

EC 771B Spring 2000 Problem Set 4

Christopher F. Baum

April 28, 2000

Due at classtime, Tuesday 2 May 2000

Set up a Stata program (do-file) to provide the empirical results requested. Hand in a copy of the program, annotated with your comments as warranted. The comments may be handwritten on the printout if they are clearly legible.

All commands referred to below are installed on fmrisec Stata. If you are using desktop Stata, you may have to install them (use `archinst` command; `webseek archinst` for help). These exercises cannot be performed with Stata version 5.

1. Use the Mills interest rate dataset, available from within Stata via the command

use `http://fmwww.bc.edu/ec-p/data/mills2d/rs-r20.dta`

This dataset contains monthly observations for 1951-1995 on the following variables:

- | | |
|-----------------------|--------------------------------------|
| 1. <code>month</code> | month identifier |
| 2. <code>rs</code> | UK short-term Treasury rate |
| 3. <code>r20</code> | UK long-term (20 year) Treasury rate |

In all estimations, use the sample 1957-1993 (use `"if tin(1957m1,1993m12)"` on each estimation command) so that you can perform out-of-sample forecasts for 1994 and 1995.

Generate the interest rate spread as the difference between long and short rates. Evaluate the competing models:

- Fit the spread as an AR(2) model; generate predictions for 1994-1995.
- Fit the spread as an AR(3) model; generate predictions for 1994-1995.
- Fit the spread as an ARMA(2,1) model (hint: see `arima`, and note that to fit an AR(2) you must specify the option as `ar(1 2)`), and generate predictions for 1994-1995.
- Fit the spread as an ARIMA(1,1,0) model (see `arima`), and generate predictions for 1994-1995 (hint: use the `'y'` option on `predict` to get predictions of the original level variable).
- Compare these four models in terms of their ability to predict the spread *ex ante*. Which model has the lowest mean *ex ante* prediction error?
- Apply the `dfgls` and `kpss` unit root tests to the spread series (assuming difference stationarity, that is, no trend, for both tests) and evaluate your findings.

2. Use the Mills S&P dataset, available from within Stata via the command

use `http://fmwww.bc.edu/ec-p/data/mills2d/sp500a.dta`

This dataset contains annual observations for 1871-1997 on the following variables:

- | | |
|----------------------|-----------------|
| 1. <code>year</code> | year identifier |
|----------------------|-----------------|

2. sp500a S&P 500 index, end of year
3. sp500ar S&P 500 index, annual percentage return

Generate the log of the S&P index (lsp) and use that series in your investigation.

a) Use the dfgls unit root test (allowing for trend stationarity) to determine whether the lsp series is I(1) or I(0).

b) Setting Tb=1929, generate $t = \text{year} - 1870$, $du = 1$ if $\text{year} > Tb$, and $dts = \text{year} - Tb$ if $\text{year} > Tb$, as suggested by Perron (Econometrica, 1989). Regress lsp on t, du and dts. Use the Box-Pierce Q statistic to evaluate the errors. Are they white noise?

c) Reestimate the equation allowing for an AR(2) error process (hint: use arima). How do you interpret this model? What was the growth rate of stock prices pre-1929? Post-1929? Was the change in growth rate significant? Save the residuals from this model. Are they white noise according to the Q statistic?

d) Given the form of this model, Perron has shown that the critical values for unit root tests on the residuals should be modified from those of the "Dickey-Fuller distribution." Run a standard augmented Dickey-Fuller test on the residuals, using 11 lags, and compare with the critical values from Perron Table VI.B (Model C) for $\lambda=0.5$ (breakpoint roughly halfway through the sample):

Sig. level	Critical value
10%	-3.95
5%	-4.22
1%	-4.81

What are your conclusions regarding the stochastic properties of the series?

3. Use the Mills daily FX dataset, available from within Stata via the command
use <http://fmwww.bc.edu/ec-p/data/mills2d/exchd.dta>

This dataset contains business-daily observations for 1974-1994 on the following variables:

1. day observation number
2. exchd exchange rate, \$/pound sterling

To reduce computation time, drop the first 4000 observations (drop if $_n \leq 4000$) and generate the difference of the exchange rate (dex). Use that series in your investigation.

a) Regress dex on its first and fifth lags (only), and test the errors of this model for ARCH of orders 5, 10, 20 (hint: see archlm).

b) Estimate this model with ARCH(5) errors (hint: see arch, and note that the order must be specified as a numlist, such as 1/5).

c) Reestimate the model with GARCH(1,1) errors.

d) Reestimate the model with GARCH(1,1)-in-mean errors. How do you interpret the ARCH-in-mean term?

Compare these three models' results and evaluate their ability to capture the conditional heteroskedasticity present in the data, as represented by the log-likelihood function values of the models.

4. Use the Mills interest rate dataset referenced above as well as the Mills FT dataset, available from within Stata via the command

use <http://fmwww.bc.edu/ec-p/data/mills2d/fta.dta>

This dataset contains monthly observations for 1965-1995 on the following variables:

- | | |
|-----------|--|
| 1. month | month identifier |
| 2. ftap | Financial Times (FT) stock price index |
| 3. ftadiv | Dividend yield on FT price index |
| 4. ftaret | Percentage returns on FT price index |
| 5. rpi | UK Retail price index |

You will have to merge the two datasets (see the merge command), and should create a time trend variable.

a) Transform the four variables ftap, ftadiv, rs, r20 into logarithms, and create differences. Fit a 4-variable VAR of these differenced series on six lags, using sureg (hint: place the list of regressors (using lag notation) in a global macro, and use the corr and notable options). How well does this model explain the comovements of the series?

b) Use Johansen and Juselius ML cointegration procedure (mlcoint) to estimate cointegrating relations in the four log-level variables. Use six lags, as you did for the VAR, and enter a time trend in the model using the static() option. For a four-variable system, the 0.05 critical values of Johansen's trace statistic are (from Hamilton, Time Series Analysis, Table B-10):

# of Random walks	# of CI vectors	Critical value
4	0	47.181
3	1	29.509
2	2	15.197
1	3	3.962

Determine the number of cointegrating vectors in the system via the trace test. Given your findings, is it appropriate to estimate the VAR in differences in (a)?