Multivariate Time Series Analysis: Brief Review and Recent Developments

Ruey S. Tsay
Booth School of Business
University of Chicago

December 2015

Objective

Analysis of multivariate time-series data: finite dimensional case

- Challenges and available models
- Obtaining parsimonious and identifiable models for estimation
- Dimension reduction: Extracting "useful" information when the dimension is high
- Handling count data
- Modeling multivariate volatility

Purposes: (a) Finding relationships (linear) among variables, (b) predictions, (c) asset allocations, etc.

Outline

- Review: multivariate time series analysis
  1. Linear models: VAR, VMA, VARMA, Seasonal VARMA, VARX, Multivariate linear regression with time-series errors, transfer function models, etc.
  2. Unit-root nonstationarity and co-integration
- Dimension reduction: Factor models, PCA, and beyond
  1. Many factor models available
  2. Exact lagged linear relationship: PCA
  3. A motivating example
- Analysis of count data
  1. Poisson conditional autoregressive models
  2. Simple illustration and applications
- Multivariate volatility
  1. What is a volatility matrix?
  2. Why is it important?

Basic concepts

Let $\boldsymbol{Z}_t = (z_{1t}, \ldots, z_{kt})'$ be a k-dimensional time series observed at equally spaced time intervals.

- Strong stationarity: distributions are time invariant
- Weak stationarity: first 2 moments are finite & time-invariant.
- Linearity:

$$\boldsymbol{Z}_t = \boldsymbol{C} + \sum_{i=0}^{\infty} \psi_i \boldsymbol{a}_{t-i},$$

  $\boldsymbol{C}$ is a constant vector, $\psi_0 = \boldsymbol{I}$, ψ_i are $k \times k$ real matrices, $\{\boldsymbol{a}_t\}$ are k-dimensional iid noises with mean zero & cov $=$ $\boldsymbol{\Sigma}_a > 0$.
- Invertibility:

$$\boldsymbol{Z}_t = \boldsymbol{C} + \sum_{i=1}^{\infty} \boldsymbol{\pi}_i z_{t-i} + \boldsymbol{a}_t,$$

  where $\boldsymbol{\pi}_i$ are $k \times k$ real matrices.

Parameterization

Consequences of parameterization:

- Why $\boldsymbol{\Sigma}_a > 0$?
- Use Cholesky decomposition: $\boldsymbol{\Sigma}_a = \boldsymbol{L}\boldsymbol{\Omega}\boldsymbol{L}'$, where $\boldsymbol{\Omega}$ is a diagonal matrix, $\boldsymbol{L}$ is lower triangular with 1 on the diagonal. Define $\boldsymbol{b}_t = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. The series can be rewritten as

$$\boldsymbol{Z}_t = \boldsymbol{C} + \sum_{i=0}^{\infty} \psi_i^* \boldsymbol{b}_{t-i},$$

  where $\psi_i^* = \psi_i \boldsymbol{L}$ and $\text{cov}(\boldsymbol{b}_t) = \boldsymbol{\Omega}$ is diagonal, and $\psi_0^* = \boldsymbol{L}$, a lower triangular matrix with unit diagonal elements.
- The model can also be written as

$$\boldsymbol{L}^{-1}\boldsymbol{Z}_t = \boldsymbol{C}^* + \boldsymbol{b}_t + \sum_{i=1}^{\infty} \omega_i \boldsymbol{b}_{t-i},$$

  where $\boldsymbol{L}^{-1}$ is lower triangular and $\omega_i = \boldsymbol{L}^{-1}\psi_i^*$.

Available models

Two difficulties encountered in MTS modeling:

1. Too many parameters
2. Model identification: identifiability

Vector autoregressive moving-average (VARMA) model:

$$\phi(B)(\boldsymbol{Z}_t - \boldsymbol{\mu}) = \boldsymbol{\theta}(B)\boldsymbol{a}_t,$$

$\boldsymbol{\mu}$ a constant vector, $\boldsymbol{a}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_a)$, and

$$\phi(B) = \boldsymbol{I} - \sum_{i=1}^{p} \phi_i B^i, \quad \boldsymbol{\theta}(B) = \boldsymbol{I} - \sum_{i=1}^{q} \boldsymbol{\theta}_i B^i,$$

with B denoting the back-shift (or lag) operator

Identifiability

Two simple examples with $k = 2$: First,

$$\boldsymbol{Z}_t = \boldsymbol{a}_t - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \boldsymbol{a}_{t-1} \Longleftrightarrow \boldsymbol{Z}_t - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \boldsymbol{Z}_{t-1} = \boldsymbol{a}_t$$

That is, VMA(1) $=$ VAR(1). Next,

$$\boldsymbol{Z}_t - \begin{bmatrix} 0.8 & 2 \\ 0 & 0 \end{bmatrix} \boldsymbol{Z}_{t-1} = \boldsymbol{a}_t - \begin{bmatrix} 0.3 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{a}_{t-1}$$

is the same as

$$\boldsymbol{Z}_t - \begin{bmatrix} 0.8 & 2+\omega \\ 0 & \beta \end{bmatrix} \boldsymbol{Z}_{t-1} = \boldsymbol{a}_t - \begin{bmatrix} 0.3 & \omega \\ 0 & \beta \end{bmatrix} \boldsymbol{a}_{t-1},$$

where ω and β are arbitrary.

Model continued

Assumptions for block-identifiability:

1. $\phi(B)$ and $\boldsymbol{\theta}(B)$ are left co-prime: Any common left factor must be unimodular
2. Rank of $[\phi_p, \boldsymbol{\theta}_q]$ is k.

How to overcome the identifiability problem?

1. Use VAR models
2. Structural specification: finding the hidden model structure

Two issues with using VAR models only:

1. Over-parametrization
2. Difficulty with non-invertible models, i.e. over-differencing.

Multivariate time series analysis in practice

Reference: Tsay (2014), **Multivariate Time Series Analysis with R and Financial Applications**, Wiley.

- An associated R package, called `MTS`
- Can handle the models discussed
- A illustrative example
  1. The quarterly GDP of United Kingdom, Canada, and United States
  2. Sample period: 1980.I to 2011.II
  3. Data downloaded from FRED (Federal Reserve Economic Data). They are also available in MTS by using the command `data(‘‘mts-examples’’,package=‘‘MTS’’)`

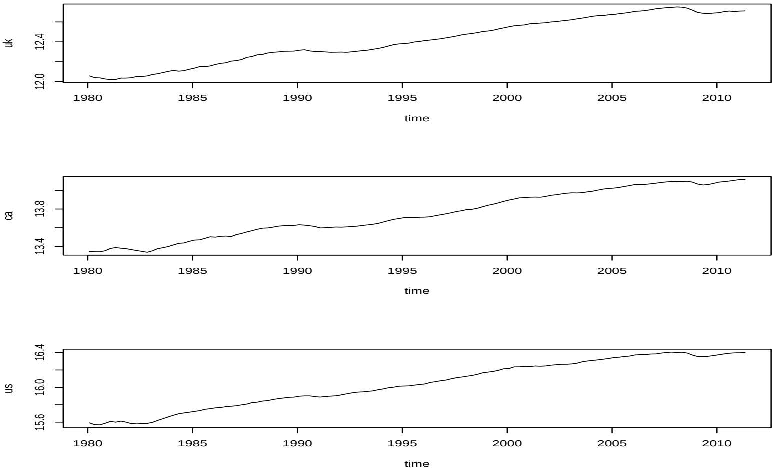

Figure: Time plots of the logarithms of quarterly real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011.

Analysis GDP growth rates

A general **Procedure**

1. Preliminary analysis: plot, ccm
2. Order selection: various criteria available
3. Estimation: Gaussian maximum likelihood or OLS
4. Refinement, including Granger causality
5. Model checking
6. Prediction
7. Impulse response functions

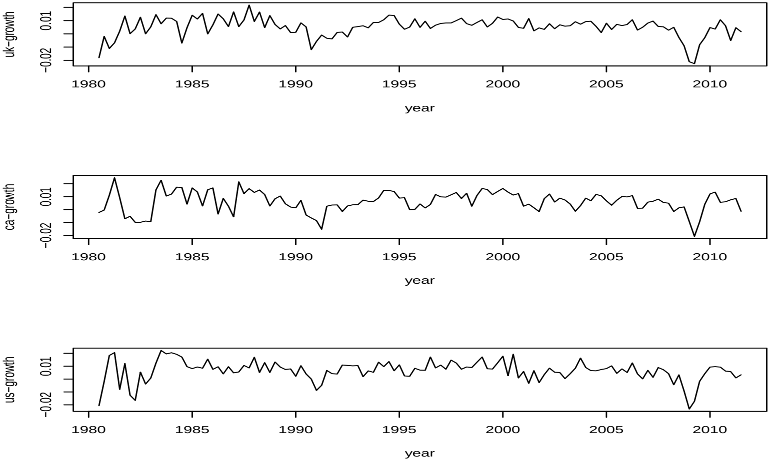

Figure: Time plots of the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011.

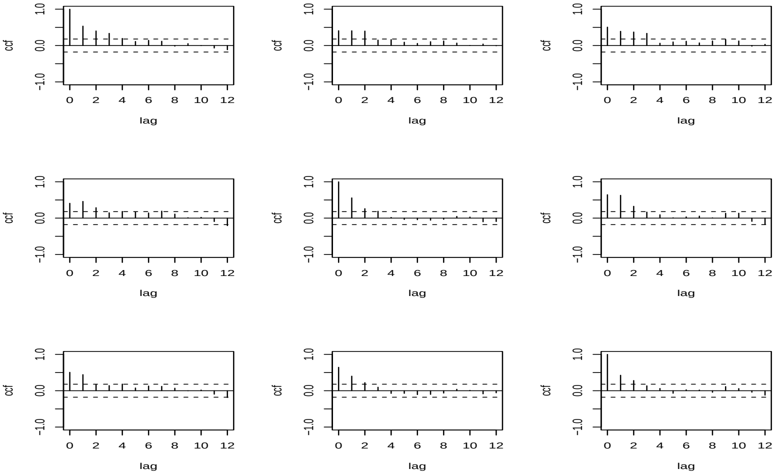

Figure: Cross-correlation matrices of the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011.

Another Example

What is the impact of Chinese economy on world markets?

1. Quarterly China GDP growth rates
2. Quarterly growth rates of crude oil prices: Western Texas Intermediate
3. Data span: 1994.II to 2015.II

A simple VAR(5) model:

$$
\begin{aligned}
\boldsymbol{Z}_t &= \begin{bmatrix} 1.35 \\ 0 \end{bmatrix} + \begin{bmatrix} .707 & .0179 \\ 4.788 & .2818 \end{bmatrix} \boldsymbol{Z}_{t-1} - \begin{bmatrix} 0 & 0 \\ 5.74 & .412 \end{bmatrix} \boldsymbol{Z}_{t-2} \\
&+ \begin{bmatrix} 0 & 0 \\ 4.14 & 0 \end{bmatrix} \boldsymbol{Z}_{t-3} + \begin{bmatrix} .296 & 0 \\ 0 & -.193 \end{bmatrix} \boldsymbol{Z}_{t-4} \\
&- \begin{bmatrix} .157 & .023 \\ 2.82 & .166 \end{bmatrix} \boldsymbol{Z}_{t-5} + \boldsymbol{a}_t, \quad \boldsymbol{\Sigma} = \begin{bmatrix} .627 & .259 \\ .259 & 167.61 \end{bmatrix}.
\end{aligned}
$$

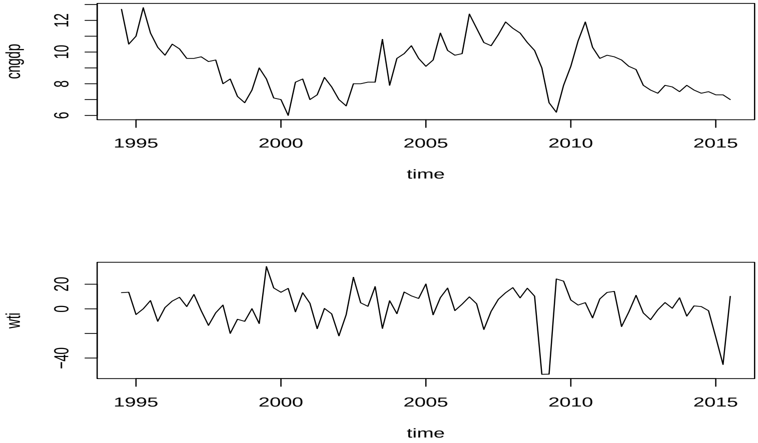

Figure: Time plots of the quarterly growth rates of Chinese GDP and the crude oil prices (WTI) from 1994.II to 2015.II

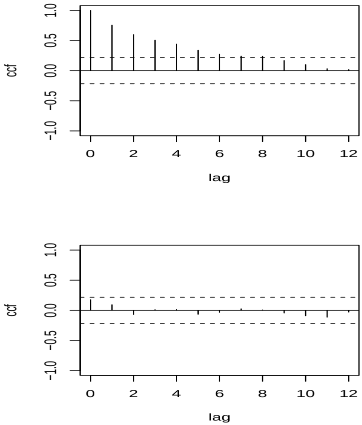

Figure: Cross-correlations between the quarterly growth rates of China GDP and WTI.

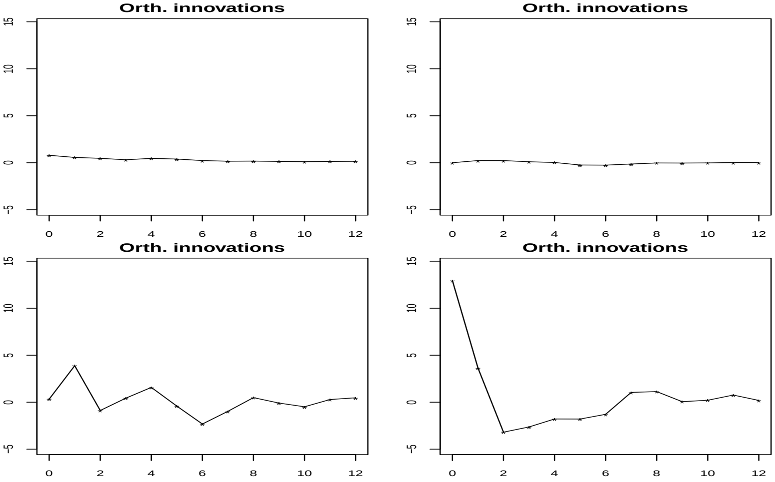

Figure: Impulse response functions of a fitted VAR(5) model for the growth rates of China GDP and WTI.

Model specification

- VAR models
  1. Information criteria: AIC, BIC, and HQ
  2. Tiao-Box sequential chi-square statistics: M-stat
- VARMA models
  1. ECCM: extended cross-correlation matrices
  2. SCM: scalar component models
  3. Kronecker index

  The last two are for structural specification, resulting in identifiable models

Brief descriptions

Consider a k-dimensional series Z_t.
ECCM method:

1. Requirement: The model is block identifiable: (a) $\phi(B)$ and $\boldsymbol{\theta}(B)$ are left co-prime, i.e. $\phi(B) = \boldsymbol{U}(B)\phi^*(B)$ and $\boldsymbol{\theta}(B) = \boldsymbol{U}(B)\boldsymbol{\theta}^*(B)$, then $|\boldsymbol{U}(B)|$ is a constant. (b) $\text{Rank}[\phi_p, \boldsymbol{\theta}_q] = k$.
2. Idea: (a) Seek consistent LS estimates of $\phi(B)$. (b) Transform Z_t by $\boldsymbol{W}_t = \hat{\phi}(B)Z_t$. (c) Use cross-correlation matrices of $\boldsymbol{W}_t$ to specify q.
3. Procedure: (a) is achieved by iterated vector autoregressions.

Illustration

Consider $k = 1$ and Z_t follows $Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1}$. The auto-regression

$$Z_t = \phi^{(0)} Z_{t-1} + \epsilon_t^{(0)}$$

gives $\phi^{(0)} = \rho_1$, which is not ϕ_1. However, the iterated auto-regression

$$Z_t = \phi^{(1)} Z_{t-1} + \beta \epsilon_{t-1}^{(0)} + \epsilon_t^{(1)}$$

gives $\phi^{(1)} = \rho_2 / \rho_1 = \phi_1$, and

$$W_t = Z_t - \phi^{(1)} Z_{t-1} = a_t - \theta_1 a_{t-1}$$

is an MA(1) series the ACF of which can be used to identify MA order 1.

SCM approach

1. Scalar component model: y_t is a SCM of order (r, s) if (a) $y_t = \boldsymbol{v}_0' \boldsymbol{Z}_t$, (b) there exists $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ with $\boldsymbol{v}_r \neq 0$ such that $w_t = y_t + \sum_{i=1}^{r} \boldsymbol{v}_i' \boldsymbol{Z}_{t-i}$ is correlated with $\boldsymbol{Z}_{t-s}$, but not with $(\boldsymbol{Z}_{t-s-1}, \boldsymbol{Z}_{t-s-2}, \ldots)$.
2. Ideas: Seek k linearly independent SCM (p_i, q_i) such that $p_i + q_i$ are as small as possible. Then, $p = \max\{p_i\}$, $q = \max\{q_i\}$, and can easily identifiable redundant parameters. For any two SCMs, number of redundant parameters is $\delta = \min\{p_1 - p_2, q_1 - q_2\}$. Tiao and Tsay (1989, JRSSB).
3. Procedure: Use canonical correlation analysis between some expanded series of $\boldsymbol{Z}_t$.

SCM concept

Consider the canonical correlation analysis between

$$\boldsymbol{Z}_t \quad \text{and} \quad \boldsymbol{Z}_{m,t} = (\boldsymbol{Z}'_{t-1}, \ldots, \boldsymbol{Z}'_{t-m})'.$$

A zero canonical correlation implies a linear combination of $\boldsymbol{Z}_t$ which is not correlated with the first m lags of the past. For sufficiently large m, this means the linear combination is a white noise, SCM(0,0). **Next**, consider

$$(\boldsymbol{Z}'_t, \boldsymbol{Z}'_{t-1})' \quad \text{and} \quad \boldsymbol{Z}_{m,t} = (\boldsymbol{Z}'_{t-1}, \ldots, \boldsymbol{Z}'_{t-m})'.$$

Zero cano. corre. implies a linear combination of $(\boldsymbol{Z}'_t, \boldsymbol{Z}'_{t-1})'$ is uncorrelated with the past, SCM(1,0).
Complication arises and need to sort them out.

Kronecker index

1. Define $\boldsymbol{P}_{t-1} = (\boldsymbol{Z}'_{t-1}, \boldsymbol{Z}'_{t-2}, \ldots)'$ and $\boldsymbol{F}_t = (\boldsymbol{Z}'_t, \boldsymbol{Z}'_{t+1}, \boldsymbol{Z}'_{t+2}, \ldots)'$ as the Past and Future vectors.
2. Define the Hankel matrix

$$\boldsymbol{H}_\infty = \mathrm{Cov}(\boldsymbol{F}_t, \boldsymbol{P}_t),$$

   which is in the Toeplitz form.
3. Ideas: Rank of $\boldsymbol{H}_\infty$, say m, and its first m linearly independent rows determine the structure of $\boldsymbol{Z}_t$. The result is an identifiable VARMA(p, p) type of model, where $p = \max\{k_i\}$ with k_i being the Kronecker index of z_{it}.
4. Procedure: Approximate $\boldsymbol{P}_t$ by truncation. Use canonical correlation analysis between $\boldsymbol{P}_t$ and subsets of $\boldsymbol{F}_t$ to specify k_i.

Estimation

- VAR models
  1. OLS $=$ GLS
  2. MLE
  3. Bayesian approach
- VARMA models
  1. Conditional MLE
  2. Exact MLE

VARMA estimation is time consuming. MTS package needs upgrade such as using C++.

Unit-root and co-integration

Idea

1. Unit roots: strong serial dependence ($= 1$ in theory in all lags), variance goes to infinity
2. Co-integration: common sources of strong serial correlation

Linear combinations of unit-root series become stationary (without unit root)

Modeling:

1. Unit root: differencing, e.g. $x_t = y_t - y_{t-1}$, i.e. increment
2. Co-integration: Error-correction models

Consequence of improper handling: Leads to non-invertible models, cannot be approximated by VAR models

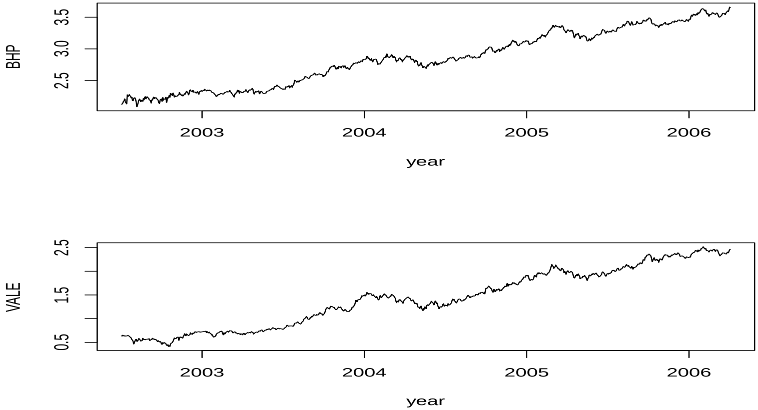

Figure: Time plots of log prices of BHP and VALE stocks: 2002.7 to 2006.12

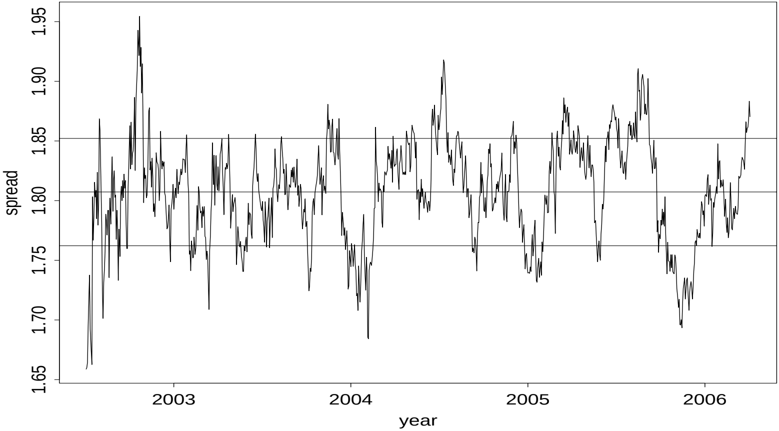

Figure: Example of statistical arbitrage

Factor models

Studying high-dimensional series: to achieve dimension reduction and ease in interpretation

Basic model: The orthogonal factor model

$$\boldsymbol{x}_t = \boldsymbol{L}\boldsymbol{f}_t + \epsilon_t,$$

where $\boldsymbol{L}$ is an $N \times m$ loading matrix, $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{mt})'$ is the m-dimensional common factors, $\boldsymbol{f}_t$ and ϵ_t are orthogonal. $\mathrm{Cov}(\epsilon_t)$ is diagonal. Often, assume Gaussian distribution.

Estimation: Principal component or maximum likelihood method

Approximate factor models

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{L}\boldsymbol{f}_t + \epsilon_t \\ y_{t+h} &= \beta' \boldsymbol{f}_t + \gamma' \boldsymbol{w}_t + v_{t+h} \end{aligned}$$

where $\boldsymbol{x}_t$ is an N-dimensional random vector, $\boldsymbol{L}$ is an $N \times r$ loading matrix, $\boldsymbol{f}_t$ is the r-dimensional common factors, $\boldsymbol{w}_t$ is a pre-determined vector that may contain lagged values of y_t, $h > 0$ is the forecast horizon, ϵ_t and v_t are the noise terms, respectively. Usual assumptions:

- All variables have zero means.
- $E(\boldsymbol{f}_t \boldsymbol{f}_t') = \boldsymbol{I}_r$.
- $E(\epsilon_t \epsilon_t') = \boldsymbol{\Psi}$ (positive definite)
- $E(\boldsymbol{f}_t \epsilon_t') = \boldsymbol{0}$, $E(\boldsymbol{f}_t v_{t+h}) = \boldsymbol{0}$, & $E(\boldsymbol{w}_t v_{t+h}) = \boldsymbol{0}$.
- $\text{Rank}(\boldsymbol{L}) = r$ and $\frac{1}{N}\boldsymbol{L}'\boldsymbol{L}$ positive definite as $N \to \infty$.
- Additional conditions needed if $\boldsymbol{\Psi}$ is not diagonal, i.e. bounded eigenvalues.

Discussions

This is the diffusion index approach of Stock and Watson.
Some difficulties often encountered when N is large:

- Hard to understand or interpret the estimated common factors.
- Does a large N produce more accurate forecasts? (Not necessarily so)
- y_t plays no role in factor estimation.
- Does not make use of any prior information or theory or past experience.

Recent research focuses on overcoming these weaknesses.

Constrained factor model

$\boldsymbol{H}$ is an $N \times m$ matrix of **known** constraints. The model becomes

$$\boldsymbol{x}_t = \boldsymbol{H}\boldsymbol{\omega}\boldsymbol{f}_t + \epsilon_t$$

where $\boldsymbol{\omega}$ is an $m \times r$ matrix, $\text{Rank}(\boldsymbol{H}) = m$ and $\text{Rank}(\boldsymbol{\omega}) = r$. Typically, $r \leq m << N$. Tsai & Tsay (2010, JASA)

Examples:

- For stock returns, columns of $\boldsymbol{H}$ may indicate the industrial sectors of the stock.
- For interest rates, columns $\boldsymbol{H}$ may indicate *level, slope* and *curvature* of the yield curve.

Motivating example

Monthly excess returns of 10 stocks: (less 3-month T bill)

(a) Pharmaceutical: Abbott Labs, Eli Lilly, Merck, and Pfizer
(b) Auto: General Motors and Ford
(c) Oil: BP, Chevron, Royal Dutch, and Exxon-Mobil

Sample period: January 1990 to December 2003 for 168 observations.

Example continued: traditional factors

Results of traditional PCA using correlations:

- Eig. Values: 3.890, 1.971, 1.498, 0.586, 0.498, ..., 0.242
- first 3 vectors:

abt	0.280	-0.355	0.1196
lly	0.244	-0.463	0.0110
mrk	0.296	-0.432	0.0462
pfe	0.337	-0.337	0.1115
gm	0.249	0.007	-0.6311
f	0.180	0.070	-0.7030
bp	0.351	0.326	0.1977
cvx	0.376	0.346	0.1318
rd	0.411	0.244	0.1366
xom	0.364	0.261	0.0574

Example continued.

Make use of the knowledge of three industries:

$$\boldsymbol{H}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Perform a constrained analysis: (least-squares estimates)
Eigen Values:

- Constrained space: 3.813, 1.917, 1.362
- Residual space: 0.660, 0.575, 0.517, ..., 0.256.

Example continued: Loading matrix

stock	Unconstrained $\boldsymbol{L}$			Constrained $\boldsymbol{H\omega}$		
abt	0.551	-0.497	0.141	0.568	-0.556	0.074
lly	0.480	-0.649	0.013	0.568	-0.556	0.074
mrk	0.583	-0.605	0.054	0.568	-0.556	0.074
pfe	0.663	-0.471	0.131	0.568	-0.556	0.074
gm	0.490	0.009	-0.744	0.423	0.071	-0.783
f	0.353	0.098	-0.829	0.423	0.071	-0.783
bp	0.690	0.457	0.233	0.736	0.409	0.168
cvx	0.739	0.485	0.155	0.736	0.409	0.168
rd	0.809	0.342	0.161	0.736	0.409	0.168
xom	0.715	0.365	0.068	0.736	0.409	0.168

Example continued.

Discussions:

- Constrained model is more parsimonious (10×3 vs. 3×3)
- Sector variations explain the variability in the excess returns (equal loading for stocks in the same industry)
- The spaces spanned by the common factors are essentially the same with/without constraints
  Canonical correlations between the two sets of common factors are
  $$0.9997, 0.9990, 0.9952.$$
- Both maximum likelihood and least squares estimations available
- Test is available for checking the constraints. Tsai and Tsay (2010, JASA)

Partially constrained factor models

In practice, it is likely that only partial constraints are available.

$$
\begin{aligned}
\boldsymbol{x}_t &= \boldsymbol{H}\omega\boldsymbol{f}_t + \boldsymbol{L}\boldsymbol{g}_t + \epsilon_t, \\
y_{t+h} &= \beta_1'\boldsymbol{f}_t + \beta_2'\boldsymbol{g}_t + v_{t+h}, \quad t = 1, \ldots, T,
\end{aligned}
$$

where $\boldsymbol{L}$ is an $N \times p$ unconstrained loading matrix of rank p and $\boldsymbol{g}_t$ is a p-dimensional unconstrained common factors.
Additional assumptions:
$E(\boldsymbol{g}_t) = \boldsymbol{0}$, $E(\boldsymbol{g}_t\boldsymbol{g}_t') = \boldsymbol{I}_p$, $E(\boldsymbol{f}_t\boldsymbol{g}_t') = 0$ and $\boldsymbol{H}'\boldsymbol{L} = \boldsymbol{0}$.
$E(\boldsymbol{g}_t v_{t+h}) = \boldsymbol{0}$

Doubly constrained factor models: an extension

Data matrix in the form

$$\boldsymbol{Z} = \boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}' + \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2' + \boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}' + \boldsymbol{E},$$

where $\boldsymbol{Z}$ is a $T \times N$ data matrix, $\boldsymbol{H}$ is the column constraints, $\boldsymbol{G}$ denotes row constraints, $\boldsymbol{F}_i$ are common factors, and $\boldsymbol{\omega}_i$ are parameters of the loading matrices.

See Tsai, Tsay, Lin and Cheng (2013) with applications to monthly U.S. regional housing starts. Needs further study when $N \to \infty$.

Dynamic factor models

Proposed by Forni, Hallin, Lippi, and Reichlin (2000, 2004, 2005)

$$\boldsymbol{x}_t = L(B)\boldsymbol{u}_t + \boldsymbol{\epsilon}_t,$$

where $L(B) = \boldsymbol{L}_0 + \boldsymbol{L}_1 B + \cdots + \boldsymbol{L}_p B^p$ is a matrix polynomial, $\boldsymbol{u}_t$ is a white-noise series, $\boldsymbol{\epsilon}_t$ is defined as before, and $\boldsymbol{u}_t$ and $\boldsymbol{\epsilon}_t$ are orthogonal.

- Identification
- Hard to estimate
- Deserve further study

How many common factors?

1. Extensively studied in the literature
2. Is it relevant?

Consider an example

$$\boldsymbol{x}_t = \frac{2}{1-0.8B} f_t + \epsilon_t,$$

where f_t is a scalar variable. Since

$$\frac{1}{1-0.8B} = 1 + 0.8B + 0.64B^2 + 0.8^3 B^3 + \cdots$$

The model becomes

$$\boldsymbol{x}_t = 2f_t + 1.6f_{t-1} + 1.28f_{t-2} + \cdots + \epsilon_t$$

which has infinite many common factors under the usual framework of factor model.

Principal Component Analysis (PCA)

PCA can be applied to the observed series or residuals of a fitted model.

Example. Consider the 4-dimensional monthly time series $\boldsymbol{Z}_t = (z_{1t}, \ldots, z_{4t})'$ of U.S. manufacturers data on durable goods,

1. z_{1t}: New orders (NO),
2. z_{2t}: Total inventory (TI),
3. z_{3t}: Unfilled orders (UO),
4. z_{4t}: Values in shipments (VS),

in billions of U.S. dollars and the data are seasonally adjusted. The sample period is from February 1992 to July 2012 for 246 observations.

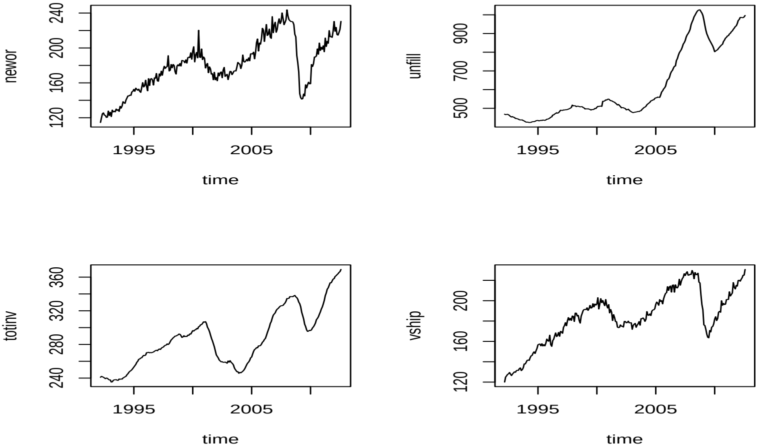

Figure: Monthly series of U.S. manufacturers data on durable goods from February 1992 to July 2012: (a) new orders, (b) total inventory, (c) unfilled orders, and (d) values of shipments. Data are in billions of dollars and seasonally adjusted.

Table: Summary of PCA Applied to the Monthly U.S. Manufacturers Data of Durable Goods From 1992.2 to 2012.7. $\hat{\boldsymbol{a}}_{p,t}$ denotes the residuals of a VAR(p) model.

Series	Variable	Principal Components			
$\boldsymbol{Z}_t$	Stand. Dev.	197.00	30.700	12.566	3.9317
	Proportion	0.9721	0.0236	0.0040	0.0004
$\hat{\boldsymbol{a}}_{1,t}$	Stand. Dev.	8.8492	3.6874	1.5720	0.3573
	Proportion	0.8286	0.1439	0.0261	0.0014
$\hat{\boldsymbol{a}}_{2,t}$	Stand. Dev.	8.3227	3.5233	1.1910	0.2826
	Proportion	0.8327	0.1492	0.0171	0.0010
$\hat{\boldsymbol{a}}_{3,t}$	Stand. Dev.	8.0984	3.4506	1.0977	0.2739
	Proportion	0.8326	0.1512	0.01530	0.0010
$\hat{\boldsymbol{a}}_{4,t}$	Stand. Dev.	7.8693	3.2794	1.0510	0.2480
	Proportion	0.8386	0.1456	0.0140	0.0008

Table: Loadings of PCA Applied to the Monthly U.S. Manufacturers Data of Durable Goods, where ts stands for Time Series.

ts	Loading matrix				ts	Loading matrix			
$\boldsymbol{Z}_t$	0.10	0.71	0.342	0.604	$\hat{\boldsymbol{a}}_{1,t}$	0.79	0.16	0.066	0.583
	0.15	0.32	-0.928	0.129		0.06	-0.11	-0.990	0.063
	0.98	-0.18	0.098	-0.006		0.55	-0.59	0.060	-0.588
	0.10	0.60	0.110	-0.786		0.26	0.78	-0.106	-0.557
$\hat{\boldsymbol{a}}_{2,t}$	0.80	0.15	0.017	0.587	$\hat{\boldsymbol{a}}_{3,t}$	0.80	0.14	0.009	0.586
	0.03	-0.07	-0.997	0.012		0.02	-0.06	-0.998	0.007
	0.54	-0.60	0.049	-0.585		0.54	-0.61	0.044	-0.583
	0.27	0.78	-0.055	-0.560		0.27	0.78	-0.048	-0.563

Eigenvectors of the smallest e.v

The eigenvector associated with the 4th can be written as
$\boldsymbol{h}_4 \approx (1, 0, -1, -1)'$.
Next, the fitted VAR(1) model is

$$\boldsymbol{Z}_t = \begin{bmatrix} 0.01 \\ -0.13 \\ -8.35 \\ 2.80 \end{bmatrix} + \begin{bmatrix} 0.686 & -0.027 & -0.001 & 0.357 \\ 0.116 & 0.995 & -0.000 & -0.102 \\ 0.562 & -0.023 & 0.995 & -0.441 \\ 0.108 & 0.023 & -0.003 & 0.852 \end{bmatrix} \boldsymbol{Z}_{t-1} + \hat{\boldsymbol{a}}_{1,t}. \tag{1}$$

Pre-multiplying Equation (1) by $\boldsymbol{h}_4'$, we have

$$\boldsymbol{h}_4'\boldsymbol{z}_t \approx 5.55 + (0.015, -0.027, -0.994, -0.054)\boldsymbol{z}_{t-1} + \boldsymbol{h}_4'\hat{\boldsymbol{a}}_{1,t}.$$

The information points to $\boldsymbol{h}_4'\hat{\boldsymbol{a}}_{1,t} \approx 0$. (Why?) Consequently, the prior equation implies

$$\mathrm{NO}_t - \mathrm{UO}_t - \mathrm{VS}_t + \mathrm{UO}_{t-1} \approx c_4,$$

where c_4 denotes a constant. In other words, PCA of the residuals of the VAR(1) model reveals a stable relation

$$\mathrm{NO}_t - \mathrm{VS}_t - (\mathrm{UO}_t - \mathrm{UO}_{t-1}) \approx c_4 \tag{2}$$

Results continue to hold for higher VAR models.
Consider the VAR(2) model,

$$\boldsymbol{Z}_t = \widehat{\phi}_{2,0} + \widehat{\boldsymbol{\Phi}}_{2,1}\boldsymbol{Z}_{t-1} + \widehat{\boldsymbol{\Phi}}_{2,2}\boldsymbol{Z}_{t-1} + \hat{\boldsymbol{a}}_{2,t}. \tag{3}$$

From PCA of the residuals $\hat{\boldsymbol{a}}_{2,t}$, the smallest eigenvalue is close to zero with eigenvector $\boldsymbol{h}_4 \approx (1, 0, -1, -1)'$.
Pre-multiplying Equation (3) by $\boldsymbol{h}_4'$, we get

$$\boldsymbol{h}_4'\boldsymbol{Z}_t \approx 2.21 + (.59, -.08, -1.57, -.61)\boldsymbol{Z}_{t-1} + (.01, .07, .57, -.01)\boldsymbol{Z}_{t-2}.$$

Consequently, we have

$$z_{1t} - z_{3t} - z_{4t} - 0.59 z_{1,t-1} + 1.57 z_{3,t-1} + 0.61 z_{4,t-1} - 0.57 z_{3,t-2} \approx c_1,$$

where c_1 denotes a constant. Rearranging terms, the prior equation implies

$$z_{1t} - z_{3t} - z_{4t} + z_{3,t-1} - (0.59 z_{1,t-1} - 0.57 z_{3,t-1} - 0.61 z_{4,t-1} + 0.57 z_{3,t-2}) \approx c$$

This approximation further simplifies as

$$(z_{1t} - z_{3t} - z_{4t} + z_{3,t-1}) - 0.59(z_{1,t-1} - z_{3,t-1} - z_{4,t-1} + z_{3,t-2}) \approx c,$$

where c is a constant.

Table: Summary of the VAR(2) Model for the 4-Dimensional Time Series of Monthly Manufacturers Data on Durable Goods

Parameter	Estimates			
$\widehat{\phi}_{2,0}'$	-0.221	3.248	-6.267	3.839
$\widehat{\mathbf{\Phi}}_{2,1}$	1.033	1.012	-0.638	-0.108
	-0.445	1.549	0.441	0.537
	0.307	0.645	1.005	-0.120
	0.141	0.452	-0.072	0.619
$\widehat{\mathbf{\Phi}}_{2,2}$	0.243	-1.028	0.634	-0.115
	0.064	-0.568	-0.438	-0.166
	0.247	-0.663	-0.010	-0.336
	-0.016	-0.440	0.070	0.227

Models for time series of count data

Two basic categories

1. Parameter driven
2. Observation driven

Consider the trading intensity in high-frequency finance. For example, the number of trades in 30 seconds of a given asset.

Two references

- Regression Analysis of Count Data by A. C. Cameron and P.K. Trivedi (2013), 2nd edition, Cambridge Press
- *Econometric Analysis of Count Data* by R. Winkelmann (2010), 5th edition, Springer

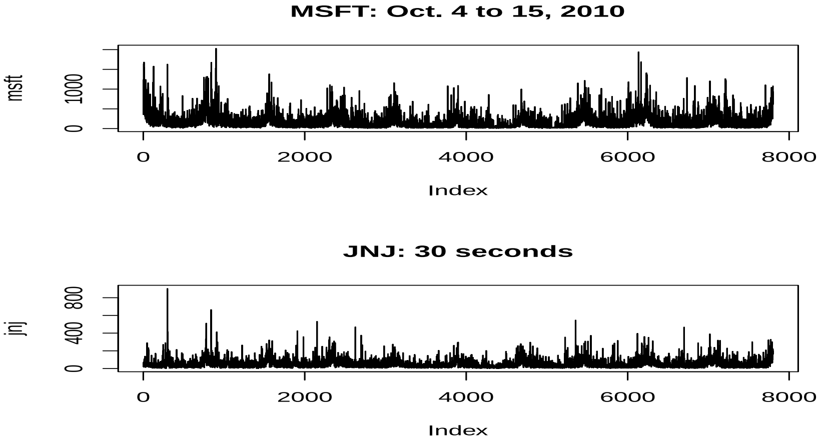

Figure: Time series plots of the number of trades of MSFT and JNJ, 30 seconds, October 4 to 15, 2010

Univariate autoregressive intensity models

Let F_{t-1} denote the information available at time $t-1$ and y_t be the time series of counts. A simple autoregressive conditional Poisson model is

$$\begin{aligned} y_t | F_{t-1} &\sim Po(\lambda_t) \\ \lambda_i &= \omega + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} \end{aligned}$$

where

$$Po(\lambda) = \frac{\exp(-\lambda)\lambda^{y_t}}{y_t!}.$$

Key features:

$$E(y_t | F_{t-1}) = \mathrm{Var}(y_t | F_{t-1}) = \lambda_t.$$

Generalizations

1. Negative Binomial (r, p): $\lambda = rp/(1-p)$ so that

   $$E(y_t|F_{t-1}) = rp_t/(1-p_t) = \lambda_t, \quad \mathrm{Var}(y_t|F_{t-1}) = \lambda_t + \lambda_t^2/r$$

   where $p_t = \lambda_t/(r + \lambda_t)$.
2. Double Poisson of Efron (1986)

   $$y_t|F_{t-1} \sim DPo(\lambda_t, \gamma),$$

   where

   $$E(y_t|F_{t-1}) = \lambda_t, \quad \mathrm{Var}(y_t|F_{t-1}) \approx \lambda_t/\gamma.$$

Use of exogenous variables

To handle the diurnal pattern in HF trading, use

$$y_t | F_{t-1} \sim Po(\lambda_t \exp(s_t)),$$

where

$$s_t = \sum_{i=1}^{k} \theta_i x_{i,t},$$

with $x_{i,t}$ being a given function, e.g. indicator or sine (co-sine) function.

Multivariate generalization: possible with some modifications. See, for example, Tsay (2014, JSM meeting) with momentum effect.

Johnson and Johnson trading intensity

1. Sample period: October 4 to October 15, 2010 for 10 trading days.
2. Time interval: 30 seconds.
3. Sample size: 7800
4. Models entertained: Conditional autoregressive models with Poisson, Negative binomial, and double Poisson
5. Order: (1,1)
6. Seasonality: indicators of the first 4 and last 4 time intervals.

The model is in the form

$$\text{rate}_t = \lambda_t \exp\left(\sum_i \theta_i x_{i,t}\right),$$

$$\lambda_t = \omega + \alpha y_{t-1} + \beta \lambda_{t-1}.$$

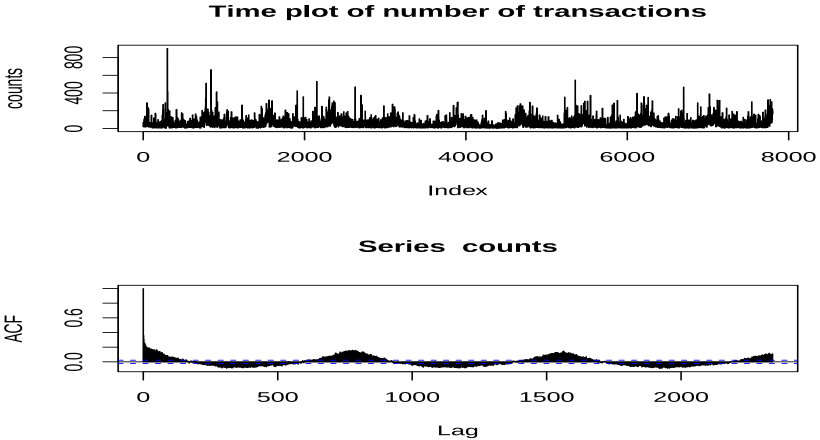

Figure: Time series plot and ACF of the number of trades of JNJ, 30 seconds, October 4 to 15, 2010

Estimation results of PCA models: JNJ data

Par	Po		NB		DP	
	est	se	est	se	est	se
ω	1.56	0.05	1.11	0.20	2.27	0.29
α	0.12	.002	0.10	.008	0.11	.008
β	0.85	.002	0.88	0.01	0.85	0.01
γ			1.73	0.03	0.039	0.001
θ_1	-0.26	0.03	-0.15	0.26	-0.24	0.16
θ_2	-0.26	0.03	-0.24	0.24	-0.23	0.16
θ_3	-0.18	0.03	-0.10	0.24	-0.16	0.15
θ_4	-0.15	0.03	-0.16	0.24	-0.13	0.16
θ_5	0.68	0.01	0.69	0.24	0.69	0.10
θ_6	0.48	0.02	0.48	0.24	0.49	0.11
θ_7	0.51	0.02	0.51	0.24	0.52	0.11
θ_8	0.32	0.03	0.31	0.24	0.33	0.13

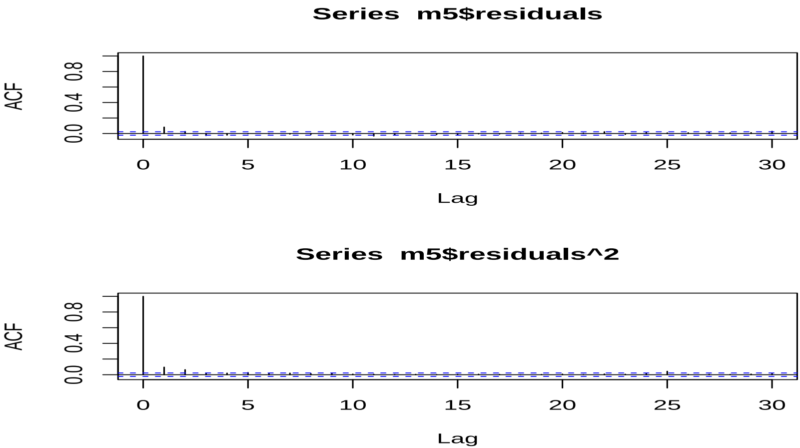

Figure: ACF of the residuals and squared residuals of fitted PCA model with double Poisson innovations, 30 seconds, October 4 to 15, 2010

Summary of empirical analysis

1. The models fit the data reasonably well
2. The trading intensity has high serial dependence (persistent)
3. Trading is not a Poisson process (over-dispersion)
4. The opening dummies become statistically insignificant when over-dispersion is entertained. The closing dummies, however, remain significant.

How about multivariate analysis?

Some thoughts

1. Limited works available
2. *Dynamic factor models for multivariate count data*, Jung et al. (2009, JBES). Use efficient important sampling method.
3. How to model dependence (serial and cross section)?
   - Common factor in intensity functions λ_{it}
   - Common factor in observations

$$y_{it} = y_{c,t} + \cdots$$

where $y_{c,t}$ is a latent series common to all individual components, where $y_{c,t} \leq \min\{y_{it}\}$

Prediction with many predictors

Approaches proposed

1. Diffusion index: generalization of principal component regression
2. LASSO-family: penalized likelihood approach
3. Partial least squares (PLS)
4. Model-based clustering approach
5. Group diffusion or group PLS

Predicting the VIX index

VIX index is commonly known as the U.S. fear factor in stock market. It is the daily volatility index of Chicago Board Options Exchange.

Problem of interest: long-term forecasts of VIX. Let T be the forecast origin. Interested in predicting

$$V_{T+h},$$

for $h = 60$ or 120 (3 month or 6 month ahead)
Data available, including

- Interest rate volatility
- FX volatility
- equity volatility

Volatility matrix

Let $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{kt})'$ be the returns of k assets at time t, and F_t be the public information available at time t.
Assume $E(\boldsymbol{r}_t | F_{t-1}) = \boldsymbol{0}$.
Definition: Volatility matrix $\boldsymbol{\Sigma}_t = \text{Cov}(\boldsymbol{r}_t | F_{t-1})$, conditional covariance matrix.

Why is it important?

1. Use it to quantify joint financial risk
2. Needed in asset allocation (portfolio re-balancing)

Challenges

Difficulties

1. There are $k(k+1)/2$ processes of variance and covariances
2. Time-varying
3. $\boldsymbol{\Sigma}_t$ must be positive definite almost surely

Multivariate volatility models

Models available in the MTS package

1. BEKK(1,1) models for $k = 2$ and 3 ONLY
2. Dynamic conditional correlation (DCC) models of Engle (2002) and Tse and Tsui (2002)
3. Cholesky decomposition model
4. Some copula models with multivariate Student-t innovations

Other possibilities (not in MTS package):

1. Stochastic volatility models
2. Factor structure

Remarks on volatility modeling

1. Available models are either highly structured or computational expensive
2. Some directions of current research
   - 2.1 Use high-frequency, transaction-by-transaction, data
   - 2.2 Use common factors, including independent component analysis
   - 2.3 Use hyper-spherical coordinates to put parameter constraints