

Second and final test in Introductory
Econometrics

Robert M. Kunst

November 21, 2018

1. For a static linear regression in a time-series context $y_t = \beta_0 + \beta_1 x_{t,1} + \dots + \beta_k x_{t,k} + u_t$, five assumptions are often invoked in order to guarantee nice properties of the OLS estimator.[6+9=15 points]
 - (a) Write down the five assumptions, mark those that are needed for unbiasedness and (we know some additional technicalities must be assumed here but we ignore this feature for the moment) for consistency.
 - (b) If y_{t-1} is inserted as an additional regressor, the regression becomes dynamic. Which of the assumptions listed in (a) cannot hold any more in this context, and which assumption is used instead?
 - (c) Even with the modified assumption, OLS will be biased, but consistency is preserved. When is an estimator called consistent? (It is enough to explain the concept, a formal statistical definition is not required)
 - (d) When is a variable ('process') called (covariance-) stationary?

2. The accompanying printout shows a dynamic regression model that targets an explanation of goods imports (lm in logarithmic form). You can see that imports are explained by their past value $L1.lm$, by current and lagged values of domestic demand (vd and $L1.vd$), and by current and lagged values of the so-called ‘terms of trade’, which is just the ratio of import and export prices (tt and $L1.tt$ and $L2.tt$). (You may assume that all our assumptions hold here, including normal distribution, maybe excepting no serial correlation, as it is tested in (c)) [5+5+4=14 points]
- (a) The researcher is not quite satisfied with this specification, as the significance of two regressors is poor. For this reason, they are subjected to an F-test that is shown after the main regression printout. If a 10% significance level is used throughout, can the null be rejected? Should we eliminate the two regressors? [for simplicity, assume the statistic is F-distributed under its null, although we should know that the F-distribution is not guaranteed in a dynamic regression]
- (b) For this F-test, formulate the null hypothesis and alternative in our ‘beta’ notation.
- (c) The researcher knows that the dynamic regression equation should be well specified and tests for serial correlation. A Breusch-Godfrey test is reported (called *bgod*) for three lags, and the corresponding p-value. Again using 10%, does the test reject? Does this mean that it is OK to use the equation?

```
. reg lm l.lm l(0/1).vd l(0/2)tt
```

Source	SS	df	MS	Number of obs	=	21
Model	1.02545925	6	.170909876	*****		
Residual	.010075115	14	.000719651	*****		
Total	1.03553437	20	.051776718	R-squared	=	0.9903
				Adj R-squared	=	0.9861
				Root MSE	=	.02683

lm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lm						
L1.	.6276227	.2279925	2.75	0.016	.1386274	1.116618
vd						
--.	2.057892	.5876279	3.50	0.004	.7975553	3.318228
L1.	-1.179823	.8315508	-1.42	0.178	-2.963322	.6036761
tt						
--.	1.561905	.6538597	2.39	0.032	.1595153	2.964294
L1.	-2.603739	.8516193	-3.06	0.009	-4.430281	-.7771972
L2.	1.204985	.6350317	1.90	0.079	-.1570227	2.566992
_cons	-3.220029	3.013543	-1.07	0.303	-9.683435	3.243377

```
. test l.vd l2.tt
```

```
( 1) L.vd = 0
( 2) L2.tt = 0
```

```
F( 2, 14) = 2.15
p-value = 0.1537
```

```
. estat bgod,lags(3)
```

```
Breusch-Godfrey LM test
```

lags(p)	chi2	df	p-value
3	5.741	3	0.1249

3. A researcher is interested in whether happiness is rather determined by income and economics or rather by other factors. The dependent variable Y is a happiness index. Explanatory variables are household income INC , a dummy for marriage status MAR , age A , and a psychological summary index of personality traits $BIG5$. Observations correspond to persons. [10 points]
- (a) First consider the errors from such a regression u . Can you name some effects that may be contained in these errors?
 - (b) Someone thinks that INC may be endogenous, as u may contain the health status, and healthy people are rewarded by a higher salary on average. What conditions must a variable for non-wage income $INCP$ fulfill in order to be a valid instrument for INC ? [$MAR, A, BIG5$ are assumed exogenous.]
 - (c) After running the IV regression, the researcher runs the over-identification test and is disappointed. Why?
 - (d) There is also another test, Hausman's endogeneity test. For the IV regression to make sense, should this Hausman test rather reject or not?

4. Two basic and important conceptual issues.[6 points]

- (a) If two regressors are correlated, this will usually not result in a violation of the non-multicollinearity condition of Gauss-Markov and related theorems. Explain in brief why not.
- (b) What is the difference between errors u and residuals \hat{u} ?