EC4090: Quantitative Methods
EXERCISE 2
Question 2

Mario Hofer

February 12, 2006

In this exercise we estimate the influence of share (SP) and bond prices (BP) on the personal share holdings (AS). Plotting each of these variables against time clearly indicates that those data series follow an upward trend. In order to avoid a spurious regression a deterministic trend (TIME) will be included in the model. In order to answer the questions it is convenient to formulate the model in a more general way:

$$\mathbf{y} = \gamma_1 \boldsymbol{\iota} + \gamma_2 \mathbf{T} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{1}$$

where $\boldsymbol{\iota}$ is a unit vector representing the constant in the regression, $\mathbf{T} = [\,1 \vdots 2 \vdots 3 \vdots 4 \vdots \ldots\,]$ and $\mathbf{X}$ is the matrix with the regressors.

In order to give a meaningful interpretation of the coefficients of the following regressions we would need more information on the variables, e.g. the measurement units. Therefore the main emphasis in this exercise is put on the last question, which is the comparison of the coefficients and the residuals of the two regressions. The detailed output of the all regressions can be found in the appendix.

1. Create a deterministic trend (TIME) and a constant (CON). Regress AS on CON TIME BP SP.

   The coefficients of all variables (TIME, BP and SP) are significant on a 10% significance level. Furthermore those three variables are also jointly significant on a 1% significance level. Because there is no information available on the measurement units of the variables it is hard to tell whether the coefficients show any economic significance.

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
CON	-1480.8	5331.2	-.27777[.783]
TIME	143.6120	82.0954	1.7493[.090]
SP	30.4988	2.6230	11.6275[.000]
BP	110.8780	60.8705	1.8215[.078]
R-Squared	.97213		
Normality	2.1098[.348]		

The normality test conducted by Microfit uses the skewness and the kurtosis of a random variable to test whether a random variable is distributed normally. H_0 states that the random variable is distributed normally. As one can observe from the regression results, we can not reject this null hypothesis neither on a 5% nor on a 10% significance level.

The R^2 in this regression is extremly high. But this is not uncommon for time-series data, especially if the dependent variable follows a trend (cf. Wooldridge (2003), p351). Wooldridge (2003) uses the adjusted R^2 to show why this is the case.

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}_u^2}{\hat{\sigma}_y^2}$$

with $\hat{\sigma}_u^2$ being an unbiased estimator of the error variance and $\hat{\sigma}_y^2 = \frac{SST}{(n-1)}$. If the dependent variable follows a linear trend then $\frac{SST}{(n-1)}$ is no longer unbiased (cf. Wooldridge (2003), p352).

2. Run auxiliary regressions AS on CON TIME, BP on CON TIME and SP on CON TIME and store the residuals respectively RAS, RBP and RSP. In each regression the coefficient on TIME is statistially significant at the 1% level.

3. Run the regression RAS on RBP and RSP

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
RSP	30.4988	2.5423	11.9967[.000]
RBP	110.8780	58.9971	1.8794[.069]

4. Compare the estimated coefficients and the residuals from 1. with those from 3..

Both the estimated coefficients and the residuals in 1. and 3. are exactly the same. The correlation coefficient of RM1 and RM2 is exactly 1. This comes from the fact the auxiliary regressions in 2. partial out the time trend in the data series. Therefore RAS, RBP and RSP can be viewed as being linearly detrended. This "detrending process" is an application of the FWL-Theorem described by Davidson and MacKinnon (2004). In the following section we will show how the FWL-Theorem can be used to detrend a data series. For this purpose it is convenient to write (1) as follows:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\Gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

with $\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$ and $\mathbf{Z} = [\boldsymbol{\iota}\ \mathbf{T}]$.

According to the FWL-Theorem the OLS estimates and the residuals from (2) are numerically identical to the ones obtained by carrying out the following regression:

$$\mathbf{M_Z y} = \mathbf{M_Z X}\boldsymbol{\beta} + residuals \tag{3}$$

(a) **Proof that the estimates of** (2) **and** (3) **are numerically identical**

According to the standard OLS formula the OLS estimates for (3) can be calculated by

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= ((\mathbf{M_Z X})^\top \mathbf{M_Z X})^{-1} (\mathbf{M_Z X})^\top \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{M_Z X})^{-1} \mathbf{X}^\top \mathbf{M_Z y} \end{aligned} \tag{4}$$

The right equality follows from the fact that the transpose of a product is the product of the transposes in reverse order and from the fact that $\mathbf{M_Z}$ is an idempotent and symmetric matrix (which we have already shown on the first exercise sheet).

With some slight modifications shown in Davidson and MacKinnon (2004) (2) can be rewritten as

$$\mathbf{y} = \mathbf{Z}\hat{\boldsymbol{\Gamma}} + \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{M_{(XZ)} y} \tag{5}$$

where $\mathbf{M_{(XZ)}}$ is the projection matrix for the combined group of regressors. Premultiplying both sides of the equation with $\mathbf{X}^\top \mathbf{M_Z}$ yields

$$\mathbf{X}^\top \mathbf{M_Z y} = \mathbf{X}^\top \mathbf{M_Z Z}\hat{\boldsymbol{\Gamma}} + \mathbf{X}^\top \mathbf{M_Z X}\hat{\boldsymbol{\beta}} + \mathbf{X}^\top \mathbf{M_Z M_{(XZ)} y} \tag{6}$$

Considering that $\mathbf{M_Z Z} = \mathbf{0}$ and $\mathbf{X}^\top \mathbf{M_Z M_{(XZ)} y} = (\mathbf{M_{(XZ)} M_Z X})^\top \mathbf{y} = (\mathbf{0X})^\top \mathbf{y} = \mathbf{0}$, equation (6) simplifies to

$$\mathbf{X}^\top \mathbf{M_Z y} = \mathbf{X}^\top \mathbf{M_Z X}\hat{\boldsymbol{\beta}} \tag{7}$$

We can now easily solve (6) for $\hat{\boldsymbol{\beta}}$.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{M_Z X})^{-1} \mathbf{X}^\top \mathbf{M_Z y} \tag{8}$$

which is exactly the same expression als (4). Therefore we have shown that (2) and (3) yield numerically identical estimates. By premultiplying (5) with $\mathbf{M_Z}$ we can easily show that the residuals are also numerically identeical. In order to explain why the regressions of 1. and 3. yield the same results we have to proof that the final regression in 3. is an FWL-regression for the model stated in 1.

(b) **Proof that the regression in 3. is an FWL-regression**

In order to accomplish this proof we analyse the individual parts of (3) which was

$$\mathbf{M_Z y} = \mathbf{M_Z X}\boldsymbol{\beta} + residuals$$

We start with the left-hand side:

$$\mathbf{M_Z y} = (\mathbf{I} - \mathbf{P_Z})\mathbf{y} = \mathbf{y} - \mathbf{P_Z y}$$

Because $\mathbf{P_Z}$ projects on to $\mathcal{S}(\mathbf{Z})$ the result of the left-hand side of above equation is a vector in $\mathcal{S}(\mathbf{Z})$, the subspace spanned by $\boldsymbol{\iota}$ and $\mathbf{T}$. Therefore we can conclude that $\mathbf{P_Z} = \mathbf{Z}\hat{\boldsymbol{\Gamma}}^{\boldsymbol{*}}$. This means that $\mathbf{P_Z}$ is from a regression of $\mathbf{y}$ on $\mathbf{Z}$. Therefore $\mathbf{M_Z y}$ are the residuals from such a regression. That corresponds to the regression of AS on CON and TIME, which is exactly the first of our auxiliary regressions.

Now we take a look at the right-hand side:

$$\mathbf{M_Z X}\boldsymbol{\beta} = (\mathbf{I} - \mathbf{P_Z})\mathbf{X}\boldsymbol{\beta} = (\mathbf{X} - \mathbf{P_Z X})\boldsymbol{\beta}$$

Similar to above $(\mathbf{X} - \mathbf{P_Z X})$ are OLS residuals, but this time they come from a regression of $\mathbf{X}$ on $\mathbf{Z}$. In our example this corresponds to the residuals from regressing BP on CON and TIME and to the residuals from regressing SP on CON and TIME.

Therefore we have shown that our final regression in 3. is an FWL-regression and therefore it yields exactly the same coefficients and residuals as our original model.

Bibliography

Davidson, R., MacKinnon, J. (2004): *Econometric Theory and Methods*, Oxford University Press

Wooldridge, J. (2003): *Introductory Econometrics*, Thomson South-Western