Tutorial 8

Factor Analysis and Principal Component Analysis The following exercises consider Factor Analysis and Principal Component Analysis applied to Time Series.

1. Let $\underline{R}_t$ denote the vector of returns of assets in the time period t. A *factor model* is:

$$\underline{R}_t = \underline{\alpha} + B\underline{F}_t + \underline{\epsilon}_t \qquad \underline{\epsilon}_t \sim \mathrm{WN}(\underline{0}, \Sigma)$$

   where B is a matrix where entry B_{ij} denotes the *factor loading* of factor j for asset i. The factors $\underline{F}_t$ are assumed to be a stationary process with mean $\mathbb{E}[\underline{F}_t] = \underline{\mu}$ and covariance structure Σ_F. Some PCA and Factor Analysis should be familiar from Statistics II.

   Consider the monthly simple and excess returns, in percentages and including dividends, of 13 stocks and the S&P 500 composite index from January 1990 to December 2008. The monthly Treasury bill rate in the secondary market is used as the risk-free interest rate to compute the excess returns. The tick symbols for the stocks are: AA, AXP, CAT, DE, F, FDX, HPQ, IBM, JNJ, KMB, MMM, PG and WFC. The data is found in the file `m-fac-ex-9008.txt`. Perform a market model analysis, assuming a *single* factor, and estimate the coefficients, β_i, variance σ_i^2 for each stock and R^2 (proportion of sum of squares due to model).

   Hint The following may be useful.

```
> m.fac.ex.9008 <- read.table("~/data/m-fac-ex-9008.txt", header=T, quote="\"")
>   View(m.fac.ex.9008)
> data<-m.fac.ex.9008
> x <- cbind(rep(1,228),data[,14])
> y<-data[,1:13]
> m <- x%*%(solve(t(x)%*%x))%*%t(x)
> y.hat <- m*y
> e.hat <- y - y.hat
> mat <- (solve(t(x)%*%x))%*%t(x)
> y<-as.matrix(y)
> bet <- mat%*%y
> beta.hat <- bet[2,]
> D.hat=diag(crossprod(e.hat)/(228-2))
```

2. Consider the monthly log stock returns, in percentages and including dividends, of Merck & Company, Johnson & Johnson, General Electric, General Motors, Ford Motor Company, and value-weighted index from January 1960 to 2008, found in the file `m-mrk2vw.txt`.

   (a) Perform a principal component analysis on the data using the sample covariance matrix.

   (b) Perform a principal component analysis on the data using the sample correlation matrix.

(c) Perform a statistical factor analysis on the data. Identify the number of common factors. Obtain estimates of the factor loadings using both principal component and maximum likelihood methods.

3. (a) The file `m-excess-c10sp-9003.txt` contains the monthly simple excess returns of 10 stocks and the S&P 500 index. The 3-month Treasury bill rate on the secondary market is used to compute the excess returns. The sample period is from January 1990 to December 2003 for the 168 observations. The 11 columns in the file contain the returns for ABT, LLY, MRK, PFE, F, GM, BP, CVX, RD, XOM and SP5 respectively. Analyse the 10 stock excess returns using a single factor model. Plot the beta estimate and R^2 for each stock and use the *global minimum variance portfolio* to compare the covariance matrices of the fitted model and the data.

   (b) Perform a Principal Component Analysis on the data and obtain the scree plot. How many common factors are there? Interpret the common factors.

   (c) Perform a statistical factor analysis. How many common factors are there if a 5% significance level is used? Plot the estimated factor loadings of the fitted model. Are the common factors meaningful?

Bayesian Methods for Multivariate Time Series We investigate the **mbsts** package for Multivariate Bayesian Structural Time Series Models in R. Here Y_t is an m-variate time series which has decomposition

$$Y_t = \mu_t + \tau_t + \omega_t + \xi_t + \epsilon_t$$

$\{\epsilon_t\} \sim IIDN(0, \Sigma)$, $\xi_t = (\xi_t^1, \ldots, \xi_t^m)'$ where

$$\xi_t^i = \beta_i' x_t^{(i)}$$

for regressor variables $x_t^i = (x_{t1}^i, \ldots, x_{tk_i})'$, the pool of al available predictors at time t for target series Y^i. The coefficients β are considered to be static (i.e. they do not change). τ is a *seasonal* component and ω a *cyclical* component. A *cyclical* component is a component when data exhibits ups and downs that are not of fixed periods (e.g. economic cycles).

Follow through the script, where we simulate a bi-variate time series based on this model and then try to recover the various components. We simulate a bi-variate series of length 505; we have 8 predictors.

You'll find more details in the paper

`https://arxiv.org/pdf/2106.14045.pdf`

You'll also find that the crucial Monte Carlo Markov Chain step takes a l-o-n-g time. I have found that this is generally the case; the Bayesian approach provides a framework which is theoretically sounder and more satisfying, but the down-side is that the Metropolis Hastings or Gibbs McMC samplers are computationally much more expensive than their classical counterparts.