

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/>

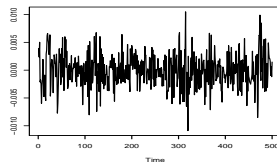
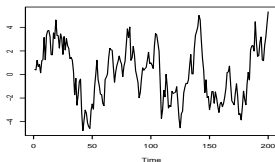
Joint work with: Ruey S. Tsay, Booth School of Business

Acknowledgments: National Science Foundation

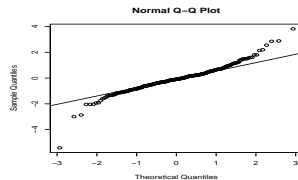
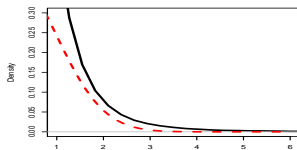
2011 April 30

Analysis of Financial and Econometric Data

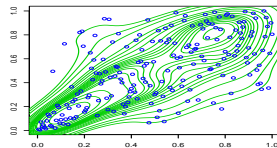
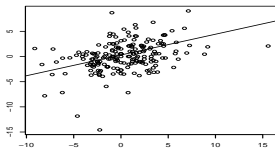
▶ Serial dependence



▶ Non-normality



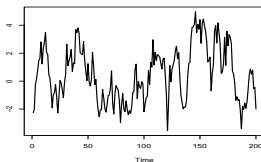
▶ Cross dependence



Univariate Serial Dependence

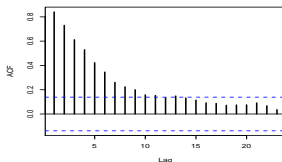
- Serial correlation

$$\text{cor}(x_t, x_{t-l})$$



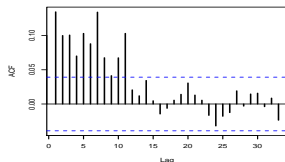
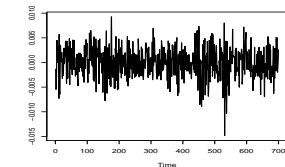
- Nonlinear correlation

$$\text{cor}(h(x_t), h(x_{t-l}))$$



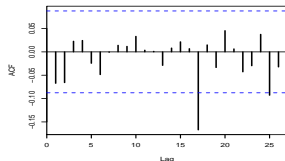
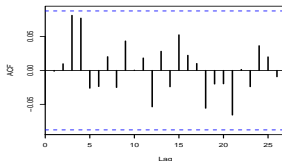
- Arbitrary dependence

$$x_t \text{ VS. } x_{t-l}$$



- Simultaneous measure

$$x_t \text{ VS. } \{x_{t-1}, \dots, x_{t-k}\}$$



Distance Covariance

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

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

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

Sample Distance Covariance

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

$$\begin{aligned} \mathcal{V}_n^2(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) &= \|\phi_{X_1, X_2}^n(\mathbf{t}_1, \mathbf{t}_2) - \phi_{X_1}^n(\mathbf{t}_1)\phi_{X_2}^n(\mathbf{t}_2)\|_\omega^2 \\ &= \frac{1}{n^2} \sum_{k, l=1}^n A_{kl} B_{kl} \end{aligned}$$

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

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

(Székely et al., 2007)

► $\lim_{n \rightarrow \infty} \mathcal{V}_n \stackrel{a.s.}{=} \mathcal{V}$, 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)

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

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

Lemma

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

A Joint Test for Serial Dependence

- ▶ $\mathcal{V}(y_t, y_{t-k})$ measures **lag- k** serial dependence
- ▶ Let $\mathbf{y}_{t-k} = \{y_{t-1}, \dots, y_{t-k}\}$
- ▶ $\mathcal{V}(y_t, \mathbf{y}_{t-k})$ **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

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

- ▶ We define our **test statistic** as

$$Q_1(\mathbf{Y}, k) = n\mathcal{V}_n^2(\hat{U}_t, \hat{\mathbf{U}}_{t-k})$$

Asymptotic Distribution

- ▶ Under H_0 ,

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

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

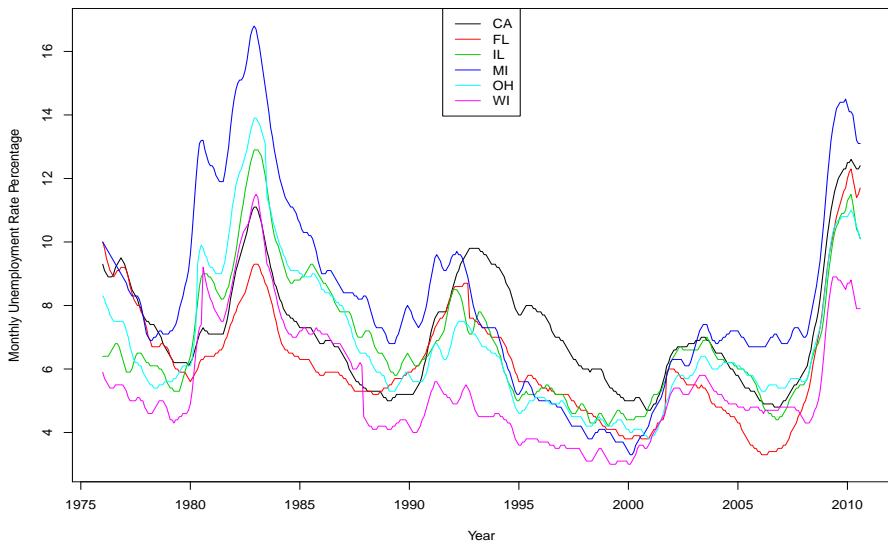
$$R_k(c, c_0) = \left(\phi_u(a - a_0) - \phi_u(a)\overline{\phi_u(a_0)} \right) \left(\phi_{u_k}(b - b_0) - \phi_{u_k}(b)\overline{\phi_{u_k}(b_0)} \right)$$

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

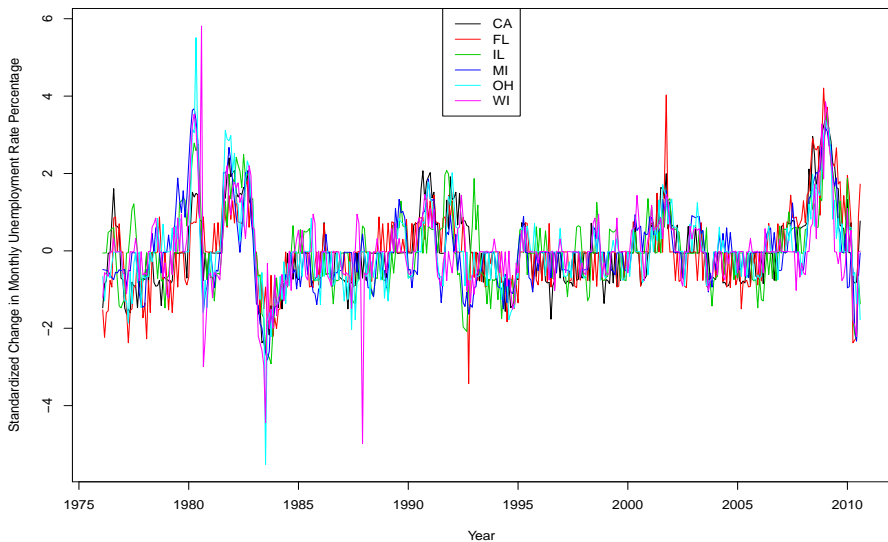
Seasonally Adjusted Monthly Unemployment Rates (%)

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
 - ▶ $Q_6(\mathbf{Y}, 12) = 121.67$ with p -value ≈ 0

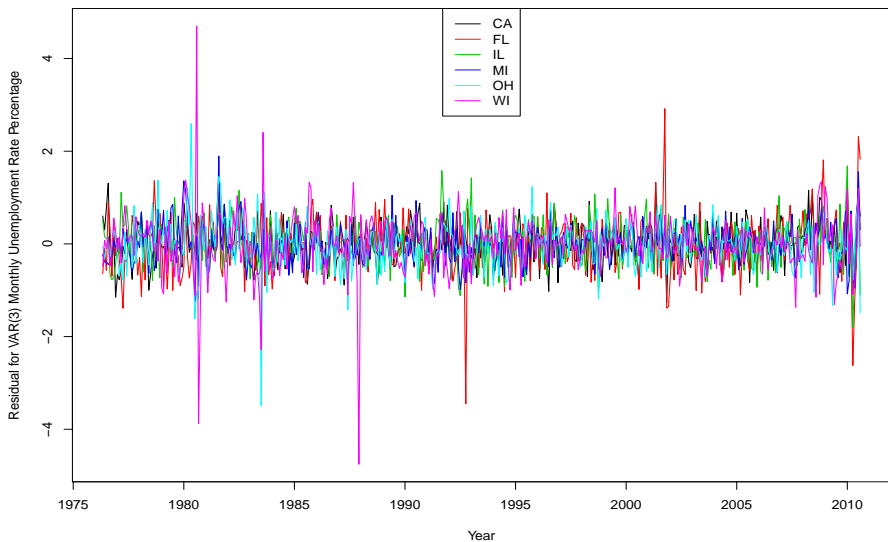
- ▶ Fit a vector autoregression of order three, via OLS

$$\mathbf{y}_t = \beta_0 + \beta_1 \mathbf{y}_{t-1} + \beta_2 \mathbf{y}_{t-2} + \beta_3 \mathbf{y}_{t-3} + \mathbf{e}_t$$

- ▶ Calculate residuals $\hat{\mathbf{e}}_t$
- ▶ $Q_6(\hat{\mathbf{E}}, 12) = 11.31$ with p -value ≈ 0.31
- ▶ \Rightarrow **Linear model is sufficient**

Residual Series, Monthly Unemployment Rate %

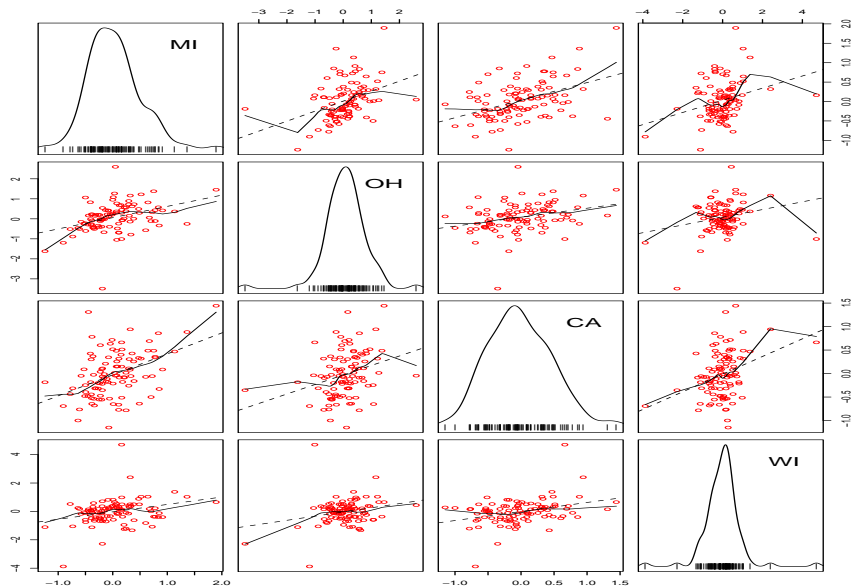
No significant serial dependence



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

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

- ▶ \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

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

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

Assumptions

- ▶ $\mathbf{y}_t = (y_{1t}, \dots, y_{dt})'$ a **d -dimensional** random vector
- ▶ \mathbf{y}_t has **continuous distribution** function
 - ▶ $\mathbf{y}_t \stackrel{iid}{\sim} F_y$
 - ▶ $E\|\mathbf{y}_t\|^2 < \infty$
 - ▶ $E(\mathbf{y}_t) = \mathbf{0}$
- ▶ $\mathbf{s}_t = (s_{1t}, \dots, s_{dt})'$ a random vector of **ICs**
 - ▶ $E\{s_{it}\} = 0$ and $\text{Var}\{s_{it}\} = 1, \forall i$
- ▶ **Separating matrix \mathbf{W}** is **orthogonal**
 - ▶ $\mathbf{I} = \text{Cov}(\mathbf{s}_t) = \mathbf{W}\text{Cov}(\mathbf{z}_t)\mathbf{W}' = \mathbf{W}\mathbf{W}'$
 - ▶ Parameterized by $p = d(d - 1)/2$ vector $\boldsymbol{\theta}$ of rotation angles, \mathbf{W}_θ

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

$$H_0 : \phi_{u_i, u_{i+}} = \phi_{u_i} \phi_{u_{i+}} \text{ for all } i = 1, \dots, d-1$$

$$H_A : \phi_{u_i, u_{i+}} \neq \phi_{u_i} \phi_{u_{i+}} \text{ for some } i = 1, \dots, d-1$$

- ▶ We define our **test statistic** as

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

- ▶ 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 $\hat{u}_{k,t}(\boldsymbol{\theta}) = \tilde{F}_{s_k(\boldsymbol{\theta})}(s_{k,t}(\boldsymbol{\theta}))$
- ▶ $i^+ = \{j : i < j \leq d\}$ $\mathcal{V}^2(u_i, \mathbf{u}_{i^+})$ dependence s_i vs $\{s_{i+1}, \dots, s_d\}$
- ▶ Components of \mathbf{s} are mutually independent iff $\mathcal{V}(u_i, \mathbf{u}_{i^+}) = 0 \forall i$
- ▶ Objective function

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

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

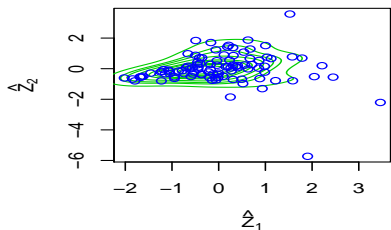
PCA vs. ICA: Filtered Unemployment Rate $\hat{\mathbf{e}}_t$

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

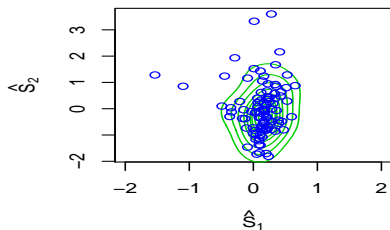
$\mathcal{U}_n(6)$	\mathbf{y}_t	$\hat{\mathbf{e}}_t$	$\hat{\mathbf{z}}_t$	$\hat{\mathbf{s}}_t$
Test Statistic	41.48	6.01	1.396	0.635
Approx. p -value	0	0	0.008	0.900

PCA vs. ICA

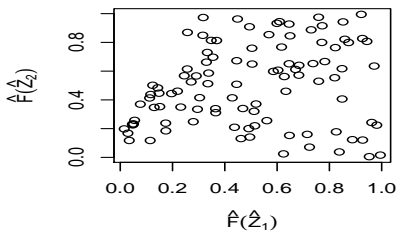
PCA



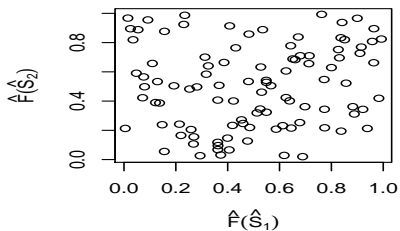
ICA



PCA



ICA



Interpretation of ICs

$$CA : e_{1t} = -0.65s_{1t} - 0.26s_{3t} + 0.66s_{5t}$$

$$FL : e_{2t} = -0.26s_{1t} + 0.94s_{2t}$$

$$IL : e_{3t} = -0.48s_{1t} + 0.83s_{4t}$$

$$MI : e_{4t} = -0.86s_{1t} - 0.25s_{4t} - 0.38s_{5t}$$

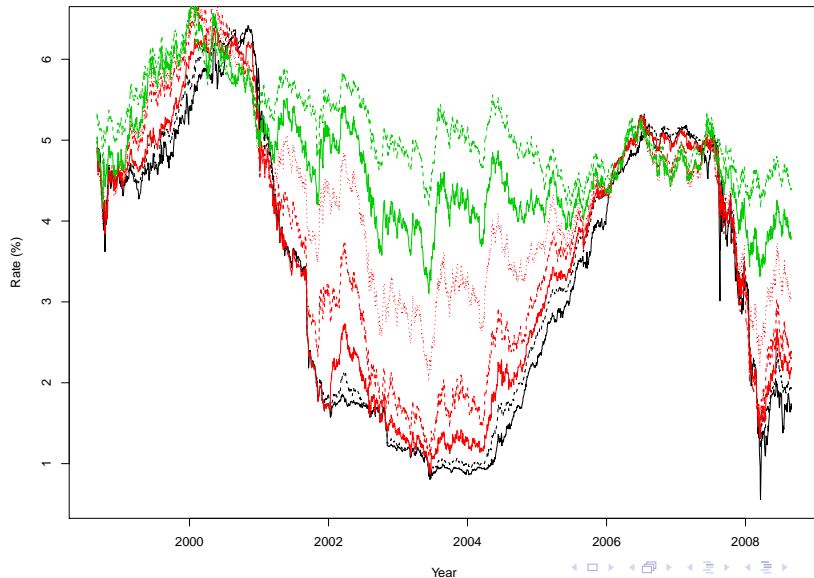
$$OH : e_{5t} = -0.46s_{1t} - 0.83s_{6t}$$

$$WI : e_{6t} = -0.13s_{1t} - 0.98s_{3t}$$

- ▶ 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.

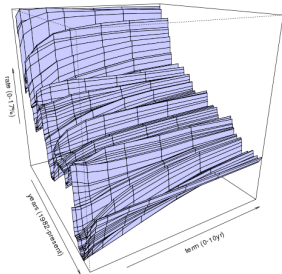
Interest Rate Yields

Daily Treasury Rates (%) 9/1998 – 8/2008

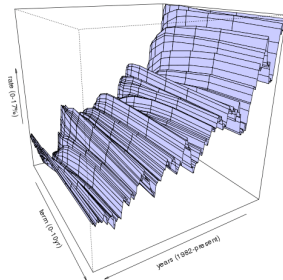


Yield Curve

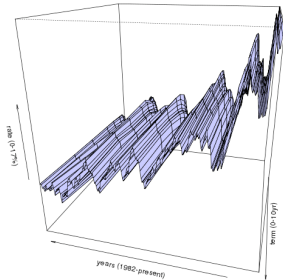
Historical Treasury Yield Curves (1982-present)



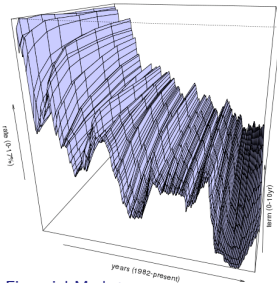
Historical Treasury Yield Curves (1982-present)



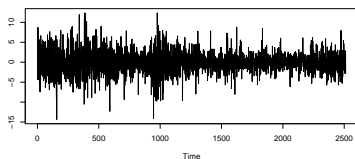
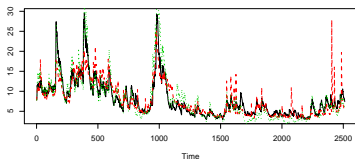
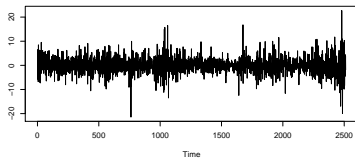
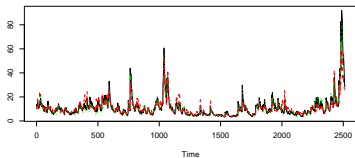
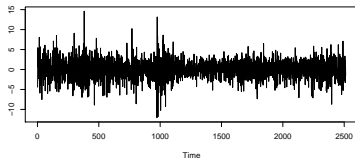
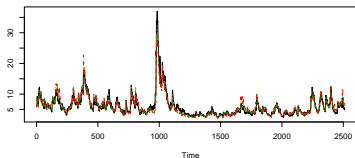
Historical Treasury Yield Curves (1982-present)



Historical Treasury Yield Curves (1982-present)



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



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 Székely, 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.