Dependence within Financial Markets

David S. Matteson
School of Operations Research and Information Engineering
& Department of Statistical Science
Cornell University

dm484@cornell.edu
http://people.orie.cornell.edu/matteson/

Joint work with: Ruey S. Tsay, Booth School of Business

Acknowledgments: National Science Foundation

2011 April 30
Analysis of Financial and Econometric Data

- Serial dependence

- Non-normality

- Cross dependence
Univariate Serial Dependence

- Serial correlation
  \[ \text{cor}(x_t, x_{t-\ell}) \]

- Nonlinear correlation
  \[ \text{cor}(h(x_t), h(x_{t-\ell})) \]

- Arbitrary dependence
  \[ x_t \text{ vs. } x_{t-\ell} \]

- Simultaneous measure
  \[ x_t \text{ vs. } \{x_{t-1}, \ldots x_{t-k}\} \]
Distance Covariance

Distance covariance $\mathcal{V}(x_1, x_2)$ measures dependence between r.v. $x_1 \in \mathbb{R}^{q_1}$ and $x_2 \in \mathbb{R}^{q_2}$, for all distributions with finite first moments

$$\mathcal{V}^2(x_1, x_2) = \|\phi_{x_1,x_2}(t_1, t_2) - \phi_{x_1}(t_1)\phi_{x_2}(t_2)\|_\omega^2$$

- $\phi_{x_1}$ and $\phi_{x_2}$ denote the characteristic functions of $x_1$ and $x_2$, resp.
- $\phi_{x_1,x_2}$ denotes the joint characteristic function
- $\omega(t_1, t_2)$ is a positive weight function
- $\mathcal{V}(x_1, x_2) = 0$ if and only if $x_1$ and $x_2$ are independent
Sample Distance Covariance

Random sample \((\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\), size \(n\), empirical distance covariance statistic

\[
\mathcal{V}^2_n \left( \mathbf{X}^{(1)}, \mathbf{X}^{(2)} \right) = \| \phi_{X_1X_2}^n(t_1, t_2) - \phi_{X_1}^n(t_1) \phi_{X_2}^n(t_2) \|^2 \omega
\]

\[
= \frac{1}{n^2} \sum_{k,l=1}^{n} A_{kl} B_{kl}
\]

\[
a_{kl} = \| X_k^{(1)} - X_l^{(1)} \|_{q_1}, \quad b_{kl} = \| X_k^{(2)} - X_l^{(2)} \|_{q_2}, \quad \text{for } k, l = 1, \ldots, n
\]

\[
A_{kl} = a_{kl} - \bar{a}_k - \bar{a}_l + \bar{a}_{..}, \quad B_{kl} = b_{kl} - \bar{b}_k - \bar{b}_l + \bar{b}_{..},
\]

(Székely et al., 2007)

\[\lim_{n \to \infty} \mathcal{V}_n \xrightarrow{a.s.} \mathcal{V}, \quad \text{and } n\mathcal{V}_n^2 \text{ convergences in distribution to a r.v.}\]
An Alternative Measure

- $\mathcal{V}(x_1, x_2)$ depends on marginal distributions
- Apply probability integral transformation (PIT)

$$\text{marginal CDF } F_X : \mathbb{R} \rightarrow [0, 1], \text{ define } u = F_X(x)$$

- $\mathcal{V}(u_1, u_2) = 0$ iff $x_1$ and $x_2$ are independent
- The $F_X$ are unknown
  - Use $\hat{F}_X$, marginal ranks
  - Let $\hat{u} = \hat{F}_X(x)$

Lemma

$$\mathcal{V}_n(\hat{U}_1, \hat{U}_2) \xrightarrow{a.s.} \mathcal{V}(u_1, u_2), \text{ and } n\mathcal{V}_n^2(\hat{U}_1, \hat{U}_2) \xrightarrow{D} \text{ r.v., as } n \rightarrow \infty$$
A Joint Test for Serial Dependence

- $\mathcal{V}(y_t, y_{t-k})$ measures lag-\(k\) serial dependence
- Let $y_{t-k} = \{y_{t-1}, \ldots, y_{t-k}\}$
- $\mathcal{V}(y_t, y_{t-k})$ jointly measures serial dependence up to lag-\(k\)
- Assuming stationarity, we use $\mathcal{V}(u_t, u_{t-k})$
- Joint hypothesis for serial dependence
  \[ H_0 : \phi_{u_t,u_{t-\ell}} = \phi_{u_t} \phi_{u_{t-\ell}} \text{ for all } \ell = 1, \ldots, k \]
- We define our test statistic as
  \[ Q_1(Y, k) = n \mathcal{V}_n^2(\hat{U}_t, \hat{U}_{t-k}) \]
Asymptotic Distribution

Under $H_0$,

$$Q_1(Y, k) \xrightarrow{D} \|\zeta_k(a, b)\|_\omega^2$$

$\zeta_k(\cdot, \cdot)$ mean zero complex Gaussian process with covariance function

$$R_k(c, c_0) = \left(\phi_u(a - a_0) - \phi_u(a)\phi_u(a_0)\right) \left(\phi_{u_k}(b - b_0) - \phi_{u_k}(b)\phi_{u_k}(b_0)\right)$$

for $c = (a, b), c_0 = (a_0, b_0) \in \mathbb{R} \times \mathbb{R}^k$

$\phi_u$ characteristic function of Uniform(0,1) r.v., and $\phi_{u_k} = \phi_u \cdots \phi_u$

Distribution Free Test

Analogous Multivariate Test
Seasonally Adjusted Monthly Unemployment Rates (%)
CA, FL, IL, MI, OH, & WI, from January 1976 through August 2010
Standardized Change in Monthly Unemployment Rate %

First difference series, scaled by monthly standard deviations

![Graph showing monthly unemployment rates for different states over time.](image-url)
Testing for Serial Dependence

- Transform to stationary $y_t$
  - First difference series, scaled by monthly standard deviations
  - $Q_6(Y, 12) = 121.67$ with $p$-value $\approx 0$

- Fit a vector autoregression of order three, via OLS
  $$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + e_t$$
- Calculate residuals $\hat{e}_t$
  - $Q_6(\hat{E}, 12) = 11.31$ with $p$-value $\approx 0.31$
  - $\Rightarrow$ Linear model is sufficient
Residual Series, Monthly Unemployment Rate %

No significant serial dependence
Financial & Economic Processes are Inherently Multivariate

- **Simultaneous Analysis is Crucial**

- **Multivariate Analysis is Difficult**

- **Independent Component Analysis (ICA)**
  - Representation of multivariate data
  - Dimension reduction
  - Simplify analysis and visualization

- **GOAL** estimation of latent independent sources $\mathbf{s}$ from observations $\mathbf{y}$
  - A novel statistical framework for ICA
  - Minimal prior assumptions on observations $\mathbf{y}$
  - A distribution-free test for the existence of independent components
Residual Distribution: Univariate and Bivariate
Independent Component Analysis

For iid vector observations $y_t$, assume independent components (ICs) $s_t$ exist, such that

$$y_t = Ms_t$$

- $M$ denotes the mixing matrix
- Validity of assumption is tested

For simplicity

- $U$, an uncorrelating matrix
- $z_t = Uy_t$, uncorrelated observations

Then

$$s_t = M^{-1}y_t = M^{-1}U^{-1}z_t \equiv Wz_t$$

in which $W = M^{-1}U^{-1}$ is referred to as the separating matrix
Assumptions

- $\mathbf{y}_t = (y_{1t}, \ldots, y_{dt})'$ a $d$-dimensional random vector
- $\mathbf{y}_t$ has continuous distribution function
  - $\mathbf{y}_t \overset{iid}{\sim} F_y$
  - $\mathbb{E}||\mathbf{y}_t||^2 < \infty$
  - $\mathbb{E}(\mathbf{y}_t) = \mathbf{0}$
- $\mathbf{s}_t = (s_{1t}, \ldots, s_{dt})'$ a random vector of ICs
  - $\mathbb{E}\{s_{it}\} = 0$ and $\text{Var}\{s_{it}\} = 1$, $\forall i$
- Separating matrix $\mathbf{W}$ is orthogonal
  - $\mathbf{I} = \text{Cov}(\mathbf{s}_t) = \mathbf{W}\text{Cov}(\mathbf{z}_t)\mathbf{W}' = \mathbf{WW}'$
  - Parameterized by $p = d(d - 1)/2$ vector $\theta$ of rotation angles, $\mathbf{W}_\theta$
Testing for the Existence of Independent Components

- Necessary & sufficient condition for mutually ICs is that

\[ \phi_s = \phi_{s_1} \cdots \phi_{s_d} \]

- Equivalent joint hypotheses WRT transformed variables \( u_i \)

\[ H_0 : \quad \phi_{u_i, u_{i+}} = \phi_{u_i} \phi_{u_{i+}} \quad \text{for all } i = 1, \ldots, d - 1 \]

\[ H_A : \quad \phi_{u_i, u_{i+}} \neq \phi_{u_i} \phi_{u_{i+}} \quad \text{for some } i = 1, \ldots, d - 1 \]

- We define our test statistic as

\[ \mathcal{U}_n(S) = n \sum_{i=1}^{d-1} \mathcal{V}_n^2(\hat{U}_i, \hat{U}_{i+}) \]

- Under \( H_0 \), asymptotic distribution only depends on the dimension \( d \)
PITdCovICA Estimator

- Let $s_t(\theta) = W_{\theta} z_t$ and $\hat{u}_{k,t}(\theta) = \tilde{F}_{sk(\theta)}(s_{k,t}(\theta))$

- $i^+ = \{j : i < j \leq d\}$ \(\mathcal{V}^2(u_i, u_{i^+})\) dependence $s_i$ vs $\{s_{i+1}, \ldots, s_d\}$

- Components of $s$ are mutually independent iff \(\mathcal{V}(u_i, u_{i^+}) = 0\) \(\forall i\)

- Objective function

\[
\hat{J}_n(\theta) = \sum_{i=1}^{d-1} \mathcal{V}^2_n(\hat{U}_i(\theta), \hat{U}_{i^+}(\theta))
\]

- Estimator: $\hat{\theta}_n = \arg\min_{\theta} \hat{J}_n(\theta)$

- Theorem: \(\tilde{\theta}_n \xrightarrow{a.s.} \theta^0\)

- Can be decomposed into \((d - 1)\) conditionally indep. optimizations
PCA vs. ICA: Filtered Unemployment Rate $\hat{e}_t$

Test statistic and approximate $p$-value for joint test of ICs

<table>
<thead>
<tr>
<th>$\mathcal{U}_n(6)$</th>
<th>$y_t$</th>
<th>$\hat{e}_t$</th>
<th>$\hat{z}_t$</th>
<th>$\hat{s}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Statistic</td>
<td>41.48</td>
<td>6.01</td>
<td>1.396</td>
<td>0.635</td>
</tr>
<tr>
<td>Approx. $p$-value</td>
<td>0</td>
<td>0</td>
<td>0.008</td>
<td>0.900</td>
</tr>
</tbody>
</table>
PCA vs. ICA

**PCA**

- $\hat{Z}_2$ vs. $\hat{Z}_1$
- $F(\hat{Z}_2)$ vs. $F(\hat{Z}_1)$

**ICA**

- $\hat{S}_2$ vs. $\hat{S}_1$
- $F(\hat{S}_2)$ vs. $F(\hat{S}_1)$
Interpretation of ICs

\[ CA : e_{1t} = -0.65s_{1t} - 0.26s_{3t} + 0.66s_{5t} \]
\[ FL : e_{2t} = -0.26s_{1t} + 0.94s_{2t} \]
\[ IL : e_{3t} = -0.48s_{1t} + 0.83s_{4t} \]
\[ MI : e_{4t} = -0.86s_{1t} - 0.25s_{4t} - 0.38s_{5t} \]
\[ OH : e_{5t} = -0.46s_{1t} - 0.83s_{6t} \]
\[ WI : e_{6t} = -0.13s_{1t} - 0.98s_{3t} \]

- $s_{1t}$ is related to each state
- $s_{1t}$ has positive relationship with seasonally adjusted GDP
- Supports hypothesis $-s_{1t}$ is national component of unemployment rate
- $s_{2t}$, $s_{3t}$ and $s_{6t}$ are specific components for FL, WI, and OH, resp.
Interest Rate Yields


David S. Matteson (dm484@cornell.edu)  Dependence within Financial Markets  2011 April 30  22 / 26
Yield Curve

Historical Treasury Yield Curves (1982-present)

Historical Treasury Yield Curves (1982-present)
Volatility: $\hat{\Sigma}_t = \hat{M}\hat{\text{Cov}}\{\hat{s}_t|\mathcal{F}_{t-1}\}\hat{M}' = \hat{M}\text{diag}\{\hat{\sigma}_{it}^2\}\hat{M}'$
Conclusions

- Joint test for multivariate serial dependence
- New approach for independent component analysis
- We combine nonparametric probability integral transformation with a generalized nonparametric whitening method
- Limiting properties of the proposed estimator under weak conditions
- A test statistic for checking the existence of independent components

Future Work

- Generalize to dependent data
- Extend to high dimensional data
- Derive asymptotic critical values for general test statistics
- Explore new applications: high-frequency time series, asset allocation


