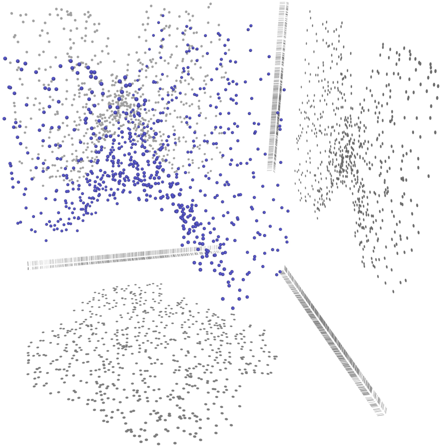
High Dimensional Nonstationary Time Series Modelling With Dynamic Factor Models

Wolfgang Karl Härdle
(P. Majer, Y. Ritov, R. Song)

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
CDA - Centre for Complex Data Analysis
The National Central University
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://www.stat.ncu.edu.tw

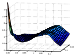
Dynamics of High Dimensional Objects



- Non-stationarity
- Spatial structure
- Many variables

Dynamics of High Dimensional Objects

- Meteorology
  - Temperature and Climate Change
- Medicine
  - Risk Perception
- Finance
  - Implied Volatility Surface
  - Limit Order Book
  - Collateralized Debt Obligation
  - CO_2 Emission Allowance
  - Empirical Pricing Kernel
  - Electricity Forward Prices
  - Yield Curve



Temperature and Climate Change

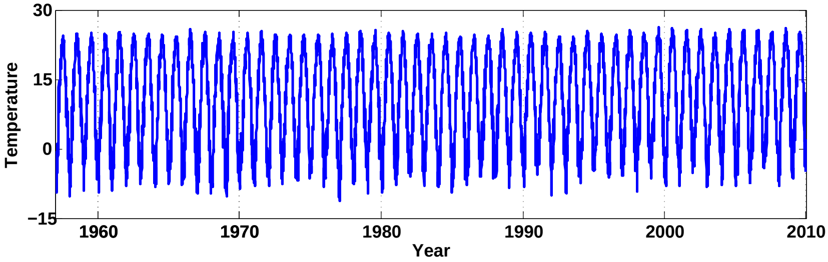


Figure 1: Daily temperature observations averaged over stations ($J = 159$) in China 19570101 - 20091231; source: China Meteorological Administration

Temperature and Climate Change

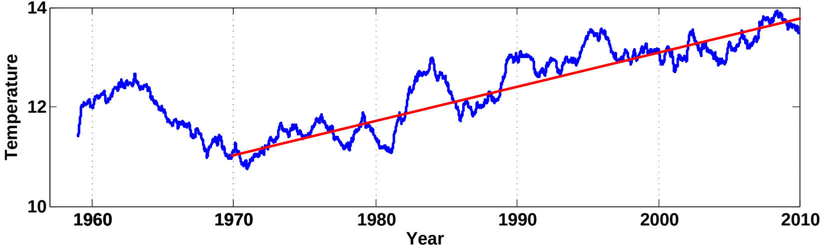


Figure 2: Moving average temperatures (730 days), linear trend (significant estimates at 1% level)

Temperature and Climate Change

- Weather Derivatives

$$
\begin{array}{ccc}
\text{Econometrics} & & \text{Fin. Mathematics} \\
T_t & & CAR(3) \\
\downarrow & & \downarrow \\
X_t = T_t - \Lambda_t & & F_{CAT(t,\tau_1,\tau_2)} \\
\downarrow & & \downarrow \\
X_{t+3} = a^\top X_t + \sigma_t \varepsilon_t & & MPR \\
\downarrow & & \\
\hat{\varepsilon}_t = \frac{\hat{X}_t}{\hat{\sigma}_t} \sim N(0,1) & &
\end{array}
$$

- Detecting complex trends, evaluating "non priced" places

Risk Perception

- functional Magnetic Resonance Imaging

- Measuring Blood Oxygenation Level Dependent (BOLD) effect every 2-3 sec
  High-dimensional, high frequency & large data set

Risk Perception

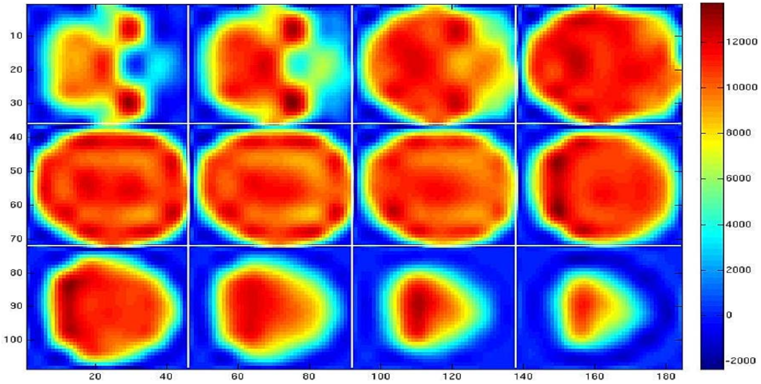


Figure 3: Example of a fMRI image at fixed time point, 12 horizontal slices of the brain's scan. fMRI

Risk Perception

- Which part is activated during *risk related decisions* ?
- Can statistical analysis help to detect this area?
- Response curve (to stimuli)? classify "risky people"?


Implied Volatility Surface

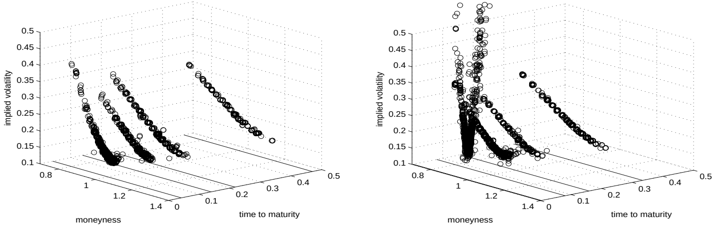


Figure 4: Left panel: observations on 20040701; right panel: observations on 20040819. Bottom solid lines indicate the observed maturities, which move towards the expiry IVS

Limit Order Book

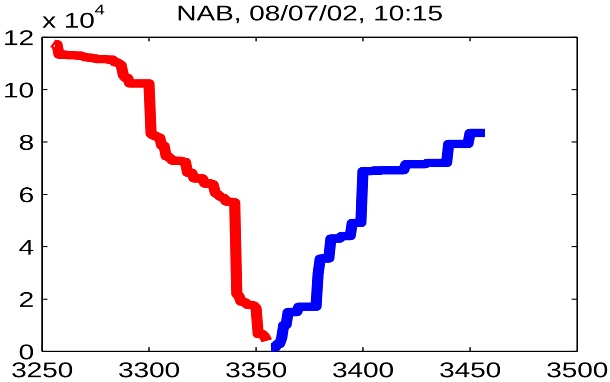


Figure 5: Bid and ask curves constructed from the order book of National Australian Bank stock prices on 20020801. LOB

Collateralized Debt Obligation

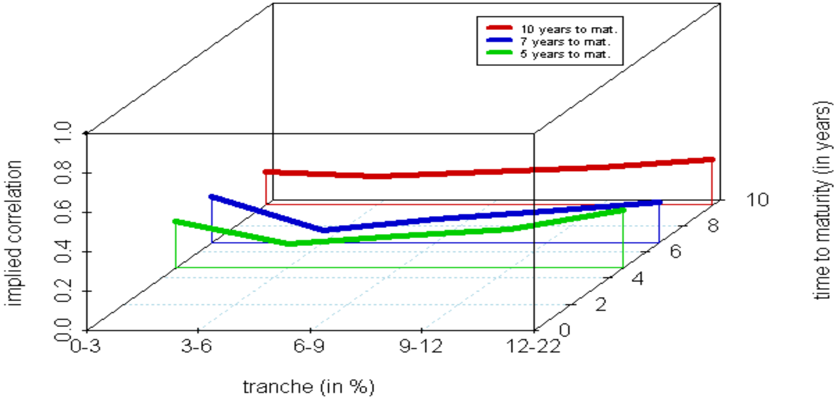


Figure 6: Compound correlations on 20070321 w.r.t. time to maturity (in years), implied correlation and tranche (in %). CDO

CO_2 Emission Allowance

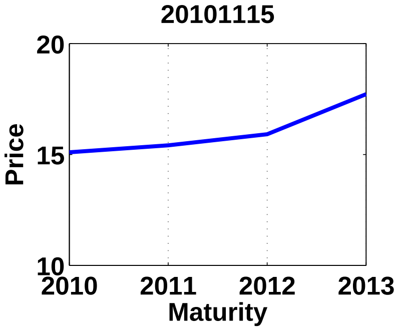
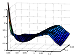


Figure 7: Term structure for CO_2 emission allowance's spot and futures prices (EUR/t), trading on 20101115 in the EEX market. CO2

- CO2 density 1.799 kg/m^3 at 25 C$^\circ$



Empirical Pricing Kernel

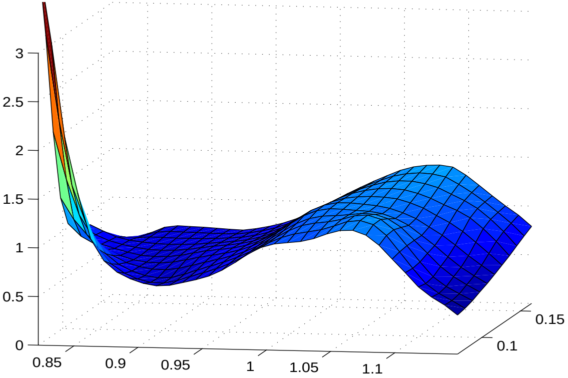


Figure 8: Estimated PK across moneyness κ and maturity τ at t = 20010710. EPK

Electricity Forward Prices

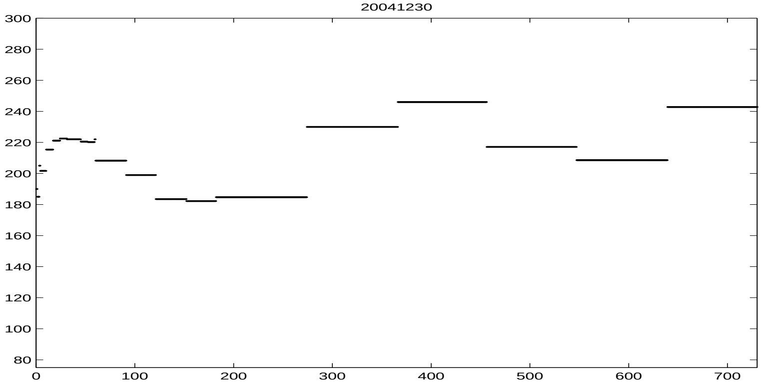


Figure 9: Term structure of the electricity prices (NOK/MWh) from the Nord Pool on 20041230. EFP

Yield Curve

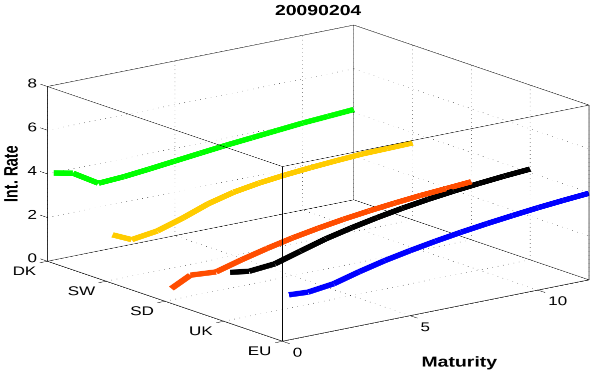


Figure 10: Yield curves of European Zone, United Kingdom, Sweden, Switzerland and Denmark on 20090204 YC

3D Challenge

Dimensionality

Dependency

Divergency

Overview

1. Motivation ✓
2. Introduction
3. Estimation
4. Some theoretical insight
5. Applications

Notation

$$\underbrace{(X_{1,1}, Y_{1,1}), \ldots, (X_{J_1,1}, Y_{J_1,1})}_{t=1}, \ldots, \underbrace{(X_{1,T}, Y_{1,T}), \ldots, (X_{J_T,T}, Y_{J_T,T})}_{t=T},$$

where:

$X_{j,t} \in \mathbb{R}^d, \; Y_{j,t} \in \mathbb{R}$

T - the number of observed time periods (days)

J_t - the number of the observations in (day) t

$\mathsf{E}(Y_t|X_t) = F_t(X_t)$.

What is $F_t(X_t)$? How does it move?

Basic Idea

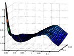
- Use a “time & space” dynamic approach
- Separate time dynamics from space functions
- Low dim time series dynamics
- High dim (time invariant) space functions
- $\#$ of factors $\nearrow J$, fitting $\nearrow$

How to penalise non interesting frequency loadings?

Dynamic Semiparametric Factor Model

$$\mathsf{E}(Y_t|X_t) = \sum_{l=1}^{L} Z_{0,t,l}\, m_l(X_t) = Z_{0,t}^{\top} m(X_t) = Z_{0,t}^{\top} A \psi(X)$$

$Z_t = (Z_{0,t,1}, \ldots, Z_{0,t,L})^{\top}$ low dim (stationary) time series
$m(\cdot)$ tuple of functions $(m_0, m_1, \ldots, m_L)^{\top}$
$\psi(x) = (\psi_1, \ldots, \psi_K)^{\top}(x)$ vector of known basis functions
A: $L \times K$ coefficient matrix



Dynamic Semiparametric Factor Model

$$
\begin{aligned}
Y_{t,j} &= \sum_{l=1}^{L}\sum_{r=1}^{R} u_r(t)\gamma_{rl}\sum_{k=1}^{K} a_{lk}\psi_k(X_{t,j}) + \varepsilon_{tj} \\
Y_t^\top &= \underbrace{U_t^\top \Gamma^*}_{Z_t^\top}\underbrace{A^*\Psi_t}_{m} + \varepsilon_t \stackrel{\mathrm{def}}{=} U_t^\top \beta^{*\top}\Psi_t + \varepsilon_t .
\end{aligned}
$$

$U_t^\top = (u_1(t), \ldots, u_R(t))$, $u_r(t)$ time basis
$\Psi_t = (\psi_1(X_t), \ldots, \psi_K(X_t))^\top$, $\psi_k(x)$ space basis
$\beta^{*\top}$ $R \times K$ matrix, $\|\beta\|_{2,1} = \sum_{r=1}^{R}\sqrt{\sum_{k=1}^{K}\beta_{rk}^2}$ (group Lasso)

Generalized DSFM

$$
\begin{aligned}
Y_t^\top &= (Z_{0,t}^\top + U_t^\top \Gamma) A \Psi_t + \varepsilon_t' = U_t^\top \Gamma A \Psi_t + (Z_{0,t}^\top A \Psi_t + \varepsilon_t') \\
&\stackrel{\text{def}}{=} U_t^\top \Gamma A \Psi_t + \varepsilon_t, \quad \textit{with} \quad \mathsf{E}(Z_{0,t} | X_t) = 0.
\end{aligned}
$$

- Stochastic evolution in time

3D Challenge

- How to fit GDSFM?
- What risk is involved?
- How to select the basis?

Dimensionality

Dependency

Divergency

GDSFM Estimation

1. Find the trend based on $Y_t^\top = U_t^\top \Gamma A \Psi_t + \varepsilon_t$
2. Based on $\widehat{Y}_t^\top \stackrel{\text{def}}{=} Y_t^\top - U_t^\top \widehat{\beta} \Psi_t$, $\widehat{A}$ and Ψ_t, obtain $\widehat{Z}_{0,t}$

Time Basis

- Global trend: 1, t, $(3t^2-1)/2, \ldots$ Legendre Polynomial
- Seasonality: $\sin\{nt/(p2\pi)\}$, $\cos\{nt/(p2\pi)\}$ Fourier Series
  Period $p = 11.8$ (fMRI), $365, 3650$ (weather)

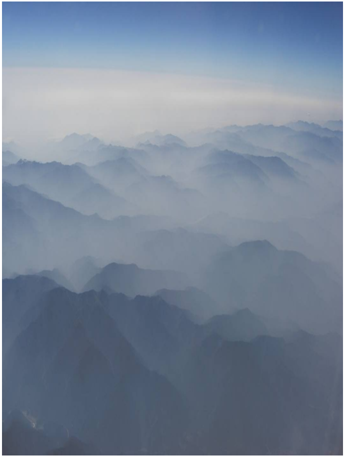
Space Basis

- B-splines $\{\psi_k\}_{k=1}^{K}$
- Eigenfunctions of Cov operator

$$Y_{tj} = \sum_{l=0}^{L} Z_{tl} m_l(X_{tj}) + \varepsilon_{tj} = M_t(X_{tj}) + \varepsilon_{tj}$$

- Functional Principal Component Analysis

Space Basis



Covariance operator

$\mathcal{C}(u, v) = \operatorname{Cov}\{M(u), M(v)\}$

Observe that $\mathcal{C}$ is a kernel:

Mercer's Theorem

$\mathcal{C}(u, v) = \sum_{k=1}^{\infty} \theta_k \psi_k(u) \psi_k(v)$

$\{\theta_k, \psi_k\}$ Eigenvalues, -functions

Space Basis

Karhunen - **Loève** expansion

$$M(u) = m_0(u) + \sum_{k=1}^{\infty} a_k \psi_k(u)$$

$$a_k = \int \{(M - m_0)\,\psi_k\}(v)dv$$

k-th PC scores, uncorrelated

Space Basis

$$\mathcal{C}(u,v) = \widehat{a}_0(u,v) - \widehat{a}(u)\widehat{a}(v)$$

$$\min_{a,b} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \left\{ Y_{tj} - a - \sum_{\nu=1}^{d} b_\nu (u_\nu - X_{tj,\nu}) \right\}^2 K_h(u - X_{tj})$$

$$\min_{a,b} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \sum_{i=1}^{J_t} \left\{ Y_{tj} Y_{ti} - a_0 - \sum_{\nu=1}^{d} b_\nu^1 (u_\nu - X_{tj,\nu}) - \sum_{\nu=1}^{d} b_\nu^2 (v_\nu^1 - X_{tj,\nu}) \right\}^2$$

$$\times K_h(u - X_{tj}) K_h(v - X_{tj})$$

Estimation Procedure

0 Space basis Ψ_t via **FPCA**

1 Time basis U_t selection via **group Lasso**; $\widehat{\beta}$ $(R \times K)$

$$\min_{\beta} (TJ)^{-1} \sum_{t=1}^{T} \left(Y_t^\top - U_t^\top \beta^\top \Psi_t \right) \left(Y_t^\top - U_t^\top \beta^\top \Psi_t \right)^\top + 2\lambda \|\beta\|_{2,1}$$

2 Split $\widehat{\beta}$ into $\widehat{\Gamma}$, $\widehat{A}$

$\widehat{\Gamma}$: first “L” eigenvectors of $\widehat{\beta}\widehat{\beta}^\top$; $\quad \widehat{A} = \widehat{\Gamma}^\top \widehat{\beta}$

Tuning parameter λ

- Take 100 equally spaced $\lambda \in [0, \max_r \| \sum_t \Psi_t Y_t U_{tr} \| / \sqrt{K}]$
- Evaluate $C_p(\lambda)$ where

$$
\begin{aligned}
C_p(\lambda) &= \frac{\sum_t \| Y_t^\top - U_t^\top \widehat{\beta}^\top \Psi_t \|^2}{\tilde{\sigma}^2} - JT + 2df \\
\tilde{\sigma}^2 &= \frac{\sum_t \| Y_t^\top - U_t^\top \widehat{\beta}_{OLS}^\top \Psi_t \|^2}{JT - df} \\
df &= \sum_r \mathbf{1}\{\| \widehat{\beta}_r \| > 0\} + \sum_r \frac{\| \widehat{\beta}_r \|}{\| \widehat{\beta}^{OLS} \|}(K-1)
\end{aligned}
$$

- Choose the minimal $C_p(\lambda)$

Sparsity of β^*

- Measured by s
- Low s implies high sparsity
- $\kappa(s)$ - decreasing function of s

Risk Bound (Gaussian)

Assume $\varepsilon \sim \mathrm{N}(0,\sigma^2)$ and $\lambda \sim (JT)^{-\frac{1}{2}}$ with probability at least $1 - R^{1-q}$ with $q = \min(A \log R, \sqrt{T})$, for all $\widehat{\beta}$:

$$(JT)^{-1} \sum_{t=1}^{T} \| \Psi_t^\top (\widehat{\beta} - \beta^*) U_t \|^2 \leqslant 64\sigma^2 s(1 + A \log R/\sqrt{T})/(\kappa^2 J),$$

$$T^{-1/2} \| \widehat{\beta} - \beta^* \|_{2,1} \leqslant 32\sigma s \sqrt{1 + A \log R/\sqrt{T}}/(\kappa^2 \sqrt{J}),$$

$$M(\widehat{\beta}) \leqslant 64 \phi_{max}^2 s/\kappa^2$$

- Dependence on R is negligible for large T

Risk Bound (Non-Gaussian)

$\varepsilon \sim (0, \sigma^2)$ and $\lambda \sim (JT)^{-\frac{1}{2}}$ with probability at least $1-(2e\log R - e)C/(\log R)^{1+\delta}$, for all $\widehat{\beta}$:

$$(JT)^{-1}\sum_{t=1}^{T} \| \Psi_t^\top (\widehat{\beta}-\beta^*)U_t \|^2 \leqslant 16\sigma^2 s(\log R)^{1+\delta}/(\kappa^2 J)$$

$$T^{-1/2}\| \widehat{\beta}-\beta^* \|_{2,1} \leqslant 16\sigma s\sqrt{(\log R)^{1+\delta}}/(\kappa^2\sqrt{J})$$

$$M(\widehat{\beta}) \leqslant 64\phi_{max}^2 s/\kappa^2$$

- Dependence on R surfaces

Risk Bound (Dependent)

Under technical assumptions and $\lambda \sim T^{-\frac{1}{2}}$ with probability at least $p(1 - R^{-\delta'})$, for $\forall\ \widehat{\beta}$:

$$(JT)^{-1} \sum_{t=1}^{T} \| \Psi_t^{\top} (\widehat{\beta} - \beta^*) U_t \|^2 \leqslant 16 \left(C' + \sqrt{\frac{\mathcal{X}^*(\mathcal{T}) \sum_t b_t^2}{(\log R)^{1-\delta'} T}} \right)^2 s/\kappa^2$$

$$T^{-1/2} \| \widehat{\beta} - \beta^* \|_{2,1} \leqslant 16 \left(C' + \sqrt{\frac{\mathcal{X}^*(\mathcal{T}) \sum_t b_t^2}{(\log R)^{1-\delta'} T}} \right) s/\kappa^2$$

$$M(\widehat{\beta}) \leqslant 64 \phi_{max}^2 s/\kappa^2$$

- Dependence level $\nearrow$, bound $\nearrow$

$\widehat{Z}_{0,t}$ not get affected

Under technical assumptions and $\lambda \sim T^{-\frac{1}{2}}$

$$\frac{1}{T}\sum_{1\leq t\leq T}\left\|\widehat{Z}_{0,t}^{\top}\widehat{A}-Z_{0,t}^{\top}A^{*}\right\|^{2}=\mathcal{O}_{P}(\rho^{2}+\delta_{K}^{2}).$$

for $\widehat{\beta}$ "close enough" to β

- Approximation $\searrow$, bound $\searrow$

Covariance Equivalence

Under technical assumptions and $\lambda \sim T^{-\frac{1}{2}}$, for $h \geq 0$

$$T^{-1} \sum_{t=\max[1,\text{-h}+1]}^{\min[T,T-h]} \widetilde{Z}_{0,t} \left(\widetilde{Z}_{0,t+h} - \widetilde{Z}_{0,t} \right)^{\top} - Z_{0,t} \left(Z_{0,t+h} - Z_{0,t} \right)^{\top} = \mathcal{O}_P$$

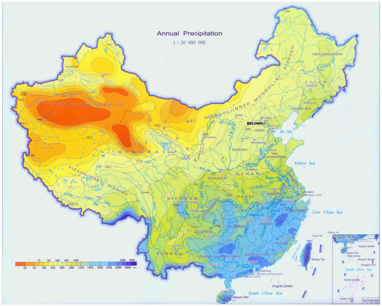
$$T^{-1} \sum_{t=\max[1,\text{-h}+1]}^{\min[T,T-h]} \widetilde{Z}_{n,t} \widetilde{Z}_{n,t+h}^{\top} - Z_{n,t} Z_{n,t+h}^{\top} = \mathcal{O}_P$$

where:

$B \stackrel{\text{def}}{=} (\sum_{t=1}^{T} Z_{0,t} \widehat{Z}_{0,t})^{-1} \sum_{t=1}^{T} Z_{0,t} Z_{0,t}^{\top}$, $\quad \widetilde{Z}_{0,t} \stackrel{\text{def}}{=} B^{\top} \widehat{Z}_{0,t}$,

$\widetilde{Z}_{n,t} \stackrel{\text{def}}{=} (T^{-1} \sum_{s=1}^{T} \widetilde{Z}_{0,s} \widetilde{Z}_{0,s}^{\top})^{-1/2} \widetilde{Z}_{0,t}$, $\quad Z_{n,t} \stackrel{\text{def}}{=} (T^{-1} \sum_{s=1}^{T} Z_{0,s} Z_{0,s}^{\top})^{-1/2} Z_{0,t}$

Temperature and Climate Change



Weather stations and China Climate Types

Temperature and Climate Change

- Space basis via **FPCA**

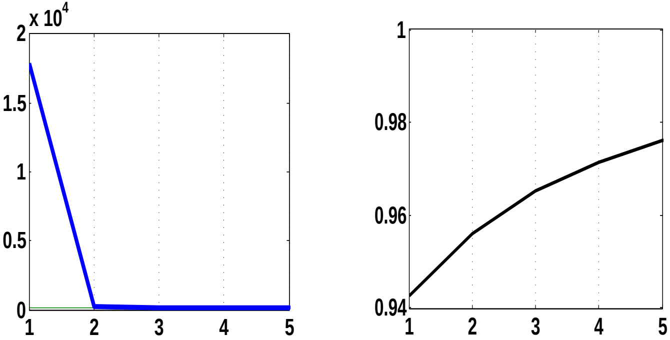


Figure 11: Distribution of the eigenvalues (left) and explained variance by the first K basis (right)

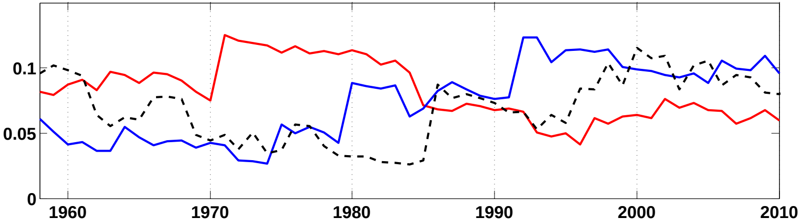
Temperature and Climate Change

- Initial time basis $(53 \cdot 3 + 20 = 179)$

	Factors		Factors
Trend	1	Large	$\sin\{2\pi t/(365 \cdot 10)\}$
(Year by Year)	t	Period	$\cos\{2\pi t/(365 \cdot 10)\}$
	$3t^2 - 1$		$\sin\{4\pi t/(365 \cdot 10)\}$
Seasonal	$\sin\{2\pi t/365\}$		$\cos\{4\pi t/(365 \cdot 10)\}$
Effect	$\cos\{2\pi t/365\}$		$\sin\{6\pi t/(365 \cdot 10)\}$
	$\ldots$		$\ldots$
	$\cos\{20\pi t/365\}$		$\cos\{20\pi t/(365 \cdot 10)\}$

Time Basis Coefficients

- Long term: linear, quadratic trend - warming effect
- Yearly variation ($p = 365$): earth rotation
- 10-year variation ($p = 3650$): solar activity



Estimated coefficients of the 1st factor $\widehat{\Gamma}_{r1}$ w.r.t. the yearly polynomial time basis (constant, linear, quadratic)

Estimated Stochastic Process $\widehat{Z}_{0,t,1}$

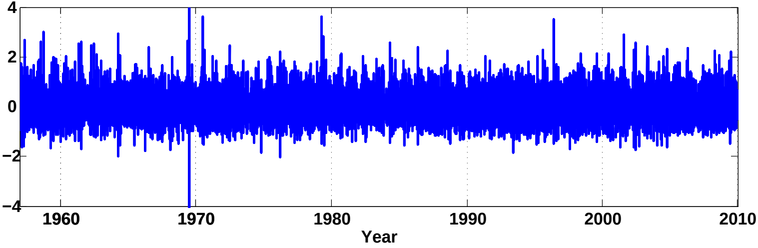


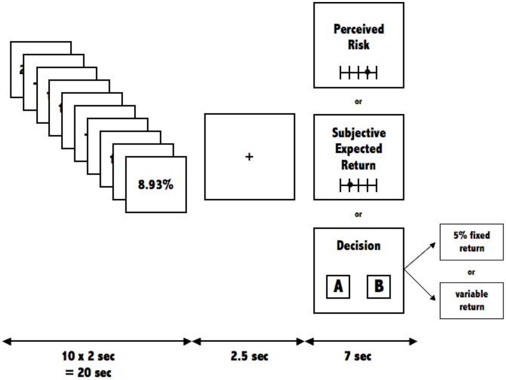
Figure 12: Estimated stochastic process $\widehat{Z}_{0,t,1}$

Modeling of $\widehat{Z}_{0,t}$

$\widehat{Z}_{0,t} = \mathcal{R}\widehat{Z}_{0,t-1} + \varepsilon_{0,t}$ with random vector $\varepsilon_{0,t}$ and estimated coefficient matrix:

$$\begin{pmatrix} 1.0003 & 0.0132 & -0.0002 & -0.3422 & 0.0786 \\ 0.0082 & 1.0111 & 0.1426 & 0.2673 & -0.0288 \\ 0.0019 & 0.0282 & 0.8362 & 0.1471 & 0.1688 \\ -0.0014 & -0.0290 & -0.1195 & 0.7185 & -0.0648 \\ -0.0017 & -0.0002 & -0.1155 & -0.0708 & 0.8151 \end{pmatrix}.$$

Risk Perception



Returns Pause Decision

Risk Perception

- 3D fMRI images data
  - Panel **GDSFM** $Y^i_{t,j}, 1 \le i \le I$

$$Y^i_{t,j} = \sum_{l=1}^{L} (\alpha^i_{t,l} + U_t^\top \Gamma^i_l) m_l(X_{t,j}) + \varepsilon_{t,j}, \;\; 1 \le j \le J_t, \; 1 \le t \le T,$$

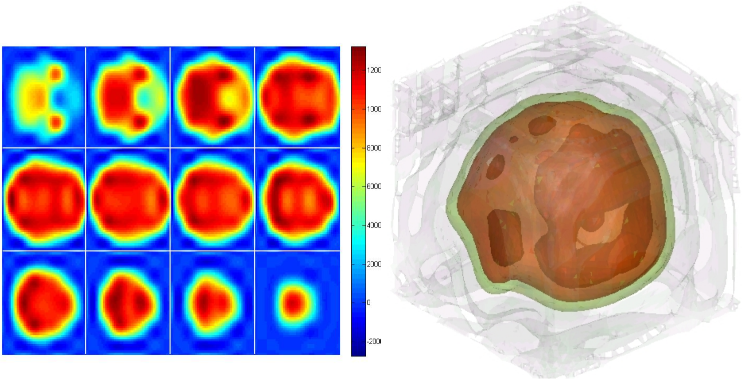
with fixed effect $\alpha^i_{t,l}$ and

$$\sum_{i=1}^{I} \left(\sum_{l=1}^{L} \alpha^i_{t,l} m_l(X_{t,j}) | X_{t,j} \right) = 0$$

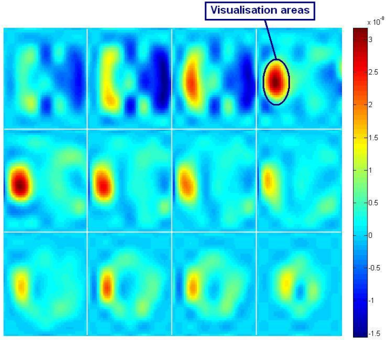
Risk Perception

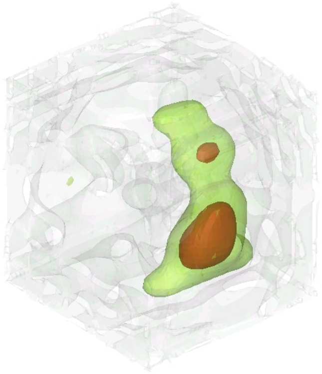
1. Average $Y^i_{t,j}$ over i and estimate factor loadings m_l
2. Given m_l, for i, estimate $Z^i_{t,l}$.

$$Y^i_{t,j} = \sum_{l=1}^{L} U_t^\top \Gamma^i_l \overline{m}_l(X_{t,j}) + \varepsilon^i_{t,j}$$

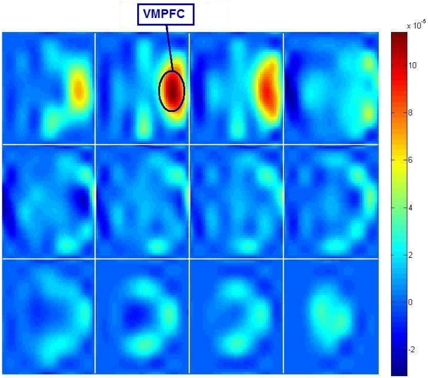


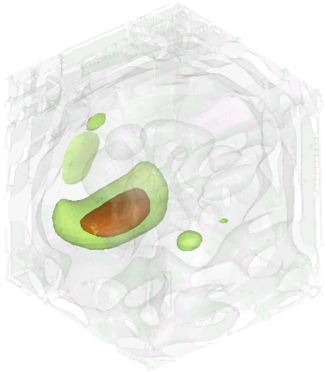
Estimated factor loading $\hat{m}_1$ with $L = 5$.



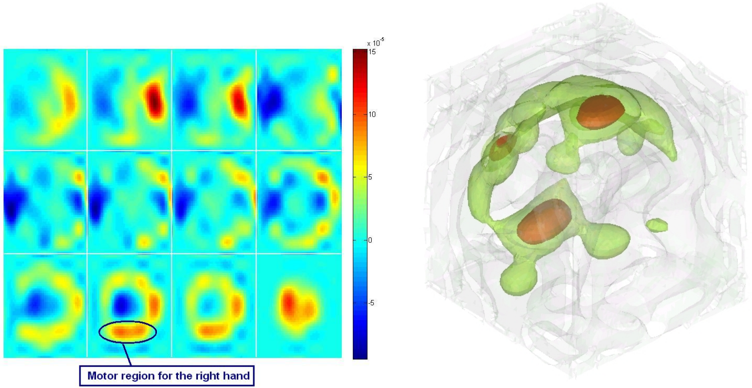


Estimated factor loading $\hat{m}_2$ with $L = 5$.

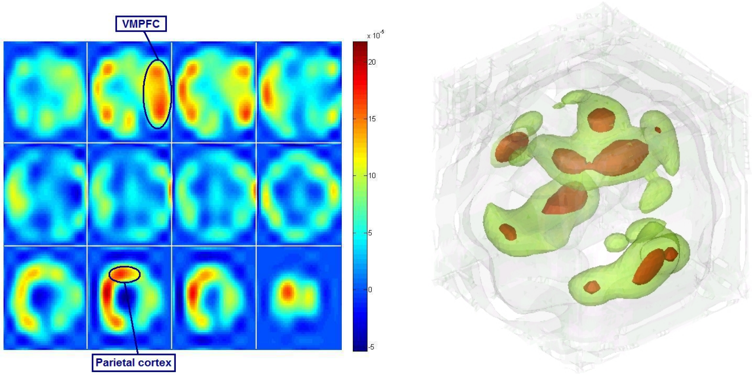




Estimated factor loading $\hat{m}_3$ with $L = 5$.
(VMPFC = Ventromedial prefrontal cortex)

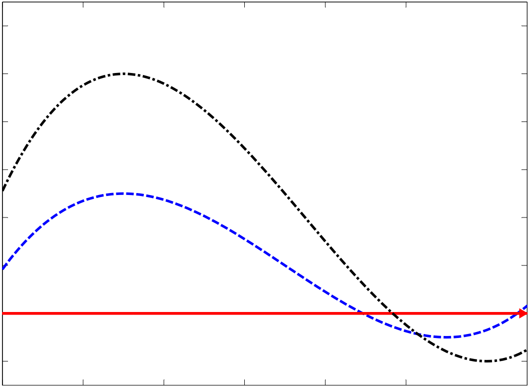


Estimated factor loading $\hat{m}_4$ with $L = 5$.



Estimated factor loading $\hat{m}_5$ with $L = 5$.

Response to Stimuli



Response curves (to stimuli) $U_t^\top \widehat{\Gamma}_2^i$ for probands $i = 9$ & $i = 19$ with periodic cubic polynomial as time basis

SVM Analysis (Risk)

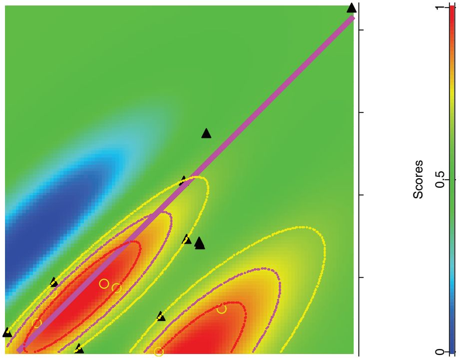
- Different subjects' response curves have different shapes
- SVM based on the $\widehat{\beta}$

MEAN		Estimated	
Data	Strongly	0.85	0.14
	Weakly	0.59	0.40

Table 1: Classification rates of the SVM method.

The rates hold over a wide range of parameters

SVM Classification



Yield Curve Modelling

- Yield Curve - $Y_{t,j} \in \mathbb{R}^{13}$
- 4 countries: AU, EU, JP and US, 20001013-20101013

- Explanatory variables
  - Time to maturity - $X_t^1 \in \mathbb{R}$
  - Real GDP changes - $X_t^2 \in \mathbb{R}$
    De-trended using Hodrick-Prescott (HP) filter

Explained Variance

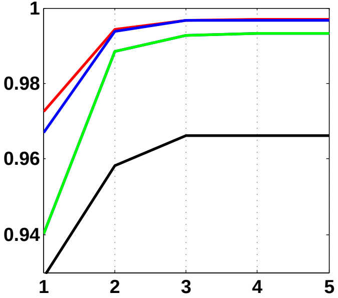


Figure 13: Explained variance for the first 5 factors for AU, EU, JP and US

Estimated First Factor, $\widehat{m}_1$

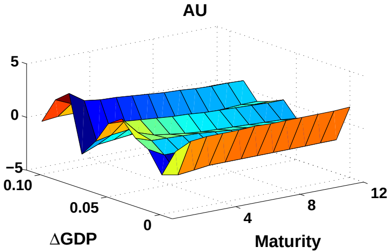


Figure 14: Estimated first factor

Estimated First Factor, $\widehat{m}_1$

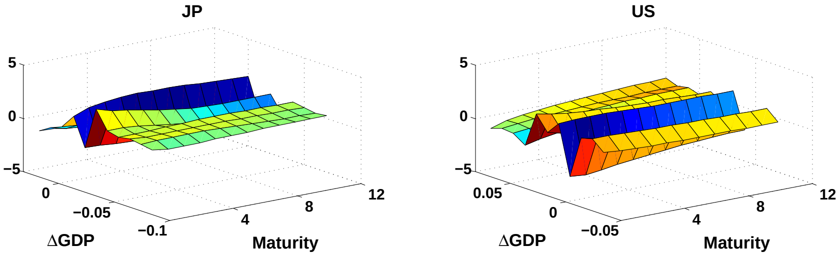


Figure 15: Estimated first factor

Estimated Factor Loadings, $\widehat{Z}_t$

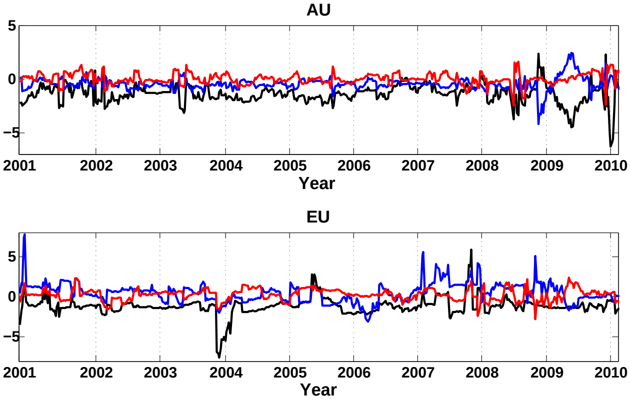


Figure 16: Estimated factor loadings: $\widehat{Z}_{1t}$, $\widehat{Z}_{2t}$ and $\widehat{Z}_{3t}$

Estimated Factor Loadings, $\widehat{Z}_t$

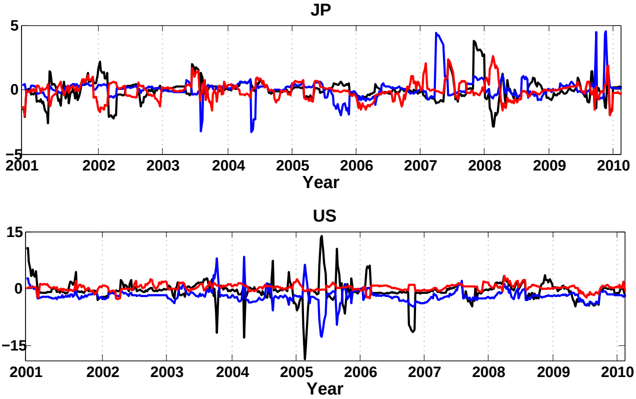


Figure 17: Estimated factor loadings: $\widehat{Z}_{1t}$, $\widehat{Z}_{2t}$ and $\widehat{Z}_{3t}$

Dynamic Semiparametric Factor Model

- Two-step estimation
- Low dimensional representation
- Applications
  - Neurobiology
  - Meteorology
  - Finance

Dimensionality ✓

Dependency ✓

Divergency ✓

High Dimensional Nonstationary Time Series Modelling With Dynamic Factor Models

Wolfgang Karl Härdle
(P. Majer, Y. Ritov, R. Song)

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
CDA - Centre for Complex Data Analysis
The National Central University
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://www.stat.ncu.edu.tw

References

Bakshi, G., Cao, C. and Chen, Z.
"Do call and underlying prices always move in the same direction?"
Review of Financial Studies, 13(3): 549–584, 2000.

Fengler, M., Härdle, W. and Mammen, E.
A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics
Journal of Financial Econometrics, 5(2): 189–218, 2007.

Hall, P., Müller, H. G. and Wang J.
Properties of principal component methods for functional and longitudinal data analysis
Ann. Statist., 34(3): 1493–1517, 2006.

References

Janson, S.
Large deviations for sums of partly dependent random variables
Random Structures Algorithms, 24(3): 234–248, 2004.

Jia, J., Rohe, K. and Yu, B.
The Lasso under Heteroscadecity
Technical Report 783, Statistics Department, UC Berkeley., 2009.

Kneip, A.
Ordered linear smoothers.
Ann. Statist., 22: 835-866.

References

Park, B., Mammen, E., Härdle, W. and Borak, S.
Time Series Modelling with Semiparametric Factor Dynamics
J. Amer. Stat. Assoc., 104(485): 284-298, 2009.

Ramsay, J. O. and Silverman, B. W.
Functional Data Analysis
New York: Springer.

Tibshirani, R.
Regression Shrinkage and Selection via the Lasso
J. R. Statist. Soc. B, 58: 267–288, 1996.

Technical assumptions

- A1 Normalization of U_t & Ψ_t: $\Psi_t \Psi_t^\top / J = I_K$, $\sum_{t=1}^T U_t^\top U_t / R = 1$
- A2 The number of nonzero β_r^*s: $M(\beta^*) \le s$
- A3 ϕ_{max} is the maximum eigenvalue of $\sum_{t=1}^T U_t U_t^\top$
- A4 The error terms $\varepsilon_1, \ldots, \varepsilon_T$ are i.i.d. **Gaussian** with mean 0 and variance $\sigma^2 I_{J \times J}$
- A5 The error terms $\varepsilon_1, \ldots, \varepsilon_T$ are independent with mean 0 and **finite** variance $\mathrm{E}(\varepsilon_{tj}^2) \leqslant \sigma^2$

Assumption

There exists a positive number $\kappa = \kappa(s)$ *such that*

$$\min\left\{\frac{\sum_t \|\Psi_t^\top \Delta U_t\|}{\sqrt{J} \parallel \Delta_{\mathcal{R}} \parallel} : |\mathcal{R}| \leqslant s, \Delta \in \mathbb{R}^{K\times R}\backslash\{0\}, \right.$$
$$\left. \parallel \Delta_{\mathcal{R}^c} \parallel_{2,1} \leqslant 3 \parallel \Delta_{\mathcal{R}} \parallel_{2,1} \right\} \geqslant \kappa.$$

- Restriction on the eigenvalues of U_t as a function of sparsity s
- Low sparsity, s big, κ small

Assumption

The matrices Ψ_t and U_t are such that

$$(JT)^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\max_r \Big| \sum_{k=1}^{K} \Psi_{tkj} U_{tr} \Big| \right)^2 \leqslant C,$$

for a constant $C > 0$.

Measure of Dependence, Jason (2004)

Given a set $\mathcal{T}$ and random variables V_t, $t \in \mathcal{T}$, we say:

- A subset $\mathcal{T}'$ of $\mathcal{T}$ is *independent* if the corresponding random variables $\{V_t\}_{t \in \mathcal{T}'}$ are independent.
- A family $\{\mathcal{T}_j\}_j$ of subsets of $\mathcal{T}$ is a *cover* of $\mathcal{T}$ if $\bigcup_j \mathcal{T}_j = \mathcal{T}$.
- A family $\{(\mathcal{T}_j, w_j)\}_j$ of pairs $(\mathcal{T}_j, w_j)$, where $\mathcal{T}_j \subseteq \mathcal{T}$ and $w_j \in [0,1]$ is a *fractional cover* of $\mathcal{T}$ if $\sum_j w_j \mathbf{1}_{\mathcal{T}_j} \geqslant \mathbf{1}_{\mathcal{T}}$, i.e. $\sum_{j: t \in \mathcal{T}_j} w_j \geqslant 1$ for each $t \in \mathcal{T}$.
- A (fractional) cover is *proper* if each set $\mathcal{T}_j$ in it is independent.
- $\mathcal{X}(\mathcal{T})$ is the size of the smallest proper cover of $\mathcal{T}$, i.e. the smallest m such that $\mathcal{T}$ is the union of m independent subsets.
- $\mathcal{X}^*(\mathcal{T})$ is the minimum of $\sum_j w_j$ over all proper fractional covers $\{(\mathcal{T}_j, w_j)\}_j$.

$\mathcal{X}^*(\mathcal{T})$: measure of dependence; $\mathcal{X}^*(\mathcal{T}) = 1$ (independent).

Assumption

With a high probability p, Ψ_t, U_t and ε_t are such that

$$
\begin{aligned}
(J^{-1} \sum_{k=1}^{K} \sum_{j=1}^{J} \Psi_{tkj} \varepsilon_{tj} U_{tr})^2 &\leqslant b_t^2 \\
\mathsf{E}(JT)^{-1} \Big\{ \sum_{t=1}^{T} (\sum_{k=1}^{K} \sum_{j=1}^{J} \Psi_{tkj} \varepsilon_{tj} U_{tr})^2 \Big\}^{1/2} &\leqslant \frac{C'}{\sqrt{T}}.
\end{aligned}
$$

for $\forall\ r$ *and some constants* $b_t, C' > 0, t = 1, \ldots, T$.

Technical assumptions

- B1 $X_{1,1},\dots,X_{T,J}$, $\varepsilon'_{1,1},\dots,\varepsilon'_{T,J}$, and $Z_{0,1}, \ldots, Z_{0,T}$ are independent.
- B2 $X_{t,1},\dots,X_{t,J}$ are identically distributed, support $[0,1]^d$ and a density f_t that is bounded from below and above on $[0,1]^d$, uniformly over $t=1,\dots,T$.
- B3 We assume that $\mathsf{E}\,\varepsilon'_{t,j}=0$ for $1\le t\le T, 1\le j\le J$, and for $c>0$ small enough $\sup_{1\le t\le T, 1\le j\le J}\mathsf{E}\exp\{c(\varepsilon'_{t,j})^2\}<\infty$.
- B4 The vector of functions $m=(m_1,\dots,m_L)^\top$ can be approximated by Ψ_k, i.e.

$$\delta_K \stackrel{\text{def}}{=} \sup_{x\in[0,1]^d} \inf_{A\in\mathbb{R}^{L\times K}} \|m(x)-A\Psi(x)\| \to 0$$

as $K\to\infty$. We denote A that fulfills $\sup_{x\in[0,1]^d}\|m(x)-A\Psi(x)\|\le 2\delta_K$ by A^*.

B5 There exist constants $0 < C_L < C_U < \infty$ such that all eigenvalues of the matrix $T^{-1}\sum_{t=1}^{T} Z_{0t} Z_{0,t}^{\top}$ lie in the interval $[C_L, C_U]$ with probability tending to one.

B6 The minimization (1) runs over all values β with

$$\sup_{x\in[0,1]^d} \max_{1\leqslant t\leqslant T} \|Z_{0,t}^{\top} A\Psi(x)\| \leqslant M_T,$$

where the constant M_T fulfils $\max_{1\leqslant t\leqslant T} \|Z_{0,t}\| \leqslant M_T/C_m$ (with probability tending to one) for a constant C_m such that $\sup_{x\in[0,1]^d} \|m(x)\| < C_m$.

B7 It holds that $\rho^2 = (K+T)M_T^2 \log(JTM_T)/(JT) \to 0$. The dimension L is fixed.

Technical assumptions

C1 $Z_{0,t}$ is a strictly stationary sequence with $\mathrm{E}(Z_{0,t}) = 0$, $\mathrm{E}(\|Z_{0,t}\|^\gamma) < \infty$ for some $\gamma > 2$. It is strongly mixing with $\sum_{i=1}^{\infty} \alpha(i)^{(\gamma-2)/\gamma} < \infty$. The matrix $\mathrm{E}\, Z_{0,t} Z_{0,t}^\top$ has full rank. The process $Z_{0,t}$ is independent of $X_{11}, \ldots, X_{TJ}, \varepsilon'_{11}, \ldots, \varepsilon'_{TJ}$.

C2 It holds that

$$[\log(KT)^2\{(KM_T/J)^{1/2} + T^{1/2} M_T^4 J^{-2} + K^{3/2} J^{-1} + K^{4/3} J^{-2/3} T^{-1/6}\} + 1] T^{1/2}(\rho^2 + \delta_K^2) = \mathcal{O}(\rho^2 + \delta_K^2)$$