Applied Panel Econometrics (MPP-E1161) - Winter Term 2015
Prof. Dr. Kerstin Bernoth

Take Home Exam 1

Max Callaghan

October 2015

1 Part I: Basic Questions [14pt: each 2pt]

Briefly explain why your chosen answer is correct.

1. **Question**

   False or true: To judge, whether adding an additional explanatory variable improves the model, I check, whether the R^2 increases.

   Answer

   This is true to an extent. The R^2 describes how much of the variation in the dependent variable is explained by variation in the independent variables. Therefore, when the R^2 increases, our model is explaining more of the variation in the independent variable. Nevertheless, one should avoid 'R^2 Maximising', especially if explanatory variables are added to the model without theoretical justification. Adding explanatory variables will *always* increase the R^2, even if this is just due to chance. The Adjusted R^2 adjusts R^2 according to the number of predictors, and can help in avoiding overfitting the model

2. **Question**

   False or true: When the panel-specific factors a_i are significant, the fixed-effects estimator will be preferred over pooled OLS.

   Answer

   True. In a fixed effects model, time-invariant factors are captured by a_i. Where these are significant we can say that the fixed effects captured by a_i are unobservable time-invariant differences across individuals. If these are present, then it is likely that the assumption of i.i.d of our error terms is very likely violated in a pooled OLS model. It is also likely, if the unobservable characteristics of each individual panel observation are correlated with the observable characteristics in our model, tat the assumption of exogeneity of the regressors is also very likely violated.

3. **Question**

   False or true: I cannot reject a Null-hypothesis on a single coefficient, when the absolute value of the t-statistic is smaller than the critical value.

   Answer

   True. At a given sample size and probability threshold, the critical value states the minimum value of t at which we can reject a Null-hypothesis. If t is above the critical value, we can reject the Null-hypothesis that the coefficient is not statistically significant, and if it is lower, we cannot reject it.

4. **Question**

   False or true: The OLS estimators β_h become more accurate, when the variance of the independent variable x_{hit} decreases

   Answer

   The accuracy of the esitamated coefficient is determined by

   $$\hat{Var}(\hat{\beta}_h) = \frac{\hat{\sigma}_\epsilon^2 \frac{1}{NT}}{(1-R)_h^2 Var(x_{it})}$$

   for $k = 1, \ldots, k$.

   Therefore, when the variance decreases, the estimators β_h become less accurate.

5. **Question**

   Calculate the R^2 of a regression, of which the variance of the realized variable y is equal to 6.4 and te variance of the fitted dependent variable $\hat{y}$ is 5.8

   Answer

6. **Question**

   False or true: Including an irrelevant variable does not lead to biased OLS coefficients.

   Answer

   Including irrelevant variables does not lead to biased OLS coefficients in the relevant variables. It can only make the model less efficient. If the true model is y

7. **Question**

   False or true: When a regressor is endogenous, we have the situation that the dependent variable y_{it} is correlated with the residual $\epsilon_{it}, cov(y_{it}, \epsilon_{it}) \neq 0$.

   Answer

2 Part 2: GDP Growth and Investment [28pt]

1. **Question**

   Calculate the variances of the GDP growth ($Var(y_{it})$) and investment ($Var(x_{it})$) and their covariance ($Cov(x_{it}, y_{it})$).

   Answer

   - Population variance of GDP Growth: $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{N}(y_{it} - \bar{y}^2) = 63.75$
   - Population variance of Investment: $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{N}(x_{it} - \bar{x}^2) = 9.93$

2. **Question**

   Calculate the OLS solution for $\hat{\beta}_0$ and $\hat{\beta}_1$ and write-up the regression equation.

   Answer

   $\hat{\beta}_1 = \frac{Cov(x_{it}, y_{it})}{Var(x_{it})}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

   $\hat{\beta}_1 = 0.335744$ and $\hat{\beta}_0 = -3.9314909$

   $GDPgrowth_{it} = -3.93 + 0.34 Investment_{it}$

3. **Question**

   Give an interpretation of the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$.

   Answer

   The estimated coefficients tell us that the model predicts that when x (Investment) is 0, y (GDP growth) will be $\hat{\beta}_0 = -3.93$. For every increase in x, y is predicted to increase by 0.34. The model estimates a positive effect of Investment on GDP growth.

4. **Question**

   Add the fitted values of $y_{it}, \hat{y}_{it} = \hat{\beta}_0 + \hat{\beta}_1 x_{it}$, to the table above.

   Answer

   See excel sheet.

5. **Question**

   Calculate the residuals and add them into the table above.

   Answer

   See excel sheet. $\epsilon_{it} = \hat{y}_{it} - y_{it}$

6. **Question**

   Draw the regression line into the graph below and mark explicitly, where the regression line cross the y axis and the slope.

   Answer

   The graph given in the exam is substituted for a similar graph produced by excel that extends the y axis so that the intercept can be drawn.

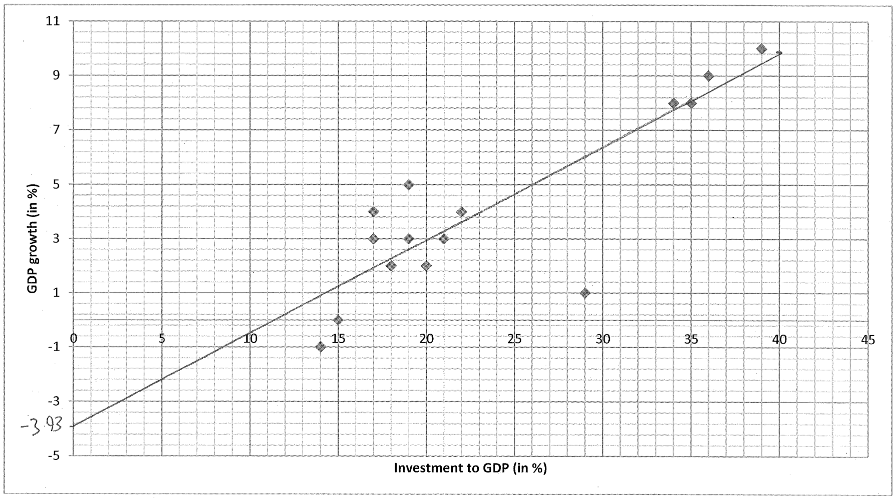

Figure 1: Intercept = -3.93, Slope = 0.34

7. **Question**

   Use graphical inspection to judge, whether this regression suffers from heteroscedasticity.

   Answer

   It is difficult to make a judgement with so few data points, but the variance at high values of x seems to be smaller than the variance at low values of x. This suggests that the data is, at least to some extent, heteroscedastic.

8. **Question**

   Calculate the standard errors of $\hat{\beta}_1$ (Note: The standard error is the square root of their variances: $s.e.\hat{\beta}_1 = \sqrt{\hat{Var}(\hat{\beta}_1)}$).

   Answer

$$s.e.\beta_1 = \sqrt{\frac{\frac{1}{n-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\epsilon}_{it}^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2}} = 0.055$$

9. **Question**

   Perform a t-test at the 5% significance level to test the hypothesis that the level of investment has a positive impact on GDP growth. Thus, the Null Hypothesis is $H_0 : \beta_1 = 0$. Interpret the result.

   Answer

   The t-statistic of our hypothesis is given by

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_1^0}{s.e.\hat{\beta}_1} = \frac{0.34 - 0}{0.055} = 6.18$$

Our degrees of freedom are $df = NT - k = 16 - 1 = 15$

With 15 degrees of freedom, the citical value for a t-test at the 5% significance level is 2.131.

Since 6.18 is larger than 2.131, we can reject the null hypothesis that the level of investment has no impact on GDP growth.

10. **Question**

    Calculate the R^2 of the regression and interpret the result (Hint: Choose the formula, which seems most convenient given the variables you already have at hand).

    Answer

    $R^2 = 1 - \frac{Var(\hat{\epsilon})}{Var(y)} = 0.72$

3 Part 3: Current Account Imbalances and Exchange Rate Regimes [37pt]

3.1 Regression preparation [6pt]

(a) **Question**

Explain briefly, why we should concentrate on the *absolute* value and not the level of the current account to measure current account imbalances.

Answer

We want to look at current account imbalances. Variation either side of 0 is an imbalance, so we look at the absolute value, and take that as the size of the current account imbalance, regardless of direction.

(b) (i) **Question**

abs_cagdp: A variable that depicts the absolute value of *cagdp* ($|cagdp|$) (Hint: Use Stata help to find out the command to calculate absolute values).

Answer

Stata code/output

```
cap gen abs_cagdp = abs(cagdp)
```

(ii) **Question**

trade_openness: A variable that measures nominal export relative to GDP plus nominal import to GDP ($imports + exports$).

Answer

Stata code/output

```
cap gen trade_openness = importsgdp + exportsgdp
```

(c) **Question**

We want to focus only on industrial countries. Thus, drop all non-industrial countries from the data set. Indicate, how many observations you have left in your data set.

Answer

Stata code/output

```
drop if ind == 0
(1,465 observations deleted)
```

3.2 Model Choice [8pt]

To test your hypothesis, you want to run a pooled OLS regresssion with *abs_gdp* as dependent variable and regime, *exportsgdp, trade_openess, finance, gdpgrowth* as independent variables

(a) **Question**

Explain, why we do not regress simply the *abs_cagdp* on *regime*, but also add other control variables to the set of regressors? What could be the consequences for the estimators, if one ignores important explanatory variables

Answer

If regime type was the only variable that we thought had an effect on *abs_cagdp*, we would only include regime type. If other variables affect our dependent variable and we do not include them, we risk *omitted variable bias*. When we miss out an independent variable that is both a determinant of the dependent variable *and* correlated with one or more of our independent variables, the coefficient for the included independent variable will be biased because it will contain part of the effects of omitted variables that are correlated with our included variable.

(a) **Question**

Explain, what the consequences are, if we have a multicollinearity problem among the explanatory variables

Answer

Multicollinearity describes a model where the independent variables are correlated with each other. In a case of perfect multicollinearity, there is a linear relationship between at least two of the variables.

The variance of the estimated coefficients is given by

$$\hat{Var}(\hat{\beta}_h) = \frac{\hat{\sigma}^2_{\epsilon}\frac{1}{NT}}{(1-R^2_h)Var(x_{it})} \text{ for } k = 1, \ldots, k$$

The greater the relationship between x_{it} and the other independent variables (given by R^2_h), the greater the variance of our estimator. Multicollinearity makes our estimators less precise.

(a) **Question**

Test, whether we have a multicollinearity problem between our explanatory variables. Explain, how you derive your conclusion.

Answer

Stata calculates the Variance Inflation Factor (VIF) for us using the formula

$$VIF(\hat{\beta}_h) = \frac{1}{1 - R_h^2}$$

High values for VIF indicate the presence of multicollinearity. 10 is sometimes used as a cutoff point.

Stata code/output

```
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
trade_open~s |    118.93    0.008408
  exportsgdp |    116.98    0.008548
     finance |      1.40    0.712628
      regime |      1.20    0.834933
   gdpgrowth |      1.13    0.886961
-------------+----------------------
    Mean VIF |     47.93
```

Unsurprisingly, we can detect multicollinearity problems in the high VIF values for both trade openness and exportsgdp. This is because we created trade openness out of a linear combination of exports and imports. So, notwithstanding some degree of variation between the balance or otherwise of exports and imports, they are highly correlated

(a) **Question**

In case you do detect multicollinear variables: What would you recomment to do? Which of the potential multicollinear variables would you drop?

Answer

We created the variable *tradeopenness*, presumably, because we thought that the combination of both exports and imports had an effect on our dependent variable. If we drop it, we are left with just imports; if we drop imports, than we still have the effect of the combination of exports and imports. An alternative would be to use imports and exports separately, but they would most likely still be highly correlated, though to a lesser extent.

3.3 Estimating a pooled OLS regression [15pt]

(a) **Question**

Continue with your preferred set of regressors (meaning: drop if necessary a variable(s) to solve for a potential multicollinearity problem (see Question 2.c)) and estimate the determinants of *abs_cagdp* with *pooled OLS*.

Answer

Stata code/output

```
reg abs_cagdp regime trade_openness finance gdpgrowth
outtex, file(pooled.tex) labels level detail ///
  legend key(stab) replace
```

Table 1: Estimation results : regress

Variable	**Coefficient**	**(Std. Err.)**
0b.regime	0.000	(0.000)
1.regime	-1.342**	(0.305)
2.regime	-1.363**	(0.316)
trade_openness	0.029**	(0.002)
finance	0.025	(0.091)
gdpgrowth	0.101*	(0.049)
Intercept	2.236**	(0.318)
N	818	
R^2	0.218	
$F_{(5,812)}$	45.368	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

(a) **Question**

Indicate, whether the 'Friedman hypothesis' is supported by the data or not.

Answer

The 'Friedman hypothesis' suggested that "fixed exchange rate ragimes should facilitate the build-up of current account imbalances" We create dummy variables of the 'intermediate' and 'fixed regime' types, compared to the reference category of 'floating'. Increases in our dependent variable *abs_cagdp* indicate that the current account imbalance is larger. Therefore, if Friedman's hypothesis were correct, the coefficient of the dummy variable for a fixed regime would be positive. The data show a statistically significant negative effect of a fixed regime type compared to a floating regime type on current account imbalance. The data therefore suggest that we reject Friedman's hypothesis.

(a) **Question**

Indicate, which of the other control variables have a significant impact on current account imbalances (assume a significance level of 5%).

Answer

Trade openness and GDP growth both have statistically significant positive effects on current account imbalances.

(a) **Question**

Give a numerical interpretation of the estimated coefficients.

Answer

The constant is 1.841728, so if all variables were equal to 0, the model would predict an absolute current account as a percentage of GDP (*abs_cagdp*) value of 1.84.

Compared to the reference category 'floating', observations that are in the category 'intermediate' are predicted to have a *abs_cagdp* that is lower by 1.34. Observations that are in the category 'fixed' are predicted to have a *abs_cagdp* that is lower by 1.36.

A one unit increase in *finance* is predicts a 0.02 increase in *abs_cagdp*

A one unit increase in GDP growth predicts a .1 unit increase in *abs_cagdp*.

(a) **Question**

Explain whther the signs of the estimated coefficients make sense from an economic point of view.

Answer

3.4 Estimating a fixed-effects regression [8pt]

(a) **Question**

Perform a Breusch-Pagan LM test for heteroscedasticity and interpret the result.

Answer

Stata code/output

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of abs_cagdp

         chi2(1)      =    48.68
         Prob > chi2  =   0.0000
```

(a) **Question**

Estimate your regression model this time with a *fixed-effects panel* estimation.

Answer

Stata code/output

```
xtreg abs_cagdp regime trade_openness finance gdpgrowth, fe
outtex, file(fe.tex) labels level detail ///
        legend key(stab) replace
```

Table 2: Estimation results : xtreg

Variable	Coefficient	(Std. Err.)
0b.regime	0.000	(0.000)
1.regime	-1.194**	(0.355)
2.regime	-0.358	(0.424)
trade_openness	0.034**	(0.008)
finance	0.279**	(0.104)
gdpgrowth	0.071	(0.045)
Intercept	1.239†	(0.649)
N	818	
R^2	0.09	
$F_{(33,784)}$	15.473	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

(a) **Question**

Indicate again, whether the 'Friedman hypothesis' is supported this time by the data or not.

Answer

In this case, the sign is still negative, contradicting the 'Friedman hypothesis'. However, the coefficient for the dummy variable 'fixed' is not statistically significant, so we cannot reject Friedman's hypothesis that there is a positive relationship.

(a) **Question**

What could explain the different results you find for the variable *regime* once you estimate a fixed-effects regression instead of pooled OLS?

Answer

4 Model Interpretation [6pt]

(a) **Question**

Answer

(a) **Question**

Answer

(a) **Question**

Answer