem, by use of robust methods for covariance estimation. Other forms of mis-specification gives more serious problems: Autocorrelated disturbances in a dynamic model may for example lead to coefficient estimators being biased, even in very large samples (i.e., they are inconsistent).

[1] An elementary exposition is found in for example Hill et al. (2008, Ch 7.51). Stewart and Wallis (1981) is an early textbook presentation of the basic form of this estimator (see page 180-184), but without the label *Difference in differences*, which is a much more recent innovation.

The following table gives and overview, and can serve as a review of what we know from elementary econometrics.

	Disturbances ε_i are:			
X_i	heteroscedastic		autocorrelated	
X_i	$\hat{\beta}_1$	$\widehat{Var}(\hat{\beta}_1)$	$\hat{\beta}_1$	$\widehat{Var}(\hat{\beta}_1)$
exogenous	unbiased consistent	wrong	unbiased consistent	wrong
predetermined	biased consistent	wrong	biased inconsistent	wrong

Here we have in mind a linear model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

and $\hat{\beta}_1$ is the OLS estimator for the slope coefficient β_1 and $\widehat{Var}(\hat{\beta}_1)$ is the standard error of the estimator. The entry "wrong" indicates that this estimator of the variance of $\hat{\beta}_1$is not the correct estimator to use, it can overestimate or underestimate the uncertainty.

We assume that we estimate by OLS because we are interested in β_1 as a parameter in the conditional expectation function. This means that we can regard X_i as exogenous in the sense that all the disturbances are uncorrelated with X_i. There is one important exception, and that is when we have time series data and X_t is the lag of Y_t,i.e., we have Y_{t-1} on the right hand side. In this case X_i in the table is not exogenous but pre-determined: It is uncorrelated with future disturbances, but not ε_{t-1}, ε_{t-2}, and so on backward.

Because of its importance in the assessment of the quality of econometric models, most programs contains a battery of mis-specification test. *PcGive* is no exception, and in fact *PcGive* reports such tests in the default output.

The output (the default) is a little different for cross section and time series models, and for simplicity we show examples of both types of models, and comment on the differences. We give reference Davidson and MacKinnon's *Econometric theory and Methods*, and to the book by Hill, Griffit and Lim used in ECON-3150/4150 and to Bårdsen and Nymoen (2011), this may be useful for Norwegian students since it has a separate chapter on mis-specification testing.

2 Mis-specification tests for cross-section data

We take the simple regression on the *konsum_sim* data set as our example, see the note called *Seminar_PcGive_intro.pdf*:

```
---- PcGive 13.20 session started at 20:49:38 on 16-08-2011 ----


EQ( 1) Modelling C by OLS-CS
     The dataset is: D:\sw20\ECON4160\H2011\Data\KonsDataSim\konsum_sim.xls
     The estimation sample is: 1960 - 2006

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant              50.9079      17.67     2.88  0.0060   0.1558
I                    0.805148    0.03685     21.8  0.0000   0.9139

sigma                 51.7618  RSS                  120567.778
R^2                  0.913855  F(1,45) =     477.4 [0.000]**
Adj.R^2              0.911941  log-likelihood        -251.161
no. of observations        47  no. of parameters            2
mean(C)               399.842  se(C)                  174.431

Normality test:   Chi^2(2)   =   3.1025 [0.2120]
Hetero test:      F(2,44)    =   3.6511 [0.0341]*
Hetero-X test:    F(2,44)    =   3.6511 [0.0341]*
RESET23 test:     F(2,43)    =   1.2975 [0.2837]
```

The default mis-specification tests are at the bottom of the screen-capture.

Normality test

The normality assumption for the disturbances is important for the exact statistical distribution of OLS estimators and the associated test statistics. Concretely: Which "p-values" to use for t-tests and F-tests and for confidence intervals and prediction intervals.

If the normality assumption holds, it is correct inference to use the t-distribution to test hypothesis about single parameters of the models, and the F-distribution to test joint hypothesis.

If the normality assumption cannot be maintained,inference with the t- and F-distribution is no longer exact, but it can still be a good approximation. And it get increasingly good with increasing sample size.

In the output above, the normality test is Chi-square distributed with two degrees of freedom, $\chi^2(2)$, reported as `Normality test: Chi^2(2)`.

This test is based on the two moments $\hat{\kappa}_3^2 = \sum \hat{\varepsilon}_i^3/\hat{\sigma}^3$ (skewness) and $\hat{\kappa}_4^2 = \sum \hat{\varepsilon}_i^4/\hat{\sigma}^4 - 3$ (kurtosis) where $\hat{\varepsilon}_i$ denote a *residual* from the estimated model. Skewness refers to how symmetric the residuals are around zero. Kurtosis refers to the "peakedness" of the distribution. For a normal distribution the kurtosis value is 3. These two moments are used to construct the test statistics

$$\chi^2_{\text{skew}} = n\frac{\hat{\kappa}_3^2}{6} \quad \chi^2_{\text{kurt}} = n\frac{\hat{\kappa}_4^2}{24} \quad \text{and, jointly} \quad \chi^2_{\text{norm}} = \chi^2_{\text{skew}} + \chi^2_{\text{kurt}}$$

with degrees of freedom 1, 1 and 2 under the null hypothesis of normality of ε_i. As you can guess, χ^2_{norm} corresponds to `Normality test: Chi^2(2)` in the Screen Capture. The p-value is in brackets and refers to the joint null of no skewness and no excess kurtosis. As you can see to reject that null you would have to accept a significance level of the test of 0.2120. Hence, there is no formal evidence of non-normality for this model.

PcGive calculates the skewness and kurtosis moments, but they not reported as part of the default output. To access the more detailed information click *Model-Test* from the main menu and then check the box for *Tests..*, click *OK* and in the next menu check for *Normality test* and click *OK*.

The χ^2_{norm}- statistics is often referred to as the *Jarque-Bera*-test due Jarque and Bera (1980).

Textbook references:

- Davidson and MacKinnon: The unnumbered section "Tests for Skewness and kurtosis" page 660—664. The exposition is rather technical. It refers for example to the concept of a Outer-Product.of the Gradient (OPG) estimator,which we do not assume familiarity with in ECON 4160. Nevertheless note equation (15.34) on page 663, and that they too refer to this joint test as the Jarque-Bera test.

- Textbook in ECON 4150: Hill, Griffits and Lim, p 89

- Textbook in Norwegian: Bårdsen and Nymoen p 199-200

Heteroscedasticity tests (White-test)

Formal tests of the homoscedasticity assumption were proposed by, White (1980), so these tests are often referred to as White-tests. In the simplest case, which we have here. the test is based on the *auxiliary regression*:

$$\hat{\varepsilon}_i^2 = a_0 + a_1 X_i + a_2 X_i^2, \;\; i = 1, 2, ..., n, \tag{4}$$

where, as stated above, the $\hat{\varepsilon}_i$'s are the OLS residuals from the model. Under the null hypothesis of homoscedasticity we have

$$H_0\text{: } a_1 = a_2 = 0$$

which can be tested by the usual F-test on (4). This statistic, which we will refer to by the symbol F_{het}, is then F-distributed with 2 and $n-3$ degrees of freedom under the H_0. n denotes the number of observations. In our example, this means that we use $F(2, 47-3)$, i.e., $F(2, 44)$ and it is reported as `Hetero test:  F(2,44)` in the screen capture. Note that you would reject the null hypothesis at the 5 % level based on this evidence, but not reject at the stricter 2.5 % level.

You will often see in textbooks that there are Chi-square distributed versions of the mis-specification tests that are based on auxiliary regressions. This is the case for White's test, which is distributed $\chi^2(2)$ in the present example. It is calculated as nR_{het}^2, where R_{het}^2 is the multiple correlation coefficient from (4). From elementary econometrics we know the F-distributed statistic can be written as

$$F_{het} = \frac{R_{het}^2}{(1-R_{het}^2)} \frac{n-3}{2}.$$

confirming that the two version of the test use the same basic information and that the difference is that the F-version "adjusts for" degrees of freedom. Usually the effect is to keep control over the level (or size of the test) so that the p-values are not overstated.

In PcGive you get the nR_{het}^2 version of the test by using *Model-Test* from the main menu and then check the box for *Tests..*, click *OK* and in the next menu check for *Heteroscedasticity test (using squares)* and click *OK*.

With two or more explanatory variables there is an extended version of White's test that includes cross-products of the regressors in the auxiliary regression. In the screen capture, this test is Hetero-X test: F(2,44). Since we have one regressor it is identical to the first test. If we include a second regressor in the model the test would be `Hetero-X test: F(5,41)` since the auxiliary regression contains X_{1i},X_{2i}, X_{1i}^2,X_{2i}^2 and $X_{1i}X_{2i}$.

Textbook references:

- Davidson and MacKinnon: Section 7.5. The Gauss-Newton Regression (GNR) they refer to is the same as what we have called the "auxilliary regression" above. The theoretical motivation for GNR is given in Chapter 6 Non-linear regression (the concept is introduced on page 236), which is very useful as a reference, although we do not require any detailed knowledge about numerical methods of otimization in this course.
- Textbook used in ECON 4150: Hill, Griffits and Lim, p 215
- Textbook in Norwegian: Bårdsen and Nymoen p 196—197

Regression Specification Error Test, RESET

The RESET test in the last line of the screen capture is based the auxiliary regression $\hat{Y}_i$

$$Y_i = a_0 + a_1 X_i + a_2 \hat{Y}_i^2 + a_3 \hat{Y}_i^3 + v_i, \quad i = 1, 2, ..., n, \tag{5}$$

where $\hat{Y}_i$ denotes the fitted values.

`RESET23 test` indicates that there is both a squared and a cubic term in (5) so that the joint null hypothesis is: $a_2 = a_3 = 0$. If you access the Model-Test menu, you also get the RESET test that only includes the squares $\hat{Y}_i^2$. Note that there are χ^2 distributed versions of both tests.

As the name suggests, the RESET test is sometimes interpreted as a test of the correctness of the model, the functional form in particular. However, most modern textbook now stress that the RESET test is nonconstructive (i.e., it gives no indication about "what to do if the test is significant"). Hence, the consensus is to interpret the RESET test as a general mis-specification test.

Textbook references:

- Davidson and MacKinnon. Section 15.2
- Textbook in ECON 4150: Hill, Griffits and Lim, p 151—152
- Textbook in Norwegian: Bårdsen and Nymoen p 197—199

3 Mis-specification tests for time series data

We can re-estimate the same regression as above but as an explicit model for time series data. Follow the instructions in *Seminar_PcGive_intro.pdf* to obtain:

The only difference is that we have two new mis-specification tests, labelled `AR 1-2 test` and `ARCH 1-1 test` in the output. This gives us a double message: First that all the mis-specification tests cross-section data, are equally relevant for models that use time series data. Second that there are special mis-specification issues for time series data. This is because of three features. First, with time series

```
EQ( 2) Modelling C by OLS
      The dataset is: D:\sw20\ECON4160\H2011\Data\KonsDataSim\konsum_sim.xls
      The estimation sample is: 1960 - 2006

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant              50.9079      17.67     2.88  0.0060   0.1558
I                    0.805148    0.03685     21.8  0.0000   0.9139

sigma                 51.7618  RSS                  120567.778
R^2                  0.913855  F(1,45) =     477.4 [0.000]**
Adj.R^2              0.911941  log-likelihood       -251.161
no. of observations        47  no. of parameters           2
mean(C)               399.842  se(C)                 174.431

AR 1-2 test:      F(2,43)   =  0.14906 [0.8620]
ARCH 1-1 test:    F(1,45)   =  0.31010 [0.5804]
Normality test:   Chi^2(2)  =   3.1025 [0.2120]
Hetero test:      F(2,44)   =   3.6511 [0.0341]*
Hetero-X test:    F(2,44)   =   3.6511 [0.0341]*
RESET23 test:     F(2,43)   =   1.2975 [0.2837]
```

data, we have a natural ordering of the observations, from the past to the present. Second, time series data are usually autocorrelated, meaning that Y_t is correlated with Y_{t-1}, Y_{t-2} and usually also longer lags (and leads). Third, unless $f(X_t)$ in (2), interpreted as a time series model, explains all the autocorrelation in Y_t, there will be residual autocorrelation in ε_t, meaning that the classical assumption about uncorrelated disturbances does not hold.

Residual autocorrelation

`AR 1-2 test` is a standard test of autocorrelation up to degree 2. It tests the joint hypothesis that $\hat{\varepsilon}_t$ is uncorrelated with $\hat{\varepsilon}_{t-j}$ for any choice of j, against the alternative that $\hat{\varepsilon}_t$ is correlated with $\hat{\varepsilon}_{t-1}$ or $\hat{\varepsilon}_{t-2}$. The test makes use of the auxiliary regression

$$\hat{\varepsilon}_t = a_0 + a_1\hat{\varepsilon}_{t-1} + a_2\hat{\varepsilon}_{t-2} + a_3 X_t + v_t \tag{6}$$

and the null hypothesis tested is

$$H_0 \, : a_1 = a_2 = 0 \, .$$

Many textbooks (Greene and Hill, Griffits and Lim also) refer to this (rather technically) as the "Lagrange multiplier test", but then one should add "for autocorrelation" since also the other tests can be interpreted statistically as Lagrange multiplier tests.

As noted by Bårdsen and Nymoen (2011), several researchers have contributed to this test for autoregression, notably Godfrey (1978) and Harvey (1981, side 173). Based on the evidence (note the F distribution again, the χ^2 form is available from the *Model-Test* menu), there is no sign of autocorrelation in this case.

This test is flexible. If you have reason to believe that the likely form of autocorrelation is of the first degree, it is efficient to base the test on an auxiliary regression with only a single lag. Extension to higher order autocorrelation is also straight forward and is easily done in the *Model-Test menu* PcGive.

Importantly, the test is also valid for dynamic model, where Y_{t-1} is among the explanatory variables. This is not the case for the older Durbin-Watson test for example (which still can be found in *Model-Test* menu though).

Textbook references:

- Davidson and MacKinnon, section 7.7
- Textbook used in ECON 4150: Hill, Griffits and Lim, p 242
- Textbook in Norwegian: Bårdsen and Nymoen p 193—196

Autoregressive Conditional Heteroscedasticity (ARCH)

With time series data it is possible that the variance of ε_t is non-constant. If the variance follows an autoregressive model of the first order, this type of heteroscedasticity is represented as

$$Var(\varepsilon_t \,|\varepsilon_{t-1}) = a_0 + \alpha_1 \varepsilon_{t-1}^2$$

The null hypothesis of constant variance can be tested by using the auxiliary regression:

$$\hat{\varepsilon}_t^2 = a_0 + a_1 \hat{\varepsilon}_{t-1}^2 + v_t, \qquad (7)$$

where $\hat{\varepsilon}_t^2$ $(t = 1, 2, ..., T)$ are squared residuals. The coefficient of determination, R_{arch}^2, from (7) is used to calculate TR_{arch}^2 which is $\chi^2(1)$ under the null hypothesis. In the same way as many of the other test, the F-form of the test is however preferred, as also the screen-capture above shows. Extensions to higher order residual ARCH are done in the *Model-Test* menu.

We use the ARCH model as a mis-specification test here, but this class of model has become widely used for modelling volatile time series, especially in finance. The ARCH model is due to Engle (1982).

Textbook references:

- Davidson and MacKinnon, section 13.6, page 589
- Textbook in ECON 4150: Hill, Griffits and Lim, p 369
- Textbook in Norwegian: Bårdsen and Nymoen p 197—199

References

Bårdsen, G. and R. Nymoen (2011). *Innføring i økonometri*. Fagbokforlaget.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Godfrey, L. G. (1978). Testing for Higher Order Serial Correlation When the Regressors Include Lagged Dependent Variables. *Econometrica*, *46*, 1303–1313.

Greene, W. (2012). *Econometric Analysis*. Pearson, 7th edn.

Harvey, A. C. (1981). *The Econometric Analysis of Time Series*. Philip Allan, Oxford.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford University Press, Oxford.

Hendry, D. F. and B. Nielsen (2007). *Econometric Modeling*. Princeton University Press, Princeton and Oxford.

Hill, R. C., W. Griffiths and G. Lim (2008). *Principles of Econometrics*. Wiley, 3rd edn.

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters*, *6*, 255–259.

Stewart, M. B. and K. F. Wallis (1981). *Introductory Econometrics*. Basil Blackwell.

White, J. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity. *Econometrica*, *48*, 817–838.