DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

Econometrics 2, Class 1

Problem Set #2
September 19, 2005

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

Remember!

- Send an email to let me know that you are following these classes:
  paul.sharp@econ.ku.dk
- That way I can contact you e.g. if I need to cancel class

#2.1 Monte Carlo Simulation

What is a Monte Carlo simulation?

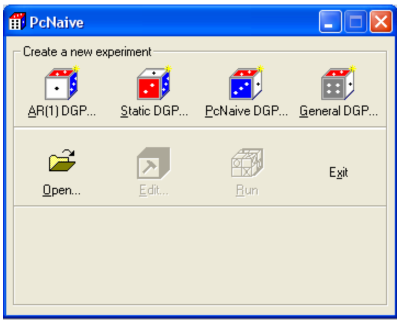
- Named after the city in Monaco and its roulette wheels (simple random number generators)
- We generate an artificial dataset with known statistical properties (or use a known distribution)
- We can then test for these properties within sub-samples using our econometric methods
- Example (Econometrics 1): We know that in the presence of endogeneity of the explanatory variables, the OLS estimators are biased. $E\left[\hat{\beta}\right] \neq \beta$

  With a Monte Carlo simulation we can generate a dataset where we know the true β, thus allowing us to check the extent of this bias by performing OLS on several sub-datasets drawn from the generated dataset.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

PcNaive

- PcNaive is a module in GiveWin (like PcGive)
- It can be accessed by clicking on Modules – Start PcNaive
- Remember licence codes!



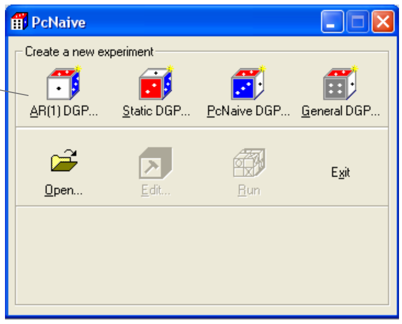
The model

- Time series, $y_1, y_2, \ldots, y_T$
- Generated by first order autoregressive, AR(1), model

$$y_t = \alpha \cdot y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1) \qquad (2.1)$$

- (2.1) is the Data Generating Process (DGP)

We need to use this one!



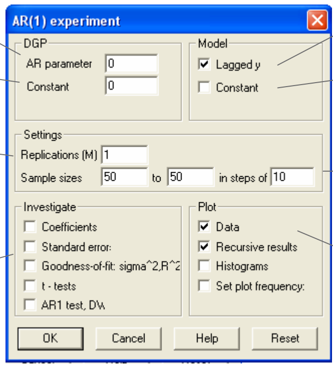
(1) Generating one realization of the DGP

This is the α parameter (=0)

We don't need a constant

Just one realization

We will use these later!



Yes!

No!

Only one sample size needed ("steps" ignored)

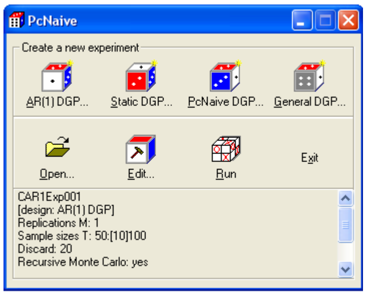
Select "data" so we can see what we have generated (the last three are only relevant for more than 1 replication!)

When you click OK you will be warned that we haven't asked for any evaluation statistics. That doesn't matter. You will also need to save the experiment – it doesn't really matter where.

Running the experiment

PcNaive now gives you the option of editing the experiment you have just created and running it.

Click "Run".



DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

Results displayed in GiveWin

```
---- PcNaive (2.00) DGP ----
y is (1 x 1), z is (0 x 1) and NOT fixed.

y[t] = e[t]

e ~ N(0,1)
20 observations discarded


---- System estimation by OLS ----
The available estimation sample is 2 - 51 (T=50)
Ya            lag 0 status Y
Ya            lag 1 status Y

---- PcNaive Monte Carlo results ----
T=51, M=1, seed=-1 (common)

Total time for experiment: 0.04
```

These are just the settings we chose. We have no additional regressors (z's). First 20 observations are discarded by default. Reduces effect of starting values on the lagged dependent variable.

There is no observation of y_{t-1} for $t=1$.

T is sample size, *M* is number of replications. The *seed* is used to generate the random process. -1 is the default. If we don't change it, we will get the same result every time.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

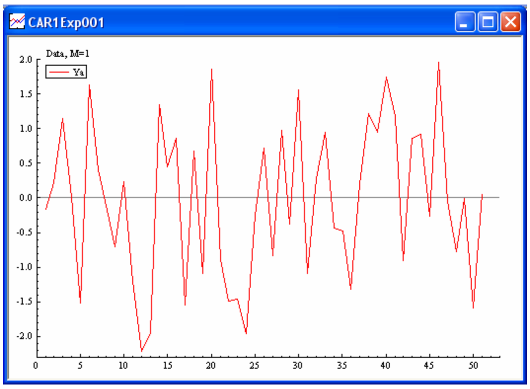
The Experimental Data

Here is the data we have generated for

$y_t = \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

If this series is to be stationary, then all observations of y_t must be drawn from the same distribution. Of course this is the case here!

Graphically this means that y_t must fluctuate around a constant level. This process looks very stationary! (White noise)



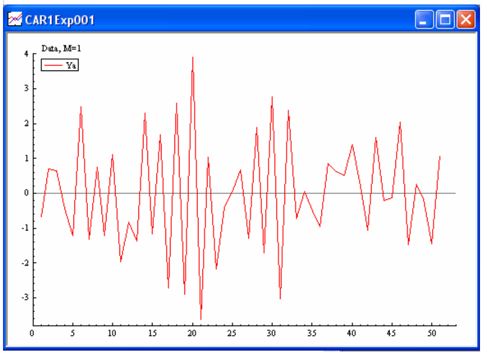
DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

(2) Alpha=-0.7 $y_t = -0.7 \cdot y_{t-1} + \varepsilon_t$

```
---- PcNaive (2.00) DGP ----
y is (1 x 1), z is (0 x 1) and NOT fixed.

y[t] = e[t] + A1 y[t-1]
A1 =
      -0.70000
```

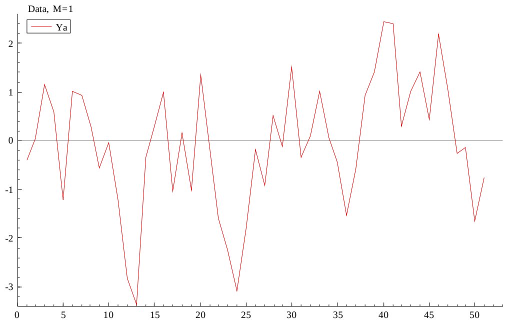
Note the change in the model! The parameter is now different from zero. We have not changed the seed, so the residuals will be the same as in the previous experiment.



We can now see that there is negative autocorrelation (a positive observation is likely to be followed by a negative observation) and *vice versa*.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

Alpha=0.5 $y_t = 0.5 \cdot y_{t-1} + \varepsilon_t$

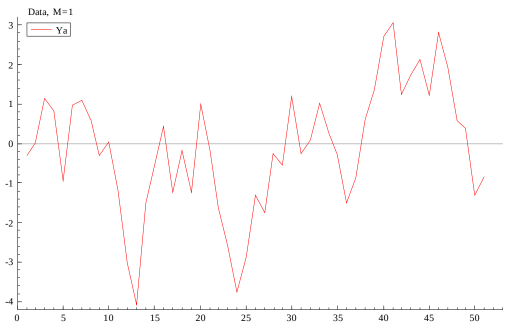


Positive autocorrelation means that high observations are likely to be followed by high observations, and low observations by low observations.

(e.g. Unemployment data)

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

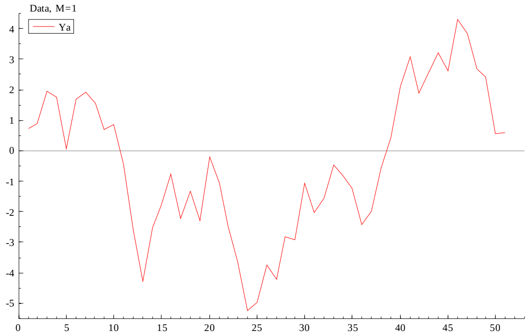
Alpha=0.7 $y_t = 0.7 \cdot y_{t-1} + \varepsilon_t$



As alpha increases, the process becomes less and less stationary, i.e. it crosses 0 less and less.

Shocks become more and more persistent.

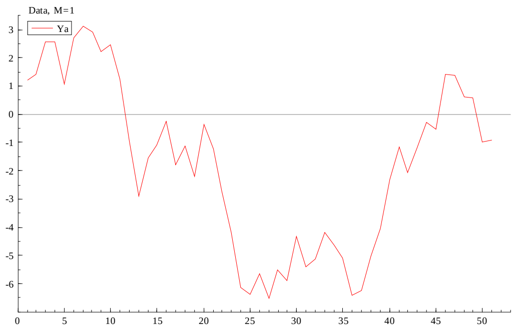
Alpha=0.9 $y_t = 0.9 \cdot y_{t-1} + \varepsilon_t$



As alpha approaches 1, the process is very nonstationary.

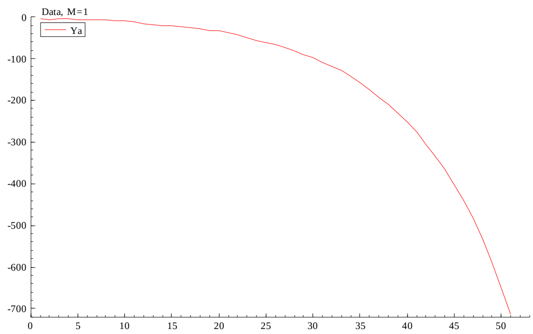
DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

Alpha = 1.0 $y_t = y_{t-1} + \varepsilon_t$



- This is a random walk.
- The best predictor of y_t is y_{t-1} .
- Nonstationary!
- This is the best model we have for fluctuations on the currency markets!

Alpha=1.1 $\quad y_t = 1.1 \cdot y_{t-1} + \varepsilon_t$



This is the extreme case, where the observations move further and further away from 0.

(3) Properties of the OLS estimator

See lecture note “Linear Regression with Time Series Data”.

Consistency: (See page 4)
OLS is consistent.

Unbiasedness: (See page 5)
OLS is only unbiased if there is *strict exogeneity*, i.e.

$$E[\epsilon_t \mid x_1, x_2, ..., x_t, ..., x_T] = 0,$$

Consider our model: $y_t = \alpha \cdot y_{t-1} + \epsilon_t,$

y_t is a function of ε_t, so ε_t cannot be uncorrelated with current and future values of the explanatory variables. The OLS estimator is therefore biased in general. (Except when the coefficient is equal to 0.)

(4) Define concepts

We perform a Monte Carlo experiment with *M* replications.
We define:

$$MEAN\left(\hat{\beta}_i\right) = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}_m$$

$$BIAS\left(\hat{\beta}_i\right) = MEAN\left(\hat{\beta}_i\right) - \beta_i$$

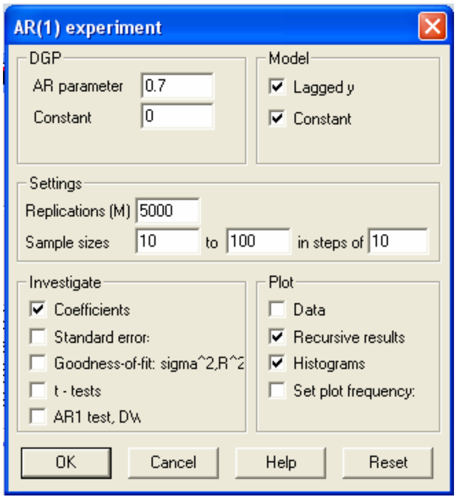
$$MCSD\left(\hat{\beta}_i\right) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\beta}_m - MEAN\left(\hat{\beta}_m\right)\right)^2}$$ Monte Carlo Standard Deviation

$$MCSE\left(\hat{\beta}_i\right) = M^{-\frac{1}{2}} \cdot MCSD\left(\hat{\beta}_i\right)$$ Monte Carlo Standard Error

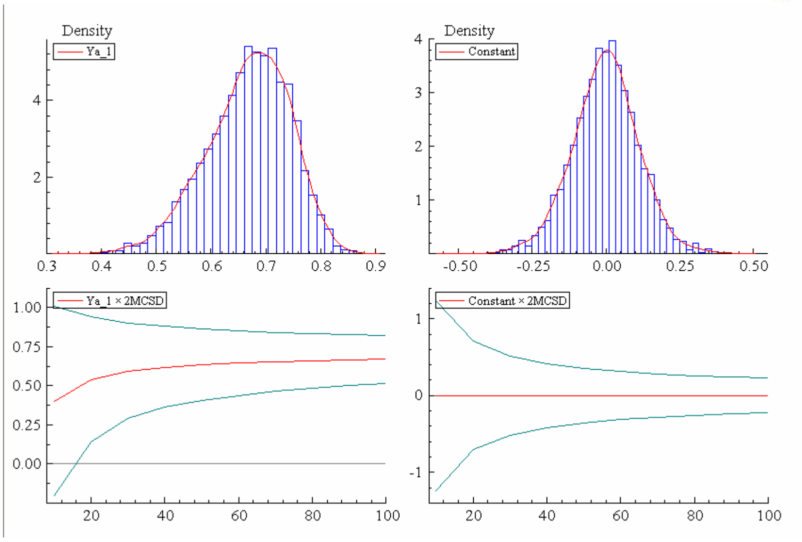
N.B. *MCSD* is a measure of the uncertainty of the OLS estimator (approaches 0 as *T* approaches infinity), *MCSE* is a measure of the uncertainty of the estimator *MEAN* in the simulation (approaches 0 as *M* approaches infinity).

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

(5) The Experiment!



Properties of $\hat{\beta}_1$ and $\hat{\beta}_0$



Properties of $\hat{\beta}_1$ and $\hat{\beta}_0$ (cont.)

```
---- PcNaive Monte Carlo results ----
T=101, M=5000, seed=-1 (common)

moments of estimates
                       mean          MCSD
Ya_1                0.66809      0.076101
Constant        -0.00048273       0.11291

biases of estimates
                  mean bias          MCSE          RMSE   true value
Ya_1              -0.031911     0.0010762      0.082521      0.70000
Constant        -0.00048273     0.0015968       0.11291      0.00000

Total time for experiment: 4.89
```

Are the estimators unbiased (i.e. is there a significant difference between the estimate and the true value)?

$\mathcal{H}_0$: MEAN$(\widehat{\beta}_0) = 0$ $\quad t = \frac{\widehat{MEAN} - 0}{MCSE} = \frac{\widehat{BIAS}}{MCSE} = \frac{-0.00048273}{0.0015968} = -0.302$ Clear accept (only because $\beta_0 = 0$)!

$\mathcal{H}_1$: MEAN$(\widehat{\beta}_1) = \alpha$. $\quad t = \frac{\widehat{MEAN} - 0.7}{MCSE} = \frac{\widehat{BIAS}}{MCSE} = \frac{-0.031911}{0.0010762} = -29.65$ Clear reject!

So there is bias, as we anticipated.

(6) Rejection frequency, size and power

Rejection frequency: Proportion of cases where a hypothesis is rejected.

$\mathcal{H}_2 \; : \; \beta_0 = 0$ TRUE. Rejection frequency $\rightarrow 5\%$

$\mathcal{H}_3 \; : \; \beta_1 = 0$ FALSE. Rejection frequency $\rightarrow 100\%$

as $T \rightarrow \infty$

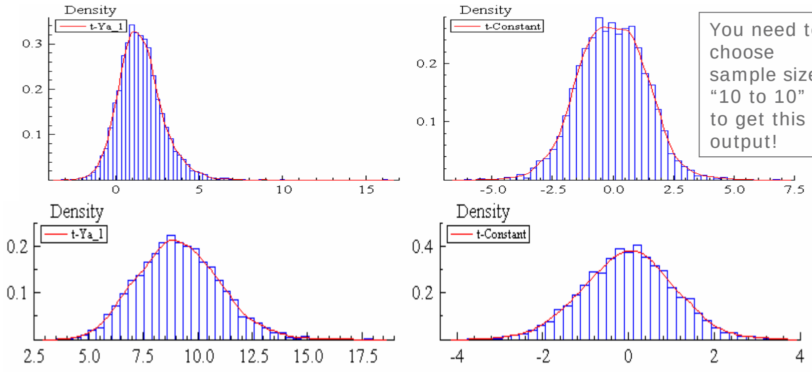
Rejecting H_2 is an example of a TYPE I error (rejecting a true null hypothesis). Probability is controlled by researcher through choice of significance level: the ***size*** of the test. As T increases, we will become less and less likely to reject the true hypothesis.

Accepting H_3 is an example of a TYPE II error (not rejecting the null when the alternative is true). Depends on the true parameter values. If hypothesis is very different from the true value, then the probability of rejection will be large. The reverse, the probability of rejecting the null when it is false, is the ***power*** of the test. As T increases, we will become more and more likely to reject the false hypothesis.

Select "t-tests" under AR(1)

T=10

T=100

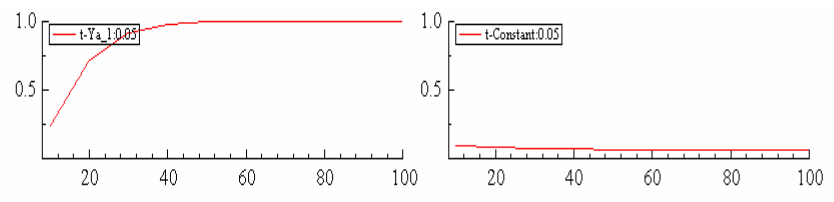


You need to choose sample size "10 to 10" to get this output!

This shows the distribution of the t-statistics for *T=10* and *T=100*. Note how the distribution for $\hat{\beta}_1$ shifts to the right for larger *T*. The bias (towards 0) is decreasing, so it is becoming more likely that we reject the false hypothesis.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

Rejection frequencies



For low *T* we are making a lot of type II errors! (i.e. accepting the null when the alternative is true)

DEPARTMENT OF ECONOMICS
UNIVERSITY OF COPENHAGEN

Computer basement (or home)

- Try to replicate these results yourself
- Try this problem set again with an intercept different from 0, i.e. $\beta_0 \neq 0$