Statistics Review

FBE 464
1/10 &1/17 2001
Professor Hans O. Mikkelsen

1

Economic Theory

- Typically, you are given a theory like:
  $Y_t = X_t$
  where Y_t and X_t are variables measured at time t.
- You want to know if empirical evidence:
  - Supports the theory.
  - With statistical confidence.
- **Time series statistics!**

2

Coefficients

- You want to test the theory, so you add a parameter and an intercept:
  $Y_t = b_0 + b_1 X_t$
- Economic theory specifies:
  $b_0 = 0$ and $b_1 = 1$
- Given some data you have to figure out what values of b_0 and b_1 are supported empirically.

3

One observation

- Suppose in Jan. 2001 Y_t=100 and X_t=100.
- This implies that $Y_t = 0 + 1*X_t$
  => Perfect fit!
- However, $Y_t = 100 + 0*X_t$ is also a perfect fit.
- Therefore, we can not identify both b_0 and b_1 with only one observation.

4

Two observations

- Suppose in Dec. 2000 Y_t=50 and X_t=50 , in Jan. 2001 Y_t=100 and X_t=100.
- Here there is only one solution:
  $Y_t = 0 + 1*X_t$
- The system is exactly identified and b_0=0 and b_1=1 is in accordance with theory.
- In general you need at least as many observations as you have parameters.

5

Many observations

- It is very imprecise to base anything on only two observations.
- Anything unusual happening during those obs., like an Asian crisis, is very unlikely to be informative about the whole population.
- Also, it can be shown that precision increases with the number of observations.
- Collect as much data as possible!

6

Independent observations

- With more observations than parameters you can not get a perfect fit, however.
  - Unless the observations are not independent (the previously mentioned two obs. are not because the second is just twice the first).
- Therefore, you expect to make an error at any point in time no matter how you choose the parameters.

7

Error term

- There are other reasons why you don't expect theory to hold at any point in time.
  - There may be measurement errors.
  - Differences in timing.
  - Model is too simple (leaves something out).

=> Add an error term, u_t:

$$Y_t = b_0 + b_1 X_t + u_t$$

8

Estimation

- How can plausible values for the parameters be obtained, that is, how can the model be estimated?
- Sample of T observations on both variables.
  - $Y_1, Y_2, ..., Y_T; X_1, X_2, ..., X_T$
- It is easy to obtain parameters so that the average error is zero.

9

Estimating the model

- It turns out that the best (easiest) estimation procedure is OLS.
- Ordinary Least Squares (OLS).
  - Minimizes the sum of squared residuals (errors)
  - => Minimizes the variance of regression errors.
- This avoids making too large negative or positive errors.

10

Estimation (cont.)

- Fortunately, if you include a constant the error term will have a mean of zero by construction.
- That is, not only do we not make too large errors, on average we make no error at all.
- There are other methods for estimation.
  - All you need to know is that for your purpose OLS is best.

11

Example

- Excel can do OLS!
- Suppose the regression output is:

  $Y_t = 0.2 + 0.7{*}X_t + u_t$
- Do you conclude that theory fails because b_0 is not 0 and b_1 not 1?
- You don't know, there is uncertainty because we only observe a snapshot, say five years, and with noise.

12

Standard errors

- Fortunately OLS (and Excel) also provides us with standard errors of the estimated parameters.
- The larger standard errors, the more uncertainty.
- If the standard error on b_1 is 0, there is no uncertainty and we can be 100% certain that b_1=0.7, and theory is rejected

13

Confidence intervals

- Assume we have many observations.
- Then, with 95% confidence we can say that a parameter is within an interval of plus, minus 2 std. dev. from its estimate.
- For b_1: $0.7 - 2*0.1 < b_1 < 0.7 + 2*0.1$
  => $0.5 < b_1 < 0.9$
- Because this interval does not include b_1=1, it is very likely that theory fails.

14

Confidence intervals

- Assume that the standard error for the estimate of b_0 is 0.5.
- Then the 95% confidence interval for b_0 is:
  $0.2 - 2*0.5 < b_0 < 0.2 + 2*0.5$
  => $-0.8 < b_0 < 1.2$
- Because this interval includes 0 it is very likely that b_0 is in accordance with theory.

15

T-test.

- It is more convenient to test theory using t-tests.
- If you want to test if an estimated parameter is significantly different from a given value, compute:

  $t = (b-b^*)/s_b$
- where b is your estimate, b* is the value of interest, and s_b is the standard error of b.

16

t-test: decision rule

- You will find all you need from the regression output.
- Two sided test:
  - Reject the null hypothesis of b=b* if |t| is large.
- How large:
  - The t-test is t-distributed, and you can find a table in your favorite text book (or the handout).

17

t-test: example.

- Estimated b_1 is 0.7
- Estimated standard error of b_1 is 0.1
- You want to test H_0: b_1=1.0
- Test statistic is:

  $t = (0.7-1.0)/0.1 = -3.0$
- If you have many obs., say 100, the 5% critical value is very close to 2.00

18

t-test: example. (cont.)

- because |t|=3.00 > 2.00, you reject the null.
- The estimated parameter is not 1.
- Had the t-test been 1.5, you would not reject that b_1=1, since |t|=1.50 < 2.00.

19

t-test: How to find critical value

- A 5% significance level in a 2-sided test means that we have to cut 2.5% in both tails of the distribution.
- We find the 97.5% percentile to be approx. 2.0
- We know that a t-distribution is symmetric. so the 2.5% percentile is -2.0
- That's why |t| is a 5% test.

20

Another example

- You want to test if the constant is significantly different from zero.
  - Estimated b_0 is 0.2
  - Std. error of b_0 is 0.5
- t = (0.2-0.0)/0.5 = 0.4
- |t|=0.4 < 2.0 => Not statistically significant.
- Cannot reject H_0: b_0=0

21

Comparison

- Confidence intervals and t-tests tell exactly the same story.
- In fact, the reason why 2 is used as critical value for the t-test in large samples is the same as why we subtract and add 2 from the estimate to get a confidence interval.

22

Sub-set of parameters

- Suppose you have several explanatory variables in your model.
- Instead of testing hypotheses on some parameters individually, you may want to test a joint hypothesis like:
  - H_0: $b_1 = b_2 = 0$
- Individual and joint test results may differ.
- Use an F-test.

23

Insignificant variables

- Get rid of them!
- In a finite sample, parameters that are not significantly different from zero:
  - Don't don't belong in the model.
  - Affects estimates of all other parameters.

=> **Only trust your results if all parameters are significantly different from zero.**

24

R^2

- R^2 is a measure of fit for the estimated model.
- R^2 = fraction of variation in Y_t that is explained by right hand side (RHS) variables.
- R^2=0 means no relation at all between Y_t and RHS => Bad model.
- R^2=1 indicates perfect fit => Perfect model.

25

R^2 - Problem

- PROBLEM: You don't know what a good fit is, and there is no test.
- R^2=0.05 may be a reasonable fit for many financial models.
- R^2=0.9 may not be good fit in other models

26

Plots

- To judge the fit of a model you can plot Y_t against the values predicted by the model.
- Y_t will always be more volatile than predictions.
- However, maybe the predictions capture the direction of most changes in Y_t.
- <u>**Always trust tests before R^2 and plots.**</u>

27

Bad news

- The bad news is that OLS only works optimally if the error term (residual) is so-called "white noise".
- There are three conditions for white noise:
  1) Mean zero: $E_t[u_{t+1}] = 0$, all t
  2) Constant variance: $E_t[u_{t+1}{}^2] = s^2$, all t
  3) No serial correlation: $E_t[u_{t+k}|u_{t-n}] = 0$, all t,n,k

28

Conditions

- The 1st condition is satisfied when using OLS, and including a constant.
- Violations of the 2nd condition are very difficult to handle.

  => Ignore them for the purpose of this course!

29

Serial correlation

- Violations of the 3rd condition are <u>very</u> serious!
- However, you can typically detect and correct for serial correlation.
- Often you find first order serial correlation:
  - $E_t[u_{t+1}|u_t]$ is different from zero.

30

Ex.: 1st order serial correlation

- $E_t[u_{t+1}|u_t] = 0.5*u_t$

  => The correlation between u_t and u_{t+1} is 0.5
- Alternatively, the first order autocorrelation coefficient of the residuals is 0.5.
- E.g., if the current error is 10% we expect the next error to be 0.5*10%=5%.

31

Implications of ser. corr.

- If the models prediction was 10% above the actual outcome, we expect it to be 5% above next period.
- That is, given this periods error we can forecast next period's error.
- The model fails to take some structure into account when there is serial correlation.
  - Model could and should be improved.

32

Rational expectations

- Often u_t is a rational expectations error.
  - Therefore, there cannot be serial correlation.
- If there was, I would be making systematic expectations errors. NOT rational!
  - If I'm 10% wrong today it is not rational to know that I will be 5% wrong tomorrow, without taken the information into account!

33

Statistical implications

- If there is serial correlation, not just first order, standard errors of parameter estimates are wrong.
  => You can't make correct inference.
- All we do involves testing some hypotheses.
  – Without correct inference this is an impossible task.

34

Durbin-Watson test

- Test for <u>first</u> order serial correlation.
- Null hypothesis: no serial correlation.
- DW = $2(1-r_1)$
  where r_1 is the first order autocorrelation coefficient of the residuals.
- The above expression is an approximation.
- Critical values for DW can be found in almost any introductory time series book.

35

Durbin-Watson: Rules

- High first order serial correlation means:
  DW=$2(1-r_1)$ is small.
- Hence, small values of DW rejects the null hypothesis of no serial correlation.
- If DW is sufficiently low you have serial correlation.
- If DW is sufficiently high you don't.

36

Durbin-Watson: example

- First order autocorrelation is 0.5
  => DW = 2(1-0.5) = 1.0
- According to my textbook:
  – DW_L=1.65 and DW_U=1.69 for our sample size of T=100, and 2 right hand side variables in my regression.
- Since DW is lower than the lowest critical value, we have detected serial correlation.

37

Other examples

- Had DW been between the two critical values, say 1.67, the test would be inconclusive.
- Had DW been above the upper critical value, say 2.00, we would have no serial correlation.

38

Notice:

- The described DW procedure is a test against positive serial correlation.
  – r_1 is positive.
- If you want to test negative serial correlation you should use: 4-DW.
- I.e. if 4-DW is sufficiently low, you have a problem of negative first order serial correlation.

39

Correcting for serial correlation

- If you detect serial correlation in the residuals, theory may already be wrong!
  - E.g., rational expectations ass. may be wrong.
- Otherwise, re-specify your model.
  - Include a lagged dependent variable on the RHS:
    $Y_t = b_0 + b_1X_t + b_2Y_{t-1} + u_t$.
- Often the serial correlation goes away.

40

Serial correlation persists

- Re-specify again!
  - include more lags of Y on the right hand side.
- If this doesn't work, transform your variables:
  - For example, estimate the model in differences.
- If nothing works there are only complicated methods:
  - Not for FBE 464!

41

Trending Variables.

- Variables that "explodes" over time.
- Maybe they:
  - get very, very high values.
  - just continue to increase.
  - never seem to revert towards a natural level.
- That gives us special statistical and numerical problems. **Variables are non-stationary and need to be transformed.**

42

Log of variables

- Instead of the variable itself, e.g. Y, we often work on ln(Y).
  => Very large numbers get smaller.
- Also, macro and financial variables are typically log normally distributed (roughly).
  => ln(variable) is normally distributed

43

Growth rates

- One way to get rid of the trends discussed before is by working on growth rates:
  Growth rate 1 of $Y_t = (Y_t - Y_{t-1})/Y_{t-1}$
- Since we often use $\ln(Y_t)$, a more convenient way to compute growth rates is:
  Growth rate 2 of $Y_t = \ln(Y_t) - \ln(Y_{t-1})$

44

2 types of growth rates

- If growth rates are small, growth rates 1 and 2 are very similar.
- If Y_{t-1}=55 and Y_t=56
  => G1: Growth rate 1 = (56-55)/55 = 0.0182
  G2: Growth rate 2 = ln(56)-ln(55) = 0.0180
- However, if Y_{t-1}=55 and Y_t=100
  => G1 = (100-55)/55 = 0.8182
  G2 = ln(100)-ln(55) = 0.5978

45

Use log growth rates

- Suppose Y_{t-2}=100, Y_{t-1}=200, and Y_t=100.
- Then $G1_{t-1}$=1.0 and $G1_t$=-0.5.

  => You will conclude that Y changes by an average of 0.25 or 25% per period, although it ends up unchanged!
- $G2_{t-1}$=0.69 and $G2_t$=-0.69.

  => You conclude that on average Y does not change at all, which is the truth!

46

The End.

47