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 \( Z_t = (z_{1t}, \ldots, z_{kt})' \) 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:**

\[
Z_t = C + \sum_{i=0}^{\infty} \psi_i a_{t-i},
\]

\( C \) is a constant vector, \( \psi_0 = I \), \( \psi_i \) are \( k \times k \) real matrices, \( \{a_t\} \) are \( k \)-dimensional iid noises with mean zero & \( \text{cov} = \Sigma_a > 0 \).

- **Invertibility:**

\[
Z_t = C + \sum_{i=1}^{\infty} \pi_i Z_{t-i} + a_t,
\]

where \( \pi_i \) are \( k \times k \) real matrices.
Parameterization

Consequences of parameterization:

▶ Why $\Sigma_a > 0$?

▶ Use Cholesky decomposition: $\Sigma_a = L\Omega L'$, where $\Omega$ is a diagonal matrix, $L$ is lower triangular with 1 on the diagonal. Define $b_t = L^{-1}a_t$. The series can be rewritten as

$$Z_t = C + \sum_{i=0}^{\infty} \psi_i^* b_{t-i},$$

where $\psi_i^* = \psi_i L$ and $\text{cov}(b_t) = \Omega$ is diagonal, and $\psi_0^* = L$, a lower triangular matrix with unit diagonal elements.

▶ The model can also be written as

$$L^{-1}Z_t = C^* + b_t + \sum_{i=1}^{\infty} \omega_i b_{t-i},$$

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

Two difficulties encountered in MTS modeling:

1. Too many parameters
2. Model identification: identifiability

Vector autoregressive moving-average (VARMA) model:

$$\phi(B)(Z_t - \mu) = \theta(B)a_t,$$

$\mu$ a constant vector, $a_t \sim N(0, \Sigma_a)$, and

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

with $B$ denoting the back-shift (or lag) operator.
Two simple examples with $k = 2$: First,

$$Z_t = a_t - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} a_{t-1} \iff Z_t - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} Z_{t-1} = a_t$$

That is, $\text{VMA}(1) = \text{VAR}(1)$. Next,

$$Z_t - \begin{bmatrix} 0.8 & 2 \\ 0 & 0 \end{bmatrix} Z_{t-1} = a_t - \begin{bmatrix} 0.3 & 0 \\ 0 & 0 \end{bmatrix} a_{t-1}$$

is the same as

$$Z_t - \begin{bmatrix} 0.8 & 2 + \omega \\ 0 & \beta \end{bmatrix} Z_{t-1} = a_t - \begin{bmatrix} 0.3 & \omega \\ 0 & \beta \end{bmatrix} a_{t-1},$$

where $\omega$ and $\beta$ are arbitrary.
Assumptions for block-identifiability:
1. $\phi(B)$ and $\theta(B)$ are left co-prime: Any common left factor must be unimodular
2. Rank of $[\phi_p, \theta_q]$ 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


- 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
  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.
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

A simple VAR(5) model:

\[
Z_t = \begin{bmatrix} 1.35 \\ 0 \end{bmatrix} + \begin{bmatrix} .707 & .0179 \\ 4.788 & .2818 \end{bmatrix} Z_{t-1} - \begin{bmatrix} 0 & 0 \\ 5.74 & .412 \end{bmatrix} Z_{t-2} \\
+ \begin{bmatrix} 0 & 0 \\ 4.14 & 0 \end{bmatrix} Z_{t-3} + \begin{bmatrix} .296 & 0 \\ 0 & -.193 \end{bmatrix} Z_{t-4} \\
- \begin{bmatrix} .157 & .023 \\ 2.82 & .166 \end{bmatrix} Z_{t-5} + a_t, \quad \Sigma = \begin{bmatrix} .627 & .259 \\ .259 & 167.61 \end{bmatrix}.
\]
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
Consider a $k$-dimensional series $Z_t$.

**ECCM method:**

1. Requirement: The model is block identifiable: (a) $\phi(B)$ and $\theta(B)$ are left co-prime, i.e. $\phi(B) = U(B)\phi^*(B)$ and $\theta(B) = U(B)\theta^*(B)$, then $|U(B)|$ is a constant. (b) $\text{Rank}[\phi_p, \theta_q] = k$.

2. Idea: (a) Seek consistent LS estimates of $\phi(B)$. (b) Transform $Z_t$ by $W_t = \hat{\phi}(B)Z_t$. (c) Use cross-correlation matrices of $W_t$ to specify $q$.

3. Procedure: (a) is achieved by iterated vector autoregressions.
Consider $k = 1$ and $Z_t$ follows $Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1}$.

The auto-regression

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

gives $\phi^{(0)} = \rho_1$, which is not $\phi_1$. However, the iterated auto-regression

$$Z_t = \phi^{(1)} Z_{t-1} + \beta \epsilon^{(0)}_{t-1} + \epsilon^{(1)}_t$$

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.
1. Scalar component model: $y_t$ is a SCM of order $(r, s)$ if (a) $y_t = v'_0 Z_t$, (b) there exists $v_1, \ldots, v_r$ with $v_r \neq 0$ such that $w_t = y_t + \sum_{i=1}^{r} v'_i Z_{t-i}$ is correlated with $Z_{t-s}$, but not with $(Z_{t-s-1}, Z_{t-s-2}, \ldots)$.

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 $Z_t$. 
Consider the canonical correlation analysis between $Z_t$ and $Z_{m,t} = (Z'_{t-1}, \ldots, Z'_{t-m})'$. 

A zero canonical correlation implies a linear combination of $Z_t$ 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 

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

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

1. Define \( P_{t-1} = (Z'_{t-1}, Z'_{t-2}, \ldots)' \) and \( F_{t} = (Z'_t, Z'_{t+1}, Z'_{t+2}, \ldots)' \) as the Past and Future vectors.

2. Define the Hankel matrix

\[
H_\infty = \text{Cov}(F_t, P_t),
\]

which is in the Toeplitz form.

3. Ideas: Rank of \( H_\infty \), say \( m \), and its first \( m \) linearly independent rows determine the structure of \( Z_t \). 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 \( P_t \) by truncation. Use canonical correlation analysis between \( P_t \) and subsets of \( F_t \) 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

\[ x_t = Lf_t + \epsilon_t, \]

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

Estimation: Principal component or maximum likelihood method
Approximate factor models

\[ x_t = Lf_t + \epsilon_t \]
\[ y_{t+h} = \beta'f_t + \gamma'w_t + \nu_{t+h} \]

where \( x_t \) is an \( N \)-dimensional random vector, \( L \) is an \( N \times r \) loading matrix, \( f_t \) is the \( r \)-dimensional common factors, \( w_t \) is a pre-determined vector that may contain lagged values of \( y_t \), \( h > 0 \) is the forecast horizon, \( \epsilon_t \) and \( \nu_t \) are the noise terms, respectively.

Usual assumptions:

- All variables have zero means.
- \( E(f_t f_t') = I_r \).
- \( E(\epsilon_t \epsilon_t') = \Psi \) (positive definite)
- \( E(f_t \epsilon_t') = 0, E(f_t \nu_{t+h}) = 0, \) and \( E(w_t \nu_{t+h}) = 0 \).
- \( \text{Rank}(L) = r \) and \( \frac{1}{N}L'L \) positive definite as \( N \to \infty \).
- Additional conditions needed if \( \Psi \) is not diagonal, i.e. bounded eigenvalues.
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

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

$$x_t = H\omega f_t + \epsilon_t$$

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

Examples:

- For stock returns, columns of $H$ may indicate the industrial sectors of the stock.
- For interest rates, columns $H$ 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:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>abt</td>
<td>0.280</td>
<td>-0.355</td>
<td>0.1196</td>
</tr>
<tr>
<td>lly</td>
<td>0.244</td>
<td>-0.463</td>
<td>0.0110</td>
</tr>
<tr>
<td>mrk</td>
<td>0.296</td>
<td>-0.432</td>
<td>0.0462</td>
</tr>
<tr>
<td>pfe</td>
<td>0.337</td>
<td>-0.337</td>
<td>0.1115</td>
</tr>
<tr>
<td>gm</td>
<td>0.249</td>
<td>0.007</td>
<td>-0.6311</td>
</tr>
<tr>
<td>f</td>
<td>0.180</td>
<td>0.070</td>
<td>-0.7030</td>
</tr>
<tr>
<td>bp</td>
<td>0.351</td>
<td>0.326</td>
<td>0.1977</td>
</tr>
<tr>
<td>cvx</td>
<td>0.376</td>
<td>0.346</td>
<td>0.1318</td>
</tr>
<tr>
<td>rd</td>
<td>0.411</td>
<td>0.244</td>
<td>0.1366</td>
</tr>
<tr>
<td>xom</td>
<td>0.364</td>
<td>0.261</td>
<td>0.0574</td>
</tr>
</tbody>
</table>
Example continued.

Make use of the knowledge of three industries:

\[
H' = \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix}.
\]

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

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

- Constrained model is more parsimonious ($10 \times 3$ vs. $3 \times 3$)
- 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.

\[ x_t = H\omega f_t + Lg_t + \epsilon_t, \]
\[ y_{t+h} = \beta'_1 f_t + \beta'_2 g_t + v_{t+h}, \quad t = 1, \ldots, T, \]

where \( L \) is an \( N \times p \) unconstrained loading matrix of rank \( p \) and \( g_t \) is a \( p \)-dimensional unconstrained common factors.

Additional assumptions:
\[ E(g_t) = 0, \quad E(g_t g_t') = I_p, \quad E(f_t g_t') = 0 \quad \text{and} \quad H'L = 0. \]
\[ E(g_t v_{t+h}) = 0. \]
Doubly constrained factor models: an extension

Data matrix in the form

\[ Z = F_1 \omega_1' H' + GF_2 \omega_2' + GF_3 \omega_3' H' + E, \]

where \( Z \) is a \( T \times N \) data matrix, \( H \) is the column constraints, \( G \) denotes row constraints, \( F_i \) are common factors, and \( \omega_i \) 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


\[ x_t = L(B)u_t + \epsilon_t, \]

where \( L(B) = L_0 + L_1 B + \cdots + L_p B^p \) is a matrix polynomial, \( u_t \) is a white-noise series, \( \epsilon_t \) is defined as before, and \( u_t \) and \( \epsilon_t \) 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

\[ x_t = \frac{2}{1 - 0.8B} f_t + \epsilon_t, \]

where \( f_t \) is a scalar variable. Since

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

The model becomes

\[ x_t = 2f_t + 1.6f_{t-1} + 1.28f_{t-2} + \cdots + \epsilon_t \]

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 \( Z_t = (z_{1t}, \ldots, z_{4t})' \) 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. $\hat{a}_{p,t}$ denotes the residuals of a VAR($p$) model.

<table>
<thead>
<tr>
<th>Series</th>
<th>Variable</th>
<th>Principal Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_t$</td>
<td>Stand. Dev. Proportion</td>
<td>197.00 30.700 12.566 3.9317</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9721 0.0236 0.0040 0.0004</td>
</tr>
<tr>
<td>$\hat{a}_{1,t}$</td>
<td>Stand. Dev. Proportion</td>
<td>8.8492 3.6874 1.5720 0.3573</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8286 0.1439 0.0261 0.0014</td>
</tr>
<tr>
<td>$\hat{a}_{2,t}$</td>
<td>Stand. Dev. Proportion</td>
<td>8.3227 3.5233 1.1910 0.2826</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8327 0.1492 0.0171 0.0010</td>
</tr>
<tr>
<td>$\hat{a}_{3,t}$</td>
<td>Stand. Dev. Proportion</td>
<td>8.0984 3.4506 1.0977 0.2739</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8326 0.1512 0.01530 0.0010</td>
</tr>
<tr>
<td>$\hat{a}_{4,t}$</td>
<td>Stand. Dev. Proportion</td>
<td>7.8693 3.2794 1.0510 0.2480</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8386 0.1456 0.0140 0.0008</td>
</tr>
</tbody>
</table>
Table: Loadings of PCA Applied to the Monthly U.S. Manufacturers Data of Durable Goods, where ts stands for Time Series.

<table>
<thead>
<tr>
<th>ts</th>
<th>Loading matrix</th>
<th>ts</th>
<th>Loading matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_t$</td>
<td></td>
<td>$\hat{a}_{1,t}$</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.71</td>
<td>0.79</td>
<td>0.16</td>
</tr>
<tr>
<td>0.15</td>
<td>0.32</td>
<td>0.06</td>
<td>-0.11</td>
</tr>
<tr>
<td>0.98</td>
<td>-0.18</td>
<td>0.55</td>
<td>-0.59</td>
</tr>
<tr>
<td>0.10</td>
<td>0.60</td>
<td>0.26</td>
<td>0.78</td>
</tr>
<tr>
<td>$\hat{a}_{2,t}$</td>
<td></td>
<td>$\hat{a}_{3,t}$</td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>0.15</td>
<td>0.80</td>
<td>0.14</td>
</tr>
<tr>
<td>0.03</td>
<td>-0.07</td>
<td>0.02</td>
<td>-0.06</td>
</tr>
<tr>
<td>0.54</td>
<td>-0.60</td>
<td>0.54</td>
<td>-0.61</td>
</tr>
<tr>
<td>0.27</td>
<td>0.78</td>
<td>0.27</td>
<td>0.78</td>
</tr>
</tbody>
</table>
The eigenvector associated with the 4th can be written as $h_4 \approx (1, 0, -1, -1)'$. 

Next, the fitted VAR(1) model is 

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

(1)
Pre-multiplying Equation (1) by $h'_4$, we have

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

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

$$NO_t - UO_t - VS_t + UO_{t-1} \approx c_4,$$

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

$$NO_t - VS_t - (UO_t - UO_{t-1}) \approx c_4$$

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

\[ Z_t = \hat{\phi}_{2,0} + \hat{\phi}_{2,1} Z_{t-1} + \hat{\Phi}_{2,2} Z_{t-1} + \hat{a}_{2,t}. \]  

(3)

From PCA of the residuals \( \hat{a}_{2,t} \), the smallest eigenvalue is close to zero with eigenvector \( h_4 \approx (1, 0, -1, -1)' \). Pre-multiplying Equation (3) by \( h_4' \), we get

\[ h_4' Z_t \approx 2.21 + (.59, - .08, -1.57, - .61) Z_{t-1} + (.01, .07, .57, - .01) Z_{t-2}. \]

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\phi}_{2,0}$</td>
<td>-0.221 3.248 -6.267 3.839</td>
</tr>
<tr>
<td>$\Phi_{2,1}$</td>
<td>1.033 1.012 -0.638 -0.108 -0.445 1.549 0.441 0.537</td>
</tr>
<tr>
<td></td>
<td>0.307 0.645 1.005 -0.120</td>
</tr>
<tr>
<td></td>
<td>0.141 0.452 -0.072 0.619</td>
</tr>
<tr>
<td>$\hat{\Phi}_{2,2}$</td>
<td>0.243 -1.028 0.634 -0.115 -0.016 0.440 0.070 0.227</td>
</tr>
</tbody>
</table>
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

- 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
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

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

$$\lambda_i = \omega + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}$$

where

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

Key features:

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

1. Negative Binomial \((r, p)\): \(\lambda = rp/(1 - p)\) so that
\[
E(y_t|F_{t-1}) = rp_t/(1 - p_t) = \lambda_t, \quad \text{Var}(y_t|F_{t-1}) = \lambda_t + \lambda_t^2/r
\]
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
\[
E(y_t|F_{t-1}) = \lambda_t, \quad \text{Var}(y_t|F_{t-1}) \approx \lambda_t/\gamma.
\]
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.
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

<table>
<thead>
<tr>
<th>Par</th>
<th>Po</th>
<th>NB</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>se</td>
<td>est</td>
</tr>
<tr>
<td>$\omega$</td>
<td>1.56</td>
<td>0.05</td>
<td>1.11</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.12</td>
<td>0.002</td>
<td>0.10</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.85</td>
<td>0.002</td>
<td>0.88</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.73</td>
<td>0.03</td>
<td>0.039</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>-0.26</td>
<td>0.03</td>
<td>-0.15</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>-0.26</td>
<td>0.03</td>
<td>-0.24</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>-0.18</td>
<td>0.03</td>
<td>-0.10</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>-0.15</td>
<td>0.03</td>
<td>-0.16</td>
</tr>
<tr>
<td>$\theta_5$</td>
<td>0.68</td>
<td>0.01</td>
<td>0.69</td>
</tr>
<tr>
<td>$\theta_6$</td>
<td>0.48</td>
<td>0.02</td>
<td>0.48</td>
</tr>
<tr>
<td>$\theta_7$</td>
<td>0.51</td>
<td>0.02</td>
<td>0.51</td>
</tr>
<tr>
<td>$\theta_8$</td>
<td>0.32</td>
<td>0.03</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Ruey S. Tsay
Booth School of Business
University of Chicago

Multivariate Time Series Analysis: Brief Review and Recent Developments
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.
Some thoughts

1. Limited works available


3. How to model dependence (serial and cross section)?
   - Common factor in intensity functions $\lambda_{it}$
   - Common factor in observations

\[ y_{it} = y_{c,t} + \cdots \]

where $y_{c,t}$ is a latent series common to all individual components, where $y_{c,t} \leq \min\{y_{it}\}$
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
Let $r_t = (r_{1t}, \ldots, r_{kt})'$ be the returns of $k$ assets at time $t$, and $F_t$ be the public information available at time $t$. Assume $E(r_t|F_{t-1}) = 0$.

**Definition**: Volatility matrix $\Sigma_t = \text{Cov}(r_t|F_{t-1})$, 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. \( \Sigma_t \) 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
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
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