MFE MATLAB Companion Course
Michaelmas and Hilary 2013–2014

Tuesday 28th January, 2014

Notes

This is the Michaelmas term MATLAB companion course. The focus this term is on core skills and MATLAB usage. The Hilary term course will focus on modeling, leveraging the skills taught this term. *Note: This document may change during the term. The latest copy can always be found at*
`http://www.kevinsheppard.com/wiki/MFE_MATLAB`

Data and Simulation **3**
- Data Set Construction . . . 3
- Simulation . . . 3

Estimation and Inference **5**
- Method of Moments . . . 5
- Maximum Likelihood . . . 5
- Bias and Verification of Standard Errors . . . 6

Linear Regression **7**
- Basic Linear Regression . . . 7
- Rolling and Recursive Regressions . . . 7
- Model Selection and Cross-Validation . . . 8

Time-series Modeling **9**
- ARMA Estimation . . . 9
- ARMA Model Selection . . . 9
- ARMA Residual Disgnostics . . . 9
- ARMA Forecasting . . . 10
- Unit Root Testing . . . 10

Volatility Modeling **11**
- ARCH Model Estimation . . . 11
- ARCH Model Selection . . . 11
- ARCH Model Forecasting . . . 11

Value-at-Risk Modeling **12**
- Value-at-Risk using Historical Simulation . . . 12
- Value-at-Risk using Filtered Historical Simulation . . . 12
- Value-at-Risk Forecast Evaluation . . . 12

Vector Autoregressions **13**
- Vector Autoregresion (VAR) Estimation . . . 13
- VAR Model Order Selection . . . 13
- Granger Causality Testing . . . 13
- Impulse Responses . . . 14
- Cointegration Testing using Engle-Granger . . . 14

Data and Simulation

Data Set Construction

Note: In all cases, convert the dates to both MATLAB date format and YYYYMMDD.

Functions

```
csvread, xlsread, weekday, textscan, str2double, c2mdate, diff
```

1. Download all daily data for the S&P 500 and the FTSE 100 from Yahoo! Finance.
   (a) Import both data sets into MATLAB, converting dates to MATLAB date format.
   (b) Construct weekly price series from each, using Tuesday prices (less likely to be a holiday).
   (c) Construct monthly price series from each using last day in the month.
2. Write a function which will return month-end prices. The function signature should be

```
function [month_end_price, month_end_date] = month_end_prices(price,date)
```

3. Import the Fama-French benchmark portfolios as well as the 25 sorted portfolios at both the monthly and daily horizon.
   `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`
4. Import daily FX rate data for USD against AUD, Euro, JPY and GBP from the Federal Reserve Economic Database (FRED)
   `http://research.stlouisfed.org/fred2/categories/94`

Note: `c2mdate` is in the MFE toolbox.

Simulation

Functions

```
randn, trnd, rng
```

1. Simulate 100 standard Normal random variables
2. Simulate 100 random variables from a $N\left(.08, .2^2\right)$
3. Simulate 100 random variables from a Students t with 8 degrees of freedom

4. Simulate 100 random variables from a Students t with 8 degrees of freedom with a mean of 8% and a volatility of 20%. Note: $\mathrm{V}[X] = \frac{\nu}{\nu-2}$ when $X \sim t_\nu$.

5. Simulate two identical sets of 100 standard normal random variables by resetting the random number generator.

6. Repeat exercise 3 using *only* `randn`.

Expectations

Functions

`randn`, `trnd`, `chi2rnd`, `exp`, `mean`, `std`, `integral`, `quadl`

1. Compute $\mathrm{E}[X]$, $\mathrm{E}\left[X^2\right]$, $\mathrm{V}[X]$ and the kurtosis of X using Monte Carlo integration when X is distributed:

   (a) Standard Normal

   (b) $N\left(0.08, 0.2^2\right)$

   (c) Students t_8

   (d) χ^2_5

2. Compute $\mathrm{E}[\exp(X)]$ when $X \sim N\left(0.08, 0.2^2\right)$. Compare this to the analytical result for a Log-Normal random variable.

3. Explore the role of uncertainty in Monte Carlo integration by using $4\times$ as many simulations as in the base case.

4. Compute the expectation in exercise 2 using quadrature. Note: This requires writing a function which will return $\exp(x) \times \phi(x)$ where $\phi(x)$ is the probability form the pdf.

5. [Much more challenging] Suppose log stock market returns are distributed according to a Students t with 8 degrees of freedom, mean 8% and volatility 20%. Utility maximizers hold a portfolio consisting of a risk-free asset paying 1% and the stock market. Assume that they are myopic and only care about next period wealth, so that

$$U(W_{t+1}) = U\left(\exp\left(r_p\right) W_t\right)$$

   and that $U(W) = \frac{W^{1-\gamma}}{1-\gamma}$ is CRRA with risk aversion γ. The portfolio return is $r_p = w r_s + (1-w) r_f$ where s is for stock market and f is for risk-free. A 4th order expansion of this utility is

$$\mathrm{E}_t\left[U(W_{t+1})\right] \approx \phi_0 + \phi_1 \mu'_1 + \phi_2 \mu'_2 + \phi_3 \mu'_3 + \phi_4 \mu'_4$$

   where

$$\phi_j = (j!)^{-1} U^{(j)}(W_t), \; U^{(j)} = \frac{\partial^j U}{\partial W^j} \text{ and } \mu'_k = \mathrm{E}_t\left[W^k_{t+1}\right].$$

   Use Monte Carlo integration to examine how the weight in the stock market varies as the risk aversion varies from 0.1 to 1. Note that when $\gamma = 1$, $U(W) = \ln(W)$. Use $W_t = 1$ without loss of generality since the portfolio problem is homogeneous of degree 0 in wealth.

Estimation and Inference

Method of Moments

Functions

```
mean, sum, subplot, plot
```

1. Estimate the mean, variance, skewness and kurtosis of the S&P 500 and FTSE 100 using the method of moments using monthly data.

2. Estimate the asymptotic covariance of the mean and variance of the two series (separately, but not the skewness and kurtosis).

3. Estimate the Sharpe ratio of the two series and compute the asymptotic variance of the Sharpe ratio. See Chapter 2 of the notes for more on this problem.

4. Plot rolling estimates of these using 120 months of consecutive data using a 4 by 1 subplot against the dates.

Maximum Likelihood

Functions

```
log, gamma, gammaln, normcdf, erf, fminunc, fmincon, trnd, var, std, normpdf
```

1. Simulate a set of i.i.d. Student's t random variables with degree of freedom parameter $\nu = 10$. Standardize the residuals so that they have unit variance using the fact that $\mathrm{V}[x] = \frac{\nu}{\nu-2}$. Use these to estimate the degree of freedom using maximum likelihood. Note that the likelihood of a standardized Student's t is

$$f(x;\nu,\mu,\sigma_t^2) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \frac{1}{\sigma_t} \frac{1}{\left(1+\frac{(x-\mu)^2}{\sigma_t^2(\nu-2)}\right)^{\frac{\nu+1}{2}}}$$

   where $\Gamma()$ is known as the gamma function.

2. Repeat the previous exercise using daily, weekly and monthly S&P 500 and FTSE 100 data. Note that it is necessary to remove the mean and standardize by the standard deviation error before estimating the degree of freedom. What happens over longer horizons?

3. Repeat the previous problem by estimating the mean and variance simultaneously with the degree of freedom parameter.

4. Simulate a set of Bernoulli random variables y_i where $p_i = \Phi(x_i)$ where $X_i \sim N(0,1)$. (Note: p_i is the probability of success and $\Phi()$ is the standard Normal CDF). Use this simulated data to estimate the Probit model where

$$p_i = \Phi(\alpha_0 + \alpha_1 x_i)$$

   using maximum likelihood.

5. Estimate the asymptotic covariance of the estimated parameters in question 3.

Bias and Verification of Standard Errors

Functions

```
sum, mean, cov, bsxfun, fminunc, normcdf, normpdf, norminv, pltdens, chi2rnd, rand, plot, title, axis,
    legend
```

1. Simulate a set of i.i.d. χ^2_5 random variables and use the method of moments to estimate the mean and variance.
2. Compute the asymptotic variance of the method of moment estimator.
3. Repeat 1 and 2 a total of 1000 times. Examine the finite sample bias of these estimators relative to the true values.
4. Repeat 1 and 2 a total of 1000 times. Compare the covariance of the estimated means and variance (1000 of each) to the asymptotic covariance of the parameters (use the average of the 1000 estimated variance-covariances). Are these close? How does the sample size affect this?
5. In the previous problem, for each parameter, form a standardized parameter estimate as

$$z_i = \frac{\sqrt{n}\left(\hat{\theta}_i - \theta_{i,0}\right)}{\sqrt{\hat{\Sigma}_{ii}}}$$

   where $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, \Sigma)$ so that $\hat{\Sigma}$ is the estimated asymptotic covariance. What percent of these z_i are larger in absolute value than 10%, 5% and 1% 2-sided critical values from a normal?
6. Produce a density plot of the z_i standardized parameters and compare to a standard normal.
7. Repeat the same exercise for the Bernoulli problem from the previous question.

Linear Regression

Basic Linear Regression

Functions

```
mean, bsxfun, repmat
```

1. Write a function to estimate the parameters of a linear regression and compute related quantities. The function signature should match
   `[b,tstat,s2,VCV,VCV_white,R2,Rbar,yhat]=ols(y,x,c)` where the inputs are

   - `y`: A n by 1 vector of dependent variables
   - `x`: A n by K vector of independent variables
   - `c`: A scalar, 1 if the model should include a constant. If not, then no constant will be added to the
   - `x` matrix.
     and where the outputs are
   - `b`: The estimated coefficients
   - `tstat`: The t-stats of the estimated coefficients
   - `s2`: The estimated variance of residual
   - `VCV`: The K by K covariance matrix of the estimated parameters
   - `VCVwhite`: The K by K heteroskedasticity robust covariance matrix of the estimated parameters
   - `R2`: The R^2 of the regression
   - `Rbar`: The $\bar{R}^2$ of the regression
   - `yhat`: The fit values $\hat{y}$.

2. Use the OLS function to estimate the coefficients of the Fama-French portfolios (monthly data) on the market, size and value factors. Use only the four extremum portfolios – that is the 1-1, 1-5, 5-1 and 5-5 portfolios.

3. Are the parameter standard errors similar using the two covariance estimators? If not, what does this mean?

4. How much of the variation is explained by these three regressors?

Rolling and Recursive Regressions

Functions

```
ols, title, datetick, legend, axis, subplot, plot, figure
```

1. For the same portfolios in the previous exercise, compute rolling βs using 60 consecutive observations.
2. For each of the four *market* βs, produce a plot containing four series:
   - A line corresponding to the constant β (full sample)
   - The β estimated on the rolling sample
   - The constant β plus 1.96 $\times$ the variance of a 60-observation estimate of β.[1]
   - The constant β minus 1.96 $\times$ the variance of a 60-observation estimate of β.
3. Do the market exposures appear constant?
4. What happens if only the market is used as a factor (repeat the exercise excluding SMB and HML).
5. In problems 1 and 2, is there any evidence of time-variation in the SMB of HML loadings?

Model Selection and Cross-Validation

Functions

```
randperm, ols, setdiff, norminv, linspace, mean
```

1. For these portfolios, and considering all 8 sets of regressors which range from no factor to all 3 factors, which model is preferred by AIC, BIC, GtS and StG?
2. Cross-validation is a method of analyzing the in-sample forecasting ability of a cross-sectional model by using α% of the data to estimate the model and then measuring the fit using the remaining $100-\alpha$%. The most common forms are 5- and 10-fold cross-validation which use $\alpha = 20\%$ and 10%, respectively. $k-$fold cross validation is implemented by randomly grouping the data into k-equally-sized groups, using $k-1$ of the groups to estimate parameters, and then evaluating using the bin that was held out. This is then repeated so that each bin is held out.
   (a) Implement cross-validation using the 5- and 10-fold methods for all 8 models.
   (b) For each model, compute the full-sample sum of squared errors as well as the sum-of-squared errors using the held-out sample. Note that all data points will appear exactly once in both of these sum or squared errors. What happens to the cross-validatedR^2 when computed on the held out sample when compared to the full-sample R^2? (k-fold cross validated SSE by the TSS).

[1]The 60-month covariance can be estimated using a full sample VCV and rescaling it by $T/60$ where T is the length of the full sample used to estimate the VCV. Alternatively, the VCV could be estimated by first estimating the 60-month VCV for each sub-sample and then averaging these.

Time-series Modeling

ARMA Estimation

Functions

`ARMAX`, `armaxfilter`

1. Download data on 1 year and 10 year US government bond rates from FRED, and construct the term premium as the different in yields on 10 year and 1 year bonds.
2. Estimate an AR(1) on the term premium, and compute standard errors for the parameters.
3. Estimate an MA(5) on the term premium, and compute standard errors for the parameters.
4. Estimate an ARMA(1,1) on the term premium, and compute standard errors for the parameters.

ARMA Model Selection

Functions

`armaxfilter`

1. Perform a model selection exercise on the term premium using
   (a) General-to-Specific
   (b) Specific-to-General
   (c) Minimizing an Information Criteria

Note: When comparing models with different AR orders, it is important to use the `holdback` input.

ARMA Residual Disgnostics

Functions

`aramxfilter`, `plot`, `lmtest1`, `ljungbox`

1. Compute the residuals from your preferred model from the previous exercise, as well as a random-walk model.
   (a) Plot the residuals
   (b) Is there evidence of autocorrelation in the residuals?
   (c) Compute the Q statistic from both sets of residuals. Is there evidence of serial correlation?
   (d) Compute the LM test for serial correlation. Is there evidence of serial correlation?

ARMA Forecasting

Functions

`arma_forecaster`, `cov_nw`, `ols`

1. Produce 1-step forecasts from your preferred model in the previous exercise, as well as a random-walk model.

   (a) Are the forecasts objectively accurate?

   (b) Compare these forecasts to the random walk models using MSE and MAE.

2. Produce 3-step forecasts from the models selected in the previous exercises as well as a random walk model.

   (a) Compare these forecasts to the random walk models using MSE and MAE.

Note: Use 50% of the sample to estimate the model and 50% to evaluate it.

Unit Root Testing

Functions

`augdf`, `augdfautolag`

1. Download data on the AAA and BAA yeilds (Moodys) from FRED and construct the default premium.

   (a) Test the default premium for a unit root.

   (b) If you find a unit root, test the change.

2. Download data on consumer prices in the UK from the ONS.

   (a) Test the log of CPI for a unit root.

   (b) If you find a unit root, test inflation for one.

Volatility Modeling

ARCH Model Estimation

Functions

`tarch`

1. Download 10 years of data for the S&P 500 and the EUR/USD exchange rate from FRED.
2. Estimate a GARCH(1,1) and a GJR-GARCH(1,1,1) to the returns of both series.
3. Comment on the asymmetry.
   (a) Compare robust and non-robust standard errors.
   (b) Plot the fit variance and fit volatility.
   (c) Plot the standardized residuals.

ARCH Model Selection

Functions

`tarch`, `egarch`

1. Which model is selected for the S&P 500 among the classes:
   (a) TARCH
   (b) GJR-GARCH
   (c) EGARCH

ARCH Model Forecasting

Functions

`tarch`, `egarch`

1. Use 50% of the sample to estimate your preferred GARCH model for returns to the S&P 500 and the EUR/USB rate, and construct forecasts for the remaining period.
2. Evaluate the accuracy of the forecasts.
3. Evaluate the accuracy of forecasts from a 2-year backward moving average variance.
4. Compare the ARCH-model forecasts to a naive 2-year backward looking moving average using QLIKE.

Value-at-Risk Modeling

Value-at-Risk using Historical Simulation

Functions

```
rand
```

1. Compute the historical simulation VaR for the S&P 500 and the EUR/USD rate.

Note: Start the historical simulation at 25% of the data, and then build the additional forecasts using a recursive scheme.

Value-at-Risk using Filtered Historical Simulation

Functions

```
tarch
```

1. Use a GARCH(1,1) model to construct filtered historical VaR for the S&P 500 and the EUR/USD exchange rate, using 10 years of data.
2. Compare this VaR to the HS VaR in the previous example.

Note: For simplicity, estimate the model on the full sample, but start the historical simulation at 25% of the data, and then build the additional forecasts using a recursive scheme.

Value-at-Risk Forecast Evaluation

Functions

```
ols
```

1. Evaluate the FHS and HS VaR forecasts constructed in the previous exercises using:
   (a) HIT tests
   (b) The Bernoulli test for unconditionally correct VaR
   (c) Christoffersen's test for conditionally correct VaR

Vector Autoregressions

Vector Autoregresion (VAR) Estimation

Functions

```
vectorar, eig
```

1. Download data on 10-year interest rates, 1-year interest rates and the GDP deflator from FRED.
2. Transform the GDP deflator to be percent returns (e.g. $\Delta \ln gdpd_t$).
3. Estimate a first-order VAR on the spread between the 10-year and 1-year ($spread$), the one-year, and the growth rate of the GDP deflator.
4. What are the "own" effects?
5. What are the cross effects between these?
6. How could you get a sense of the persistence of this system?

VAR Model Order Selection

Functions

```
vectorar
```

1. Using the same data as in the previous exercise, determine the optimal VAR order using:
   (a) AIC
   (b) HQIC
   (c) BIC
   (d) Likelihood-ratio testing using General-to-specific

Granger Causality Testing

Functions

```
grangercause
```

1. Using the data in the model selected in the previous exercise, is there evidence of Granger Causality between the series?
2. What if the 10 year and the 1 year are both used, but the spread is omitted?

Impulse Responses

Functions

`impulseresponse`

1. Plot the impulse responses from both a first order model and the model selected in the order selection exercise.
2. Which covariance factor makes sense for the impulse responses?
3. What happens when you re-order the series and use the Choleski factor in the impulse response?
   (a) Which series should be first?
   (b) Which should be last?

Cointegration Testing using Engle-Granger

Functions

`augdf`, `augdfautolag`

1. Download data on CAY from Sydney Ludvigson's site.
2. Is there evidence that these three series are cointegrated in the entire sample?
3. What about in the post-Volker era (start in 1981)?