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/

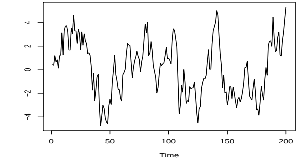
Joint work with: Ruey S. Tsay, Booth School of Business

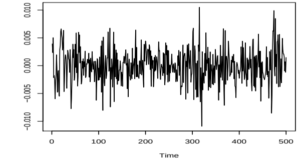
Acknowledgments: National Science Foundation

2011 April 30

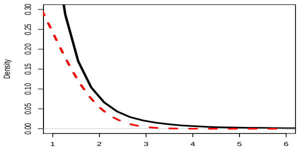
Analysis of Financial and Econometric Data

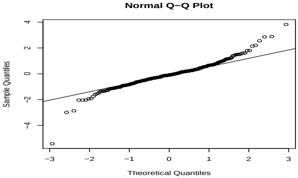
- Serial dependence



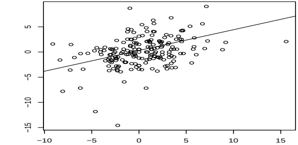


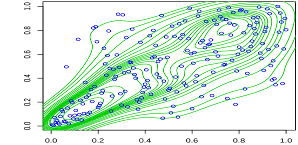
- Non-normality





- Cross dependence





Univariate Serial Dependence

- Serial correlation

  $\mathrm{cor}(x_t, x_{t-\ell})$

- Nonlinear correlation

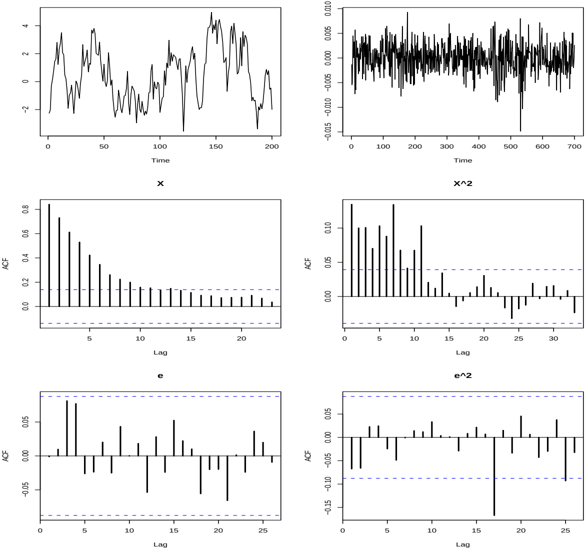
  $\mathrm{cor}(h(x_t), h(x_{t-\ell}))$

- Arbitrary dependence

  x_t vs. $x_{t-\ell}$

- Simultaneous measure

  x_t vs. $\{x_{t-1}, \ldots x_{t-k}\}$



Distance Covariance

Distance covariance $\mathcal{V}(\mathbf{x}_1, \mathbf{x}_2)$ measures dependence between r.v. $\underline{\mathbf{x}_1 \in \mathbb{R}^{q_1}}$ and $\underline{\mathbf{x}_2 \in \mathbb{R}^{q_2}}$, for all distributions with finite first moments

$$\mathcal{V}^2(\mathbf{x}_1, \mathbf{x}_2) = ||\phi_{x_1,x_2}(\mathbf{t}_1, \mathbf{t}_2) - \phi_{x_1}(\mathbf{t}_1)\phi_{x_2}(\mathbf{t}_2)||_\omega^2$$

- ϕ_{x_1} and ϕ_{x_2} denote the characteristic functions of $\mathbf{x}_1$ and $\mathbf{x}_2$, resp.
- ϕ_{x_1,x_2} denotes the joint characteristic function
- $\omega(\mathbf{t}_1, \mathbf{t}_2)$ is a positive weight function
- $\boxed{\mathcal{V}(\mathbf{x}_1, \mathbf{x}_2) = 0 \text{ if and only if } \mathbf{x}_1 \text{ and } \mathbf{x}_2 \text{ are independent}}$

Sample Distance Covariance

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

$$\begin{aligned}\mathcal{V}_n^2\left(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}\right) &= ||\phi^n_{x_1,x_2}(\mathbf{t}_1, \mathbf{t}_2) - \phi^n_{x_1}(\mathbf{t}_1)\phi^n_{x_2}(\mathbf{t}_2)||^2_\omega \\ &= \frac{1}{n^2}\sum_{k,l=1}^{n} A_{kl}B_{kl}\end{aligned}$$

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

$$A_{kl} = a_{kl} - \bar{a}_{k.} - \bar{a}_{.l} + \bar{a}_{..}, \quad B_{kl} = b_{kl} - \bar{b}_{k.} - \bar{b}_{.l} + \bar{b}_{..},$$

(Székely et al., 2007)

- $\lim_{n\to\infty} \mathcal{V}_n \overset{a.s.}{=} \mathcal{V}$, and $n\mathcal{V}_n^2$ convergences in distribution to a r.v.

An Alternative Measure

- $\mathcal{V}(\mathbf{x}_1, \mathbf{x}_2)$ depends on marginal distributions
- Apply probability integral transformation (PIT)

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

- $\boxed{\mathcal{V}(\mathbf{u}_1, \mathbf{u}_2) = 0 \text{ iff } \mathbf{x}_1 \text{ and } \mathbf{x}_2 \text{ are independent}}$
- The F_X are unknown
  - Use $\widehat{F}_X$, marginal ranks
  - Let $\hat{u} = \widehat{F}_X(x)$

Lemma

$$\mathcal{V}_n(\widehat{\mathbf{U}}_1, \widehat{\mathbf{U}}_2) \xrightarrow{a.s.} \mathcal{V}(\mathbf{u}_1, \mathbf{u}_2), \quad \textit{and} \quad n\mathcal{V}_n^2(\widehat{\mathbf{U}}_1, \widehat{\mathbf{U}}_2) \xrightarrow{\mathcal{D}} \textit{r.v., as } n \to \infty$$

A Joint Test for Serial Dependence

- $\mathcal{V}(y_t, y_{t-k})$ measures lag-k serial dependence
- Let $\mathbf{y}_{t-k} = \{y_{t-1}, \ldots, y_{t-k}\}$
- $\mathcal{V}(y_t, \mathbf{y}_{t-k})$ jointly measures serial dependence up to lag-k
- Assuming stationarity, we use $\mathcal{V}(u_t, \mathbf{u}_{t-k})$
- Joint hypothesis for serial dependence

$$H_0: \quad \phi_{u_t, \mathbf{u}_{t-\ell}} = \phi_{u_t} \phi_{\mathbf{u}_{t-\ell}} \text{ for all } \ell = 1, \ldots, k$$

- We define our test statistic as

$$\boxed{\mathcal{Q}_1(\mathbf{Y}, k) = n\mathcal{V}_n^2(\widehat{U}_t, \widehat{\mathbf{U}}_{t-k})}$$

Asymptotic Distribution

- Under H_0,

$$\mathcal{Q}_1(\mathbf{Y}, k) \xrightarrow{D} ||\zeta_k(a,b)||_\omega^2$$

$\zeta_k(\cdot,\cdot)$ mean zero <u>complex Gaussian process</u> with covariance function

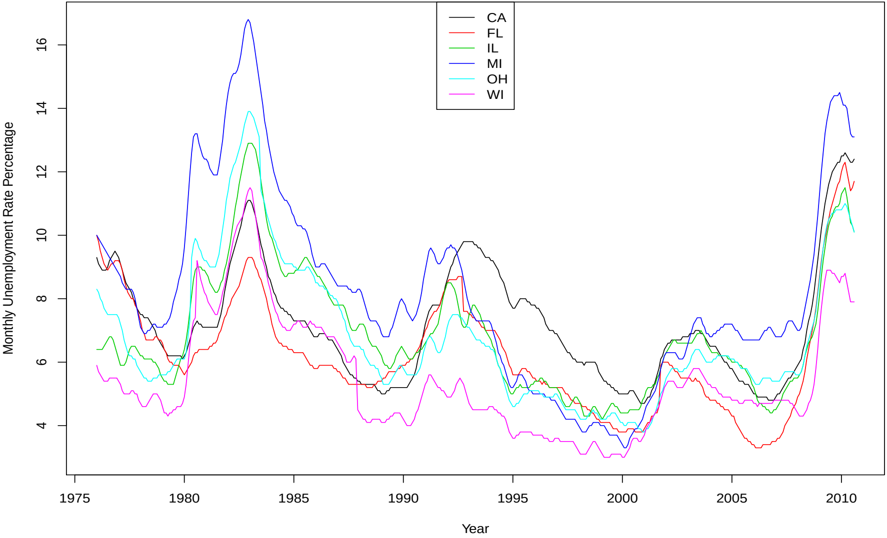
$$R_k(c, c_0) = \left(\phi_u(a - a_0) - \phi_u(a)\overline{\phi_u(a_0)}\right)\left(\phi_{u_k}(b - b_0) - \phi_{u_k}(b)\overline{\phi_{u_k}(b_0)}\right)$$

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

- ϕ_u characteristic function of Uniform(0,1) r.v., and $\phi_{u_k} = \phi_u \cdots \phi_u$
- Distribution Free Test
- Analogous Multivariate Test

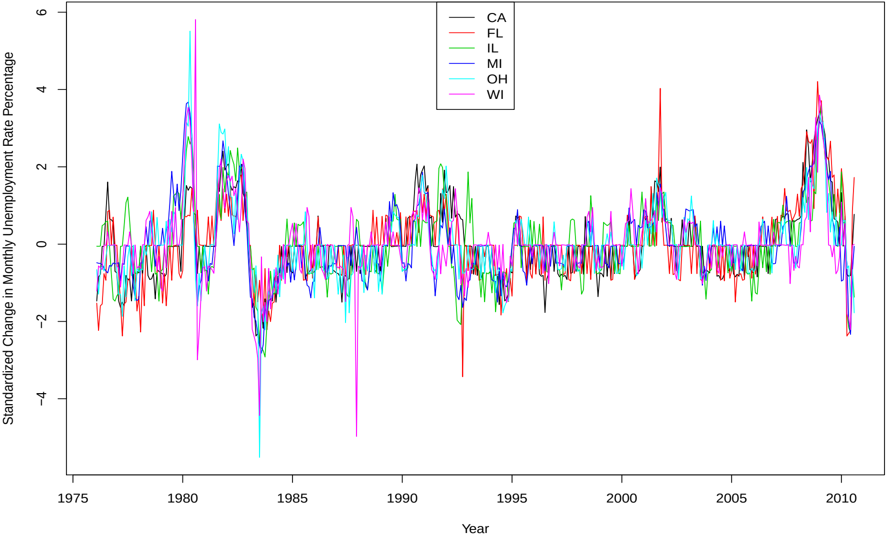
Seasonally Adjusted Monthly Unemployment Rates (%)

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



Testing for Serial Dependence

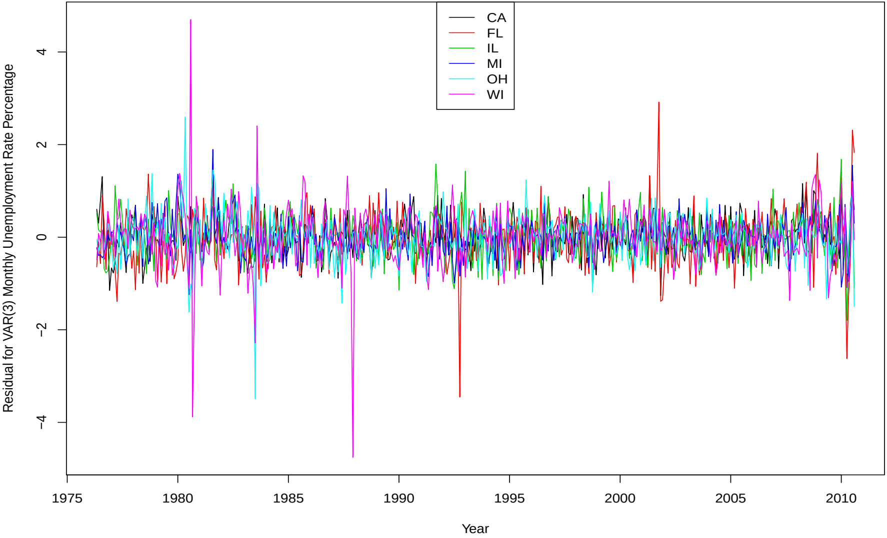
- Transform to stationary $\mathbf{y}_t$
  - First difference series, scaled by monthly standard deviations
  - $\mathcal{Q}_6(\mathbf{Y}, 12) = 121.67$ with p-value ≈ 0
- Fit a vector autoregression of order three, via OLS

$$\mathbf{y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{y}_{t-1} + \boldsymbol{\beta}_2 \mathbf{y}_{t-2} + \boldsymbol{\beta}_3 \mathbf{y}_{t-3} + \mathbf{e}_t$$

  - Calculate residuals $\hat{\mathbf{e}}_t$
  - $\mathcal{Q}_6(\widehat{\mathbf{E}}, 12) = 11.31$ with p-value ≈ 0.31
  - $\Rightarrow$ Linear model is sufficient

Residual Series, Monthly Unemployment Rate %

No significant serial dependence



Financial & Economic Processes are Inherently Multivariate

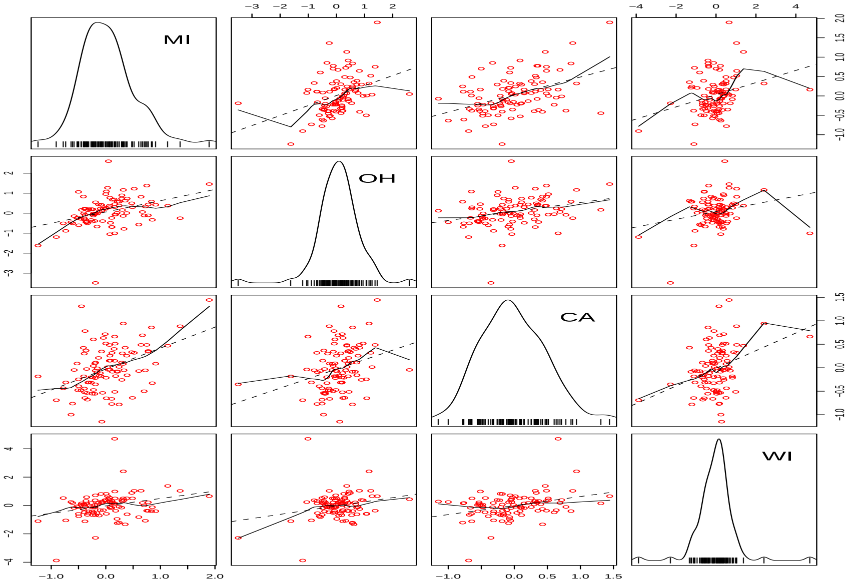
- **Simultaneous Analysis is Crucial**

- **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 $\mathbf{y}_t$, assume independent components (ICs) $\mathbf{s}_t$ exist, such that

$$\boxed{\mathbf{y}_t = \mathbf{M}\mathbf{s}_t}$$

- $\mathbf{M}$ denotes the mixing matrix
- Validity of assumption is tested

For simplicity

- $\mathbf{U}$, an uncorrelating matrix
- $\mathbf{z}_t = \mathbf{U}\mathbf{y}_t$, uncorrelated observations

Then

$$\boxed{\mathbf{s}_t = \mathbf{M}^{-1}\mathbf{y}_t = \mathbf{M}^{-1}\mathbf{U}^{-1}\mathbf{z}_t \equiv \mathbf{W}\mathbf{z}_t}$$

in which $\mathbf{W} = \mathbf{M}^{-1}\mathbf{U}^{-1}$ 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
  - $\mathbf{y}_t \overset{iid}{\sim} F_y$
  - $\mathrm{E}||\mathbf{y}_t||^2 < \infty$
  - $E(\mathbf{y}_t) = \mathbf{0}$
- $\mathbf{s}_t = (s_{1t}, \ldots, s_{dt})'$ a random vector of ICs
  - $\mathrm{E}\{s_{it}\} = 0$ and $\mathrm{Var}\{s_{it}\} = 1, \forall i$
- Separating matrix $\mathbf{W}$ is orthogonal
  - $\mathbf{I} = \mathrm{Cov}(\mathbf{s}_t) = \mathbf{W}\mathrm{Cov}(\mathbf{z}_t)\mathbf{W}' = \mathbf{W}\mathbf{W}'$
  - Parameterized by $p = d(d-1)/2$ vector $\boldsymbol{\theta}$ of rotation angles, $\mathbf{W}_{\boldsymbol{\theta}}$

Testing for the Existence of Independent Components

- Necessary & sufficient condition for mutually ICs is that

$$\phi_{\mathbf{s}} = \phi_{s_1} \cdots \phi_{s_d}$$

- Equivalent joint hypotheses WRT transformed variables u_i

$$\begin{aligned} H_0 : &\quad \phi_{u_i, u_{i+}} = \phi_{u_i}\phi_{u_{i+}} \text{ for all } i = 1, \ldots, d-1 \\ H_A : &\quad \phi_{u_i, u_{i+}} \neq \phi_{u_i}\phi_{u_{i+}} \text{ for some } i = 1, \ldots, d-1 \end{aligned}$$

- We define our test statistic as

$$\mathcal{U}_n(\mathbf{S}) = n \sum_{i=1}^{d-1} \mathcal{V}_n^2(\widehat{\mathbf{U}}_i, \widehat{\mathbf{U}}_{i^+})$$

- Under H_0, asymptotic distribution only depends on the dimension d

PITdCovICA Estimator

- Let $\mathbf{s}_t(\boldsymbol{\theta}) = \mathbf{W}_{\boldsymbol{\theta}}\mathbf{z}_t$ and $\hat{u}_{k,t}(\boldsymbol{\theta}) = \widetilde{F}_{s_k(\theta)}\left(s_{k,t}(\boldsymbol{\theta})\right)$
- $\boxed{i^+ = \{j : i < j \leq d\}}$ $\mathcal{V}^2(u_i, \mathbf{u}_{i^+})$ dependence s_i vs$\{s_{i+1}, \ldots, s_d\}$
- Components of $\mathbf{s}$ are mutually independent iff $\mathcal{V}(u_i, \mathbf{u}_{i^+}) = 0 \ \forall i$
- Objective function

$$\widehat{\mathcal{J}}_n(\boldsymbol{\theta}) = \sum_{i=1}^{d-1} \mathcal{V}_n^2(\widehat{\mathbf{U}}_i(\boldsymbol{\theta}), \widehat{\mathbf{U}}_{i^+}(\boldsymbol{\theta}))$$

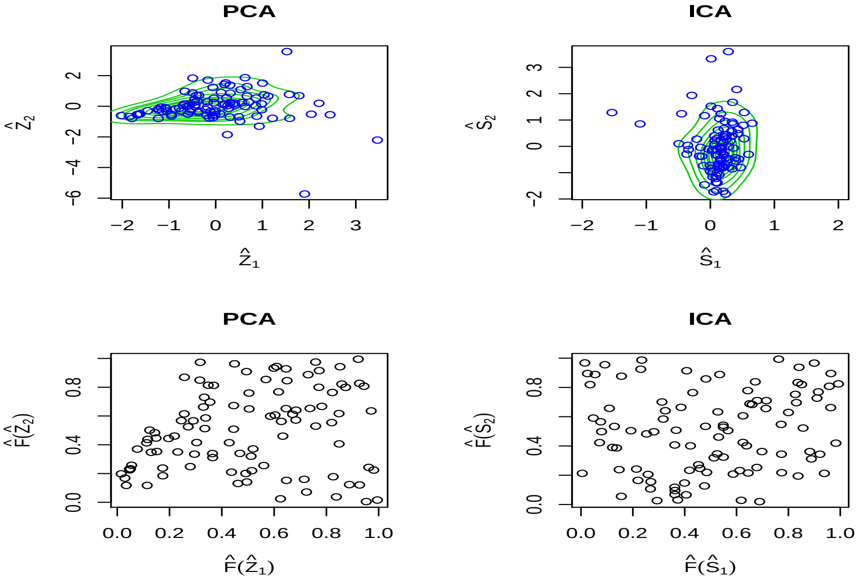
- Estimator: $\boxed{\widehat{\boldsymbol{\theta}}_n = \text{argmin}_{\boldsymbol{\theta}} \widehat{\mathcal{J}}_n(\boldsymbol{\theta})}$ Theorem: $\boxed{\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\theta}^0}$
- Can be decomposed into $(d-1)$ conditionally indep. optimizations

PCA vs. ICA: Filtered Unemployment Rate $\hat{\mathbf{e}}_t$

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

$\mathcal{U}_n(6)$	$\mathbf{y}_t$	$\hat{\mathbf{e}}_t$	$\hat{\mathbf{z}}_t$	$\hat{\mathbf{s}}_t$
Test Statistic	41.48	6.01	1.396	0.635
Approx. *p*-value	0	0	0.008	0.900

PCA vs. ICA

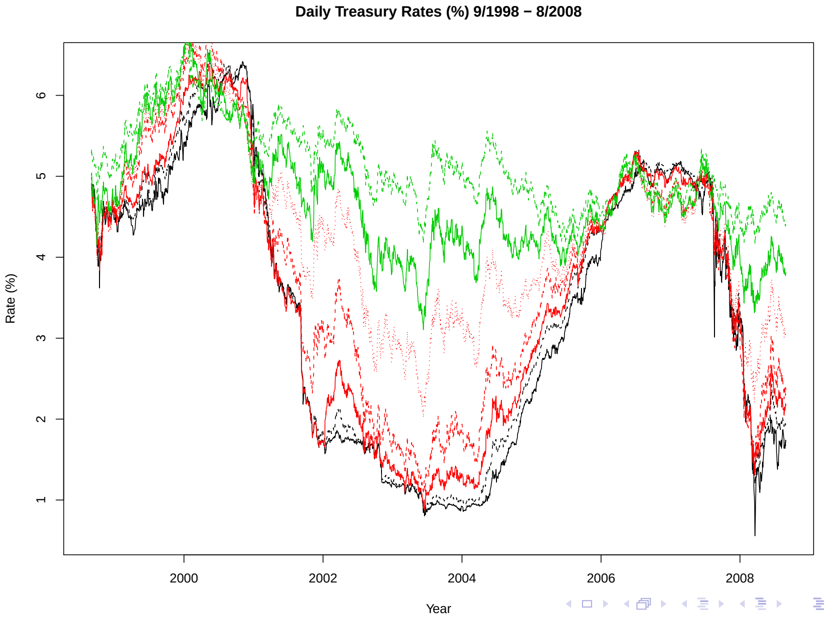

PCA
ICA
$\hat{Z}_2$
$\hat{Z}_1$
$\hat{S}_2$
$\hat{S}_1$
−6 −4 −2 0 2
−2 −1 0 1 2 3
−2 0 1 2 3
−2 −1 0 1 2
$\hat{F}(\hat{Z}_2)$
$\hat{F}(\hat{Z}_1)$
$\hat{F}(\hat{S}_2)$
$\hat{F}(\hat{S}_1)$
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8 1.0

Interpretation of ICs

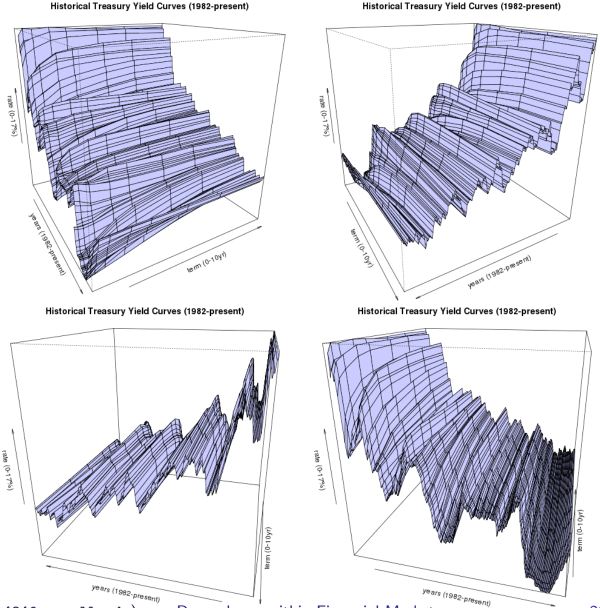
$$\begin{aligned}
CA: e_{1t} &= -0.65 s_{1t} - 0.26 s_{3t} + 0.66 s_{5t} \\
FL: e_{2t} &= -0.26 s_{1t} + 0.94 s_{2t} \\
IL: e_{3t} &= -0.48 s_{1t} + 0.83 s_{4t} \\
MI: e_{4t} &= -0.86 s_{1t} - 0.25 s_{4t} - 0.38 s_{5t} \\
OH: e_{5t} &= -0.46 s_{1t} - 0.83 s_{6t} \\
WI: e_{6t} &= -0.13 s_{1t} - 0.98 s_{3t}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

- 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.

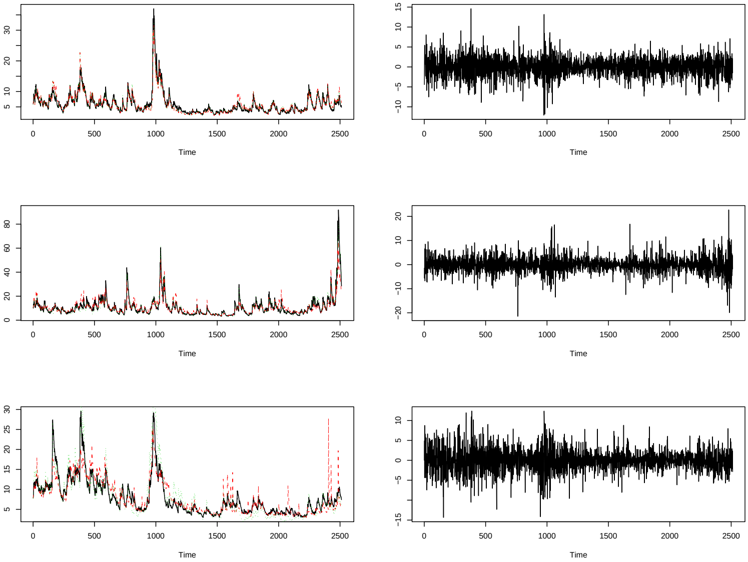
Interest Rate Yields



Yield Curve


Historical Treasury Yield Curves (1982-present)
rate (0-17%)
years (1982-present)
term (0-10yr)
Historical Treasury Yield Curves (1982-present)
rate (0-17%)
term (0-10yr)
years (1982-present)
Historical Treasury Yield Curves (1982-present)
rate (0-17%)
years (1982-present)
term (0-10yr)
Historical Treasury Yield Curves (1982-present)
rate (0-17%)
years (1982-present)
term (0-10yr)

Volatility: $\widehat{\Sigma}_t = \widehat{\mathbf{M}}\widehat{\mathrm{Cov}}\{\widehat{\mathbf{s}}_t | \mathcal{F}_{t-1}\}\widehat{\mathbf{M}}' = \widehat{\mathbf{M}}\mathrm{diag}\{\widehat{\sigma}^2_{it}\}\widehat{\mathbf{M}}'$



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*

Bibliography

Matteson, D. S., and Tsay, R. S. (2010), "Dynamic Orthogonal Components for Multivariate Time Series," *Invited Revision, JASA*, .

Matteson, D. S., and Tsay, R. S. (2011), "Independent Component Analysis via Distance Covariance," *Submitted*, .

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rizzo, M. L., and Szekely, G. J. (2008), *energy: E-statistics (Energy Statistics)*. R Package Version 1.1-0.

Székely, G. J., and Rizzo, M. L. (2009), "Brownian Distance Covariance," *Annals of Applied Statistics*, 3(4), 1236–1265.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances," *Annals of Statistics*, 35(6), 2769.