

Chao, Shih-Kang; Härdle, Wolfgang Karl; Huang, Chen

Working Paper

Multivariate factorisable sparse asymmetric least squares regression

SFB 649 Discussion Paper, No. 2016-058

Provided in Cooperation with:

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Chao, Shih-Kang; Härdle, Wolfgang Karl; Huang, Chen (2016) : Multivariate factorisable sparse asymmetric least squares regression, SFB 649 Discussion Paper, No. 2016-058, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/162500>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Multivariate Factorisable Sparse Asymmetric Least Squares Regression

Shih-Kang Chao *
Wolfgang K. Härdle *²
Chen Huang *²



* Purdue University, United States of America

*² Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



Multivariate Factorisable Sparse Asymmetric Least Squares Regression*

Shih-Kang Chao[†], Wolfgang K. Härdle[‡], Chen Huang[§]

December 23, 2016

Abstract

More and more data are observed in form of curves. Numerous applications in finance, neuroeconomics, demographics and also weather and climate analysis make it necessary to extract common patterns and prompt joint modelling of individual curve variation. Focus of such joint variation analysis has been on fluctuations around a mean curve, a statistical task that can be solved via functional PCA. In a variety of questions concerning the above applications one is more interested in the tail asking therefore for tail event curves (TEC) studies. With increasing dimension of curves and complexity of the covariates though one faces numerical problems and has to look into sparsity related issues.

Here the idea of FActorisable Sparse Tail Event Curves (FASTEC) via multivariate asymmetric least squares regression (expectile regression) in a high-dimensional framework is proposed. Expectile regression captures the tail moments globally and the smooth loss function improves the convergence rate in the iterative estimation algorithm compared with quantile regression. The necessary penalization is done via the nuclear norm. Finite sample oracle properties of the estimator associated with asymmetric squared error loss and nuclear norm regularizer are studied formally in this paper.

As an empirical illustration, the FASTEC technique is applied on fMRI data to see if individual's risk perception can be recovered by brain activities. Results show that factor loadings over different tail levels can be employed to predict individual's risk attitudes.

JEL classification: C38, C55, C61, C91, D87

Keywords: high-dimensional M -estimator, nuclear norm regularizer, factorization, expectile regression, fMRI, risk perception, multivariate functional data

*Financial support from the Deutsche Forschungsgemeinschaft via CRC 649 "Economic Risk" and IRTG 1792 "High Dimensional Non Stationary Time Series", Humboldt-Universität zu Berlin, is gratefully acknowledged. The authors would like to thank Prof. Martin A. Lindquist for the discussion and suggestions on this paper.

[†]Department of Statistics, Purdue University, 250 N University St., West Lafayette IN 47907-2066, U.S.A. Partially supported by Office of Naval Research.

[‡]C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany. Sim Kee Boon Institute for Financial Economics, Singapore Management University, 50 Stamford Road, Singapore 178899, Singapore.

[§]Corresponding author. C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany. Email: chen.huang@hu-berlin.de.

1 Introduction

Data are observed more and more in form of curves, thus prompting a joint modelling to extract common patterns and also individual curves variations. Such data curve modelling occurs e.g., in neuroeconomics, weather and climate analysis, demographics among many other disciplines. A well known tool in these situations is Functional data analysis (FDA) that studies the variation of random curve objects in a high dimensional content. Leading references are Ramsay and Silverman (2002, 2005). Treating these random objects as curves FDA provides insight into main factors, typically extracted as principal components via a Karhunen-Loève decomposition. A commonly used approach is to fit the individual observation $\mathbf{Y}_j \in \mathbb{R}^n$ (j indicates individuals) via a basis or series approximation and then to enter a spectral analysis e.g., based on the Fourier coefficients of the series expansion. This leads via inspection of the eigenvalues to a lower dimensional factor model. This approach has been successfully employed in many situations, see, e.g., Yao et al. (2003); Hall et al. (2006).

Focus of such joint variation analysis has been on fluctuations around a mean curve, a statistical task that can be solved via functional principal component analysis. However, in a variety of questions concerning the above applications one is more interested in the tail variations asking therefore for tail event curves (TEC) studies. TEC studies may be performed through smooth approximation of conditional tail probabilities. More generally though one needs to look at functions based on conditional tail events; it helps to discover "extreme curves" which are aberrant from the majorities. Modeling this way the TECs require to deviate from Hilbert L_2 geometry and to introduce asymmetric norms or loss functions, Koenker and Bassett (1978); Newey and Powell (1987); Breckling and Chambers (1988), and more recent work on principal component analysis with asymmetric norm by Tran et al. (2016). Also in climate weather analysis and electricity load forecasting, distributional forecasts characterized by tail measures are shown to be powerful, Cabrera and Schulz (2016).

In scatterplot smoothing and multivariate settings, quantile regressions have been studied under different approaches. A survey is given by Serfling (2002). Computational challenges arise in high-dimensional multi-task quantile regression due to the non-smooth absolute loss. Asymmetric least squares (ALS) regression (Efron, 1991), known as expectile regression as well, can capture the complete conditional distribution as quantile regression does. While associated with a smooth differentiable loss, it is more desirable if we have to pay attention to the computational convenience and efficiency in a high-dimensional framework. Expectile as a generalization of mean is more and more appealing in financial econometrics since it is more sensitive to the magnitude of extreme losses, Taylor (2008); Kuan et al. (2009); Xu et al. (2015). It plays a crucial role in risk

management because of its conventional interpretation: it specifies the sufficient amount of money required to maintain a position given a gain-loss ratio (Bernardo and Ledoit, 2000). For industry investors this notion of loss is certainly more attractive than the pure probability of a loss as given via the definition of quantile. Moreover, among other popular risk measures such as Value at Risk (VaR) and expected shortfall (ES), expectile is the only one enjoys elicitable law-invariant properties (Ziegel, 2016), which are desired in forecasts and risk diversification.

On the other hand, with increasing dimension of curves and complexity of the covariates though one faces numerical problems and has to look into sparsity related issues. A natural way to reduce the burden of this estimation task is to introduce a penalty term. Yuan et al. (2007) proposed a penalization approach with nuclear norm, the sum of the singular values of the coefficient matrix, as the penalty. Numerically the estimator can be readily obtained since it involves a convex optimization. Moreover, it leads via thresholdings of the eigenvalues to a low dimensional factor model. Compared with previous research such as the reduced rank approach by Izenman (1975), the number of factors does not need to be predetermined. The dimension reduction and coefficient estimation can be done simultaneously, thus leading to a handy tool in data analysis of many curves.

Following these lines of thoughts we propose Factorisable Sparse Tail Event Curves (FASTEC) via multivariate asymmetric least squares regression. We employ FISTA technique developed by Beck and Teboulle (2009) to solve the optimization. Expectile regression captures the tail moments globally and the smooth loss function improves the convergence rate in the iterative procedure compared with the quantile regression case (Chao et al., 2015). The finite sample oracle properties of the estimator are established formally.

As an empirical illustration, FASTEC is applied on functional Magnetic Resonance Imaging (fMRI) data recorded during investment decisions experiment. To be more specific, multivariate factorisable sparse asymmetric least squares regression is employed to jointly model all response curves with multivariate functional data. We expect that individual's risk perception is predictable with one's brain reactions, particularly after taking tail risks into consideration.

The rest of the paper is arranged as follows. Section 2 introduces the model setting, estimation method and finite sample oracle properties of the estimator. Section 3 illustrates the empirical application with fMRI data. Detailed proofs are provided in appendices. The codes to implement the algorithms are publicly accessible via  www.quantlet.de.

2 Model and Estimation

2.1 Model Setting

We start with defining some notations. For a matrix $\mathbf{S} = (s_{lj}) = [\mathbf{S}_{\cdot 1} \dots \mathbf{S}_{\cdot m}] \in \mathbb{R}^{p \times m}$, where $\mathbf{S}_{\cdot j} \in \mathbb{R}^p$ be the column vectors. Let $\|\mathbf{S}\|_F$, $\|\mathbf{S}\|_*$ and $\|\mathbf{S}\|$ be the matrix Frobenius, nuclear and spectral norm. Denote $\sigma_{\min}(\mathbf{S})$ and $\sigma_{\max}(\mathbf{S})$ the smallest and largest singular values. For a vector $\mathbf{v} \in \mathbb{R}^p$, $\|\mathbf{v}\|_2$ is the Euclidean norm. Define $\langle\langle \mathbf{A}, \mathbf{B} \rangle\rangle \stackrel{\text{def}}{=} \text{tr}(\mathbf{A}^\top \mathbf{B})$.

Let $\{(\mathbf{X}_i, Y_{i1}, \dots, Y_{im})\}_{1 \leq i \leq n}$ be i.i.d. samples, with $Y_{ij} \in \mathbb{R}$ and $\mathbf{X}_i \in \mathbb{R}^p$. We note that Y_{ij} and Y_{ik} may be dependent, and m and p can diverge with n . For $\tau \in (0, 1)$, the conditional expectile $e_j(\tau | \mathbf{X}_i)$ of Y_{ij} given $\mathbf{X}_i$ is defined by

$$e_j(\tau | \mathbf{X}) \stackrel{\text{def}}{=} \arg \min_{\theta} \mathbb{E}[\rho_{\tau}(Y_{ij} - \theta) | \mathbf{X}], \quad (2.1)$$

with $\rho_{\tau}(u) \stackrel{\text{def}}{=} |\tau - \mathbf{1}\{u < 0\}| |u|^2$. In particular, we assume a factor structure:

$$e_j(\tau | \mathbf{X}_i) = \sum_{k=1}^r \psi_{j,k}(\tau) f_k^{\tau}(\mathbf{X}_i), \quad (2.2)$$

where $f_k^{\tau}(\mathbf{X}_i)$ is the k th factor, r is the number of factors (much less than p) and $\psi_{j,k}(\tau)$ are the factor loadings. Furthermore, factors are constructed by linear combinations of covariates $\mathbf{X}_i$:

$$f_k^{\tau}(\mathbf{X}_i) = \mathbf{X}_i^{\top} \boldsymbol{\varphi}_k(\tau). \quad (2.3)$$

Substituting (2.3) into (2.2) yields

$$e_j(\tau | \mathbf{X}_i) = \mathbf{X}_i^{\top} \boldsymbol{\gamma}_j(\tau), \quad (2.4)$$

where $\boldsymbol{\gamma}_j(\tau) = (\sum_{k=1}^r \psi_{j,k}(\tau) \varphi_{k,1}(\tau), \dots, \sum_{k=1}^r \psi_{j,k}(\tau) \varphi_{k,p}(\tau))^{\top}$ as the unknown coefficient vector. Define $\boldsymbol{\Gamma} \stackrel{\text{def}}{=} [\boldsymbol{\gamma}_1 \dots \boldsymbol{\gamma}_m]$, the factor model (2.2) implies that $\boldsymbol{\Gamma}$ is of rank r , and the model (2.4) corresponds to a multivariate linear regression model. For standard regression with square loss, Reinsel and Velu (1998) propose to estimate $\boldsymbol{\Gamma}$ with reduced-rank regression under the knowledge of r . However, r is usually unknown in practice. Yuan et al. (2007) propose to perform the multivariate regression with nuclear norm penalty, which does not require the knowledge of r . The latter inspire the use of nuclear norm penalty in the next section. It is important to note that both methods can only apply to small number of p and m , and do not scale up to large dimensions.

Suppose an estimator $\hat{\boldsymbol{\Gamma}}$ is available, we can estimate the k th factor $\hat{f}_k^{\tau}(\mathbf{X}_i) = \mathbf{X}_i^{\top} \hat{\boldsymbol{\varphi}}_k(\tau) = \sigma_k \mathbf{X}_i^{\top} \mathbf{U}_{\cdot k}$ and the factor loadings for the j th curve $\hat{\psi}_j(\tau) = \mathbf{V}_{j\cdot}$, where $\mathbf{U}$ and $\mathbf{V}$ are

unitary matrices obtained from singular value decomposition: $\hat{\mathbf{\Gamma}} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$.

2.2 Algorithm

To estimate our model under factor model (2.2), we combine asymmetric loss with nuclear norm penalty. To be more specific, it is proposed to estimate $\mathbf{\Gamma}$ defined in Section 2.1 by solving:

$$\hat{\mathbf{\Gamma}}_\tau(\lambda) \stackrel{\text{def}}{=} \arg \min_{\mathbf{\Gamma} \in \mathbb{R}^{p \times m}} F(\mathbf{\Gamma}), \quad (2.5)$$

$$F(\mathbf{\Gamma}) \stackrel{\text{def}}{=} (mn)^{-1} \sum_{i=1}^n \sum_{j=1}^m \rho_\tau(Y_{ij} - \mathbf{X}_i^\top \mathbf{\Gamma}_{\cdot j}) + \lambda \|\mathbf{\Gamma}\|_*, \quad (2.6)$$

where λ is a tuning parameter, $\mathbf{\Gamma}_{\cdot j}$ is the j th column of $\mathbf{\Gamma}$. The second term nuclear norm $\|\mathbf{\Gamma}\|_*$ is defined by $\sum_{l=1}^{\min(p,m)} \sigma_l(\mathbf{\Gamma})$ given the singular values of $\mathbf{\Gamma}$ (square roots of non-zero eigenvalues of both $\mathbf{\Gamma}^\top \mathbf{\Gamma}$ and $\mathbf{\Gamma} \mathbf{\Gamma}^\top$): $\sigma_1(\mathbf{\Gamma}) \geq \sigma_2(\mathbf{\Gamma}) \geq \dots \geq \sigma_{\min(p,m)}(\mathbf{\Gamma})$. We note that (2.6) is a convex optimization problem that can be solved efficiently. The number of factors r in (2.2) does not need to be specified. To simplify the notation, we denote $\hat{\mathbf{\Gamma}}$ for $\hat{\mathbf{\Gamma}}_\tau(\lambda)$ hereinafter.

To solve the optimization problem (2.6), we apply the fast iterative shrinkage-thresholding algorithm (FISTA) of Beck and Teboulle (2009). FISTA is a popular algorithm for optimization problems of the form:

$$\min_{\mathbf{\Gamma}} \{g(\mathbf{\Gamma}) + h(\mathbf{\Gamma})\}, \quad (2.7)$$

where g is a smooth convex function with Lipschitz continuous gradient ∇g ,

$$\|\nabla g(\mathbf{\Gamma}_1) - \nabla g(\mathbf{\Gamma}_2)\|_F \leq L_{\nabla g} \|\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2\|_F, \forall \mathbf{\Gamma}_1, \mathbf{\Gamma}_2 \in \mathbb{R}^{p \times m}, \quad (2.8)$$

where $L_{\nabla g}$ is the Lipschitz constant of ∇g and h is a continuous convex (possibly nonsmooth) function (Ji and Ye, 2009). In view of (2.6), this corresponds to

$$g(\mathbf{\Gamma}) \stackrel{\text{def}}{=} (mn)^{-1} \sum_{i=1}^n \sum_{j=1}^m \rho_\tau(Y_{ij} - \mathbf{X}_i^\top \mathbf{\Gamma}_{\cdot j}), \quad (2.9)$$

$$h(\mathbf{\Gamma}) \stackrel{\text{def}}{=} \lambda \|\mathbf{\Gamma}\|_*. \quad (2.10)$$

The Lipschitz constant of ∇g is $L_{\nabla g} = 2(mn)^{-1} \max(\tau, 1 - \tau) \|\mathbf{X}\|_F^2$ will be calculated in

(A.4) in Appendix A.1. The FISTA algorithm is described in Algorithm 1.

Algorithm 1: Fast Iterative Shrinkage Thresholding Algorithm	
Input: $\{\mathbf{Y}_i\}_{i=1}^n, \{\mathbf{X}_i\}_{i=1}^n, \lambda$	
Output: $\hat{\mathbf{\Gamma}} = \mathbf{\Gamma}_T$	
1 Initialization: $\mathbf{\Gamma}_0 = 0, \mathbf{\Omega}_1 = 0$, step size $\delta_1 = 1$;	
2 for $t = 1, 2, \dots, T$ do	
3	$\mathbf{\Gamma}_t = \text{SVT}_\lambda(\mathbf{\Omega}_t - L_{\nabla g}^{-1} \nabla g(\mathbf{\Omega}_t));$
4	$\delta_{t+1} = \frac{1 + \sqrt{1 + 4\delta_t^2}}{2};$
5	$\mathbf{\Omega}_{t+1} = \mathbf{\Gamma}_t + \frac{\delta_t - 1}{\delta_{t+1}}(\mathbf{\Gamma}_t - \mathbf{\Gamma}_{t-1});$
6 end	

The subroutine SVT_λ in Algorithm 1 is the singular value thresholding given by $\text{SVT}_\lambda(\mathbf{S}) \stackrel{\text{def}}{=} \mathbf{U}_\mathbf{S}(\mathbf{D}_\mathbf{S} - (\lambda/L_{\nabla g})\mathbf{I}_{p \times m})_+ \mathbf{V}_\mathbf{S}^\top$, where SVD implies $\mathbf{S} = \mathbf{U}_\mathbf{S}\mathbf{D}_\mathbf{S}\mathbf{V}_\mathbf{S}^\top$, $\mathbf{I}_{p \times m}$ is a rectangular identity matrix with main diagonal elements equal to 1, and $(\mathbf{S})_+ = (\max\{0, s_{ij}\})$.

Theorem 2.1 (Bounds for loss difference and convergence rate in Algorithm 1). *Let $\{\mathbf{\Gamma}_t\}_{t=0}^T$ be the sequence obtained by the iteration of Algorithm 1. Then*

$$|F(\mathbf{\Gamma}_t) - F(\hat{\mathbf{\Gamma}})| \leq \frac{4(mn)^{-1} \max(\tau, 1 - \tau) \|\mathbf{X}\|_F^2 \|\mathbf{\Gamma}_0 - \hat{\mathbf{\Gamma}}\|_F^2}{(t + 1)^2}. \quad (2.11)$$

If for $\epsilon > 0$, $|F(\mathbf{\Gamma}_t) - F(\hat{\mathbf{\Gamma}})| \leq \epsilon$, then

$$t \geq \frac{2\sqrt{\max(\tau, 1 - \tau)} \|\mathbf{X}\|_F \|\mathbf{\Gamma}_0 - \hat{\mathbf{\Gamma}}\|_F}{\sqrt{mn\epsilon}} - 1. \quad (2.12)$$

The bound (2.11) comes from an explicit calculation of the Lipschitz constant of the gradient of g . The proof of Theorem 2.1 can be found in Appendix A.1.

Theorem 2.1 shows the convergence rate in our model is $\mathcal{O}(1/\sqrt{\epsilon})$, which is better than $\mathcal{O}(1/\epsilon)$ by quantile regression and $\mathcal{O}(1/\epsilon^2)$ by general subgradient method, see Theorem 3.2 and Remark 3.1 in Chao et al. (2015). In view of (2.15), when τ is approaching 0 or 1, the number of iteration that is required to achieve an ϵ -solution would increase.

Furthermore, utilizing the strong convexity of g , we can obtain a bound for $\|\mathbf{\Gamma}_t - \hat{\mathbf{\Gamma}}\|_F^2$. For this purpose, additional assumption on the design $\mathbf{X}$ is required.

(A1) Suppose $\mathbf{E} \mathbf{X} = 0$, $\mathbf{E} \mathbf{X} \mathbf{X}^\top = \mathbf{\Sigma}$ with $\sigma_{\min}(\mathbf{\Sigma}) > 0$ and $\sigma_{\max}(\mathbf{\Sigma}) < \infty$. for some sequence $0 < a_n < 1$, constants $c_1, c_2 > 0$,

$$\mathbf{P} \left[\sigma_{\min} \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} \right) \geq c_1 \sigma_{\min}(\mathbf{\Sigma}), \sigma_{\max} \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} \right) \leq c_2 \sigma_{\max}(\mathbf{\Sigma}) \right] \geq 1 - a_n. \quad (2.13)$$

Assumption (A1) holds for Gaussian design $\mathbf{X}$ with $c_1 = 1/9$, $c_2 = 9$ and $a_n = 4 \exp(-n/2)$. See ?.

Theorem 2.2. *Given (A1), the sequence $\mathbf{\Gamma}_t$ obtained Algorithm 1 satisfies*

$$\|\mathbf{\Gamma}_t - \hat{\mathbf{\Gamma}}\|_{\text{F}}^2 \leq \frac{36}{n(t+1)^2} \frac{\max(\tau, 1-\tau)}{\min(\tau, 1-\tau)} \frac{\|\mathbf{X}\|_{\text{F}}^2}{\sigma_{\min}(\mathbf{\Sigma})} \|\mathbf{\Gamma}_0 - \hat{\mathbf{\Gamma}}\|_{\text{F}}^2, \quad (2.14)$$

with probability greater than $1 - a_n$. If for $\epsilon > 0$, $\|\mathbf{\Gamma}_t - \hat{\mathbf{\Gamma}}\|_{\text{F}}^2 \leq \epsilon$, then

$$t \geq 6 \sqrt{\frac{\max(\tau, 1-\tau)}{\min(\tau, 1-\tau)} \frac{\|\mathbf{X}\|_{\text{F}} \|\mathbf{\Gamma}_0 - \hat{\mathbf{\Gamma}}\|_{\text{F}}}{\sqrt{n \sigma_{\min}(\mathbf{\Sigma})} \epsilon}} - 1. \quad (2.15)$$

The proof of Theorem 2.2 is in Section A.2.

2.3 Oracle Inequalities

In this section, we derive bounds for the sequence generated by Algorithm 1 $\mathbf{\Gamma}_t$ and the true matrix $\mathbf{\Gamma}$. These results heavily rely on the strong convexity of ρ_τ . The nuclear norm is decomposable with respect to two appropriately chosen subspaces in the sense that

$$\mathcal{R}(\mathbf{\Gamma} + \mathbf{\Delta}) = \mathcal{R}(\mathbf{\Gamma}) + \mathcal{R}(\mathbf{\Delta}), \forall \mathbf{\Gamma} \in \mathcal{M}, \mathbf{\Delta} \in \overline{\mathcal{M}}^\perp, \quad (2.16)$$

where

$$\begin{aligned} \mathcal{M}(U, V) &= \{\mathbf{\Theta} \in \mathbb{R}^{p \times m} \mid \text{row}(\mathbf{\Theta}) \subseteq U, \text{col}(\mathbf{\Theta}) \subseteq V\}, \\ \overline{\mathcal{M}}^\perp(U, V) &= \{\mathbf{\Theta} \in \mathbb{R}^{p \times m} \mid \text{row}(\mathbf{\Theta}) \subseteq U^\perp, \text{col}(\mathbf{\Theta}) \subseteq V^\perp\}, \end{aligned} \quad (2.17)$$

where U and V are two subspaces $U \subseteq \mathbb{R}^p$ and $V \subseteq \mathbb{R}^m$, represent the left and right singular vectors of the target matrix $\mathbf{\Gamma}$ respectively, $\text{row}(\mathbf{\Theta})$ and $\text{col}(\mathbf{\Theta})$ denote the row and column spaces of $\mathbf{\Theta}$.

We make the following assumptions.

(A2) There exists $c > 0$ such that for $u_{ij} \stackrel{\text{def}}{=} Y_{ij} - \mathbf{X}_i^\top \mathbf{\Gamma}_{\cdot j}$, $\mathbb{P}(|u_{ij}| > s) \leq \exp(1 - s^2/c^2)$, $\forall s \geq 0$ with sub-gaussian norm $\|u_{ij}\|_{\psi_2} \stackrel{\text{def}}{=} \sup_{p \geq 1} p^{-1/2} (\mathbb{E} |u_{ij}|^p)^{1/p}$, and let $K_u \stackrel{\text{def}}{=} \max_{1 \leq j \leq m} \|u_{ij}\|_{\psi_2}$.

(A3) Conditional on $\mathbf{X}_i$, u_{ij} are independent from $\mathbf{X}_i$ and independent over j .

(A2) regulates the tail of Y_{ij} . (A3) is required for obtaining bounds on tail probabilities

that are important for our main theorem. However, this assumption can be restrictive in practice.

Theorem 2.3. *Under (A1)-(A3), $\lambda = 2cm^{-1} \max(\tau, 1-\tau)K_u \sqrt{\|\Sigma\|} \sqrt{\frac{p+m}{n}}$, the sequence Γ_t obtained by Algorithm 1 satisfies*

$$\|\Gamma_t - \Gamma\|_F^2 \leq c'' \{R_t/n + 1\} \left\{ \frac{p+m}{n} \zeta_\tau^2 \dim(\overline{\mathcal{M}}) + \sqrt{\frac{p+m}{n}} \zeta_\tau \|\Gamma_{\mathcal{M}^\perp}\|_* \right\} + \frac{c'' R_t}{n} \|\Gamma_0 - \Gamma\|_F^2, \quad (2.18)$$

with probability greater than $1 - 3 \cdot 8^{-(p+m)} - a_n$, where $c'' > 0$ is an absolute constant, $R_t \stackrel{\text{def}}{=} \frac{1}{(t+1)^2} \frac{\eta_\tau \|\mathbf{X}\|_F^2}{\sigma_{\min}(\Sigma)}$, $\eta_\tau \stackrel{\text{def}}{=} \frac{\max(\tau, 1-\tau)}{\min(\tau, 1-\tau)}$, $\zeta_\tau \stackrel{\text{def}}{=} \frac{\eta_\tau \sqrt{\|\Sigma\|}}{\sigma_{\min}(\Sigma)} K_u$ and $\Gamma_{\mathcal{M}^\perp} \stackrel{\text{def}}{=} \arg \min_{\mathbf{Z} \in \mathcal{M}^\perp} \|\mathbf{Z} - \Gamma\|_F$.

The optimal bound is obtained by minimizing the right hand side of (2.18) with respect to all pairs $(\mathcal{M}, \overline{\mathcal{M}}^\perp)$, which will also balance $\dim(\overline{\mathcal{M}})$ and $\|\Gamma_{\mathcal{M}^\perp}\|_*$. When holding all other quantities fixed, as long as $p+m$ increases slower than n , the the right hand side of (2.18) goes to 0 as n tends to infinity. The quantity R_t characterizes how computational cost enters the oracle bound. We can increase the number of iteration in Algorithm 1 to shrink R_t , but this also increases the computational cost. Similar to Theorem 2.1 and B.1, when τ is approaching to the boundary of $(0, 1)$, the upper bounds will increase. Furthermore, heavier tail for Y_{ij} makes higher K_u , and leads to weaker error bounds.

Remark 2.1. As explained in Section 2.1, we estimate $\psi_{j,t}(\tau)$ by $\mathbf{V}_{j,t}$ in the SVD $\Gamma_t = \mathbf{U}_t \mathbf{D}_t \mathbf{V}_t^\top$. By Theorem 3.10 of Chao et al. (2015), we have $\psi_{j,t}(\tau)$:

$$1 - |\psi_j^\top(\tau) \psi_{j,t}(\tau)| \leq \frac{2(2\|\Gamma\| + \|\Gamma_t - \Gamma\|_F) \|\Gamma_t - \Gamma\|_F}{\min \{ \sigma_{j-1}^2(\Gamma) - \sigma_j^2(\Gamma), \sigma_j^2(\Gamma) - \sigma_{j+1}^2(\Gamma) \}}, \quad (2.19)$$

where $\widehat{\psi}_j(\tau)$ is the true loadings. Theorem 2.3 can be used with (2.19) to get an explicit bound.

3 Empirical Analysis: Predicting Risk Attitude with fMRI Data

In this section, we apply FASTEC on fMRI data to predict the risk attitude of humans on investment decisions. How human's brain responds to reward and risk is an ongoing research topic in neuropsychology, financial economics and neuroeconomics (Heekeren et al., 2008; Camerer, 2007; Schultz, 2015). Previous research mainly focuses on identifying the region of interest (ROI) using significantly positive Blood Oxygenation Level Dependent (BOLD) signal (see Schultz (2015) and the references therein). However, only

a few research uses fMRI BOLD on predicting the risk attitude of a subject or even future actions. Helfinstein et al. (2014) train support vector machines with fMRI BOLD recorded in a Ballon Analog Risk Task (BART) on several combinations of brain regions, and this classifier can predict subjects' next choice with over 70% accuracy. Majer et al. (2015) and van Bömmel et al. (2014) retrieve factor loadings from a dynamic model and apply these loadings on predicting subjects' risk attitude.

In our empirical analysis, we focus on predicting the subjects' risk attitude using the fMRI responses, but we differ from previous study in that we separately analyze the *positive* and *negative* fMRI BOLD signal observed in the cortical regions. The positive BOLD signal is known to be closely associated with increased neuronal activities, but the interpretation of large *negative* BOLD response (NBR) is still controversial. Mullinger et al. (2014) argue that the best explanation for NBR at the cortical layer might be a decrease in cerebral blood flow (CBF) with a lesser reduction in the neuronal activity, which is measured by the cerebral metabolic rate of oxygen consumption (CMRO₂). This explanation is proven to be an important complement or even a more plausible explanation than the more classical blood/vascular stealing hypothesis (see the references cited by Mullinger et al. (2014)). However, Mullinger et al. (2014) also argue that there may exist deeper neuronal reasons for NBR than simply inversion of the neurovascular coupling mechanism of positive BOLD response. Following the interpretation of NBR of Mullinger et al. (2014), we suspect that NBR also contains information for predicting the risk attitude. Using our expectile based approach, we are able to use the positive and negative BOLD response information in a very specific way.

3.1 Data

Our data come from a rapid event-related design experiment on investment decision, and this data set is firstly analyzed in Majer et al. (2015). The experiment is done as follows: 19 subjects were requested to make choices in 256 investment decision tasks and each task lasts 7 seconds. The fMRI is taken every two seconds, and there are 1400 images for each subject. We have also acquired the answer for each task from each subject. Majer et al. (2015) identify three brain regions Anterior insula (left and right aINS) and dorsomedial prefrontal cortex (DMPFC) via spectral clustering method. We will only focus on the BOLD response of the voxels in these three regions.

We integrate the information of each region (left and right aINS and DMPFC) spatially by taking quantile of the BOLD response over all voxels. At each fMRI scan i of sth subject, we take quantile with levels $\omega \in \{0.1, 0.5, 0.9\}$ of BOLD response over all voxels in the regions $b = 1$ (aINS_L), $b = 2$ (aINS_R) and $b = 3$ (DMPFC) to construct a

single time series $\nu_i(s, b, \omega)$, where $i = 1, \dots, N = 1400$. Figure 3.1 gives an illustration of the BOLD time series of each cluster. For each cluster, the series of 19 subjects at ω are averaged (the solid lines) and the band shows the dispersion of the 19 time series in ω . We observe that the series for $\omega = 0.9$ is largely positive, which summarize the information of positive BOLD response, while the series for $\omega = 0.1$ is mainly negative, which corresponds to the negative BOLD response. The series for $\omega = 0.5$ is stationary and varying around the origin.

3.2 Method

3.2.1 Factor loadings at each region b and quantile level ω

There are many ways to define Y_{ij} using BOLD series, and this can have big impact to predictive performance. For each ω and a single region b , we consider two approaches to obtain the variable Y_{ij} :

- (C1) "Whole time series": set $Y_{ij}^{b,\omega} = \nu_i(j, b, \omega)$, where $i = 1, \dots, n$ with $n = N$, $j = 1, \dots, 19$ (subject). Thus, we have $m = 19$ curves in each region b and at each quantile ω .
- (C2) "Task-wise" perspective: we divide the whole time series in each region b and at each quantile level ω into subseries based on the the start and the end of each task. Let $\mathcal{I}_q \in \{1, \dots, N\}$ be the set which contains the index of the images taken during the q th task. In our data, $\mathcal{I}_q$ usually contains 3-4 components. We interpolate the points $\{\nu_i(s, b, \omega)\}_{i \in \mathcal{I}_q}$ for each fixed s, b , and ω . Denote the value on the interpolated curve at i th point in n equally distant grid on the interval $(\min(\mathcal{I}_q), \max(\mathcal{I}_q))$ by $\tilde{\nu}_i(s, b, q, \omega)$, where $i = 1, \dots, n = 50$. Let $Y_{ij}^{b,\omega} = \tilde{\nu}_i(s, b, q, \omega)$ with $j = 256(s-1) + q$, where $s = 1, \dots, 19$ (index for subject) and $q = 1, \dots, 256$ (index for tasks) for each ω, b . Thus, there are $m = 19 * 256 = 4864$ curves in each b and ω .

The variable $\mathbf{X}_i$ needs to be flexible enough to capture the shape of the fMRI sequence. For this purpose, we use cubic B -spline basis $\{B_k\}_{k=1}^p$ with regularly spaced knots on $[0, 1]$, and set $\mathbf{X}_i = (B_1(i/n), B_2(i/n), \dots, B_p(i/n))^\top$, where $i = 1, \dots, n$ and n is subject to which approach is taken. B -splines have nice computational properties for estimating the hemodynamic response, see Degras and Lindquist (2014) for more detail. We select $p = \lceil n^{0.8} \rceil$ of basis functions in each approach above, where $\lceil \cdot \rceil$ takes the smallest integer that is greater than the argument. The power 0.8 is greater than the optimal rate 0.4, because the nuclear norm penalty potentially reduces the overfitting. As the result, there are 329 basis functions in the approach (C1) and 23 in (C2).

We compute the matrix $\widehat{\mathbf{\Gamma}}^{b,\omega}$ with expectile level $\tau = 0.1, 0.5, 0.9$ using Y_{ij} and $\mathbf{X}_i$ by Algorithm 1, where Y_{ij} is chosen under either **(C1)** or **(C2)**. We select $\lambda^{b,\omega}$ by five fold cross-validation. To be more specific, we divide the whole sample into 5 groups along $i = 1, \dots, N$, e.g., under **(C1)** each group with 280 observations would be held out as the validation group in turns. About more detailed results in the determination of tuning parameters, we refer to Appendices D.2. Applying Algorithm 1 with the selected λ , we obtain $\widehat{\mathbf{\Gamma}}^{b,\omega}$. Using SVD $\widehat{\mathbf{\Gamma}}^{b,\omega} = \widehat{\mathbf{U}}_\tau^{b,\omega} \widehat{\mathbf{D}}_\tau^{b,\omega} (\widehat{\mathbf{V}}_\tau^{b,\omega})^\top$, where $(\widehat{\mathbf{V}}_\tau^{b,\omega})^\top$ is regarded as factor loadings. We note that the size of matrix $\widehat{\mathbf{V}}_\tau^{b,\omega}$ is 19×19 if we define $Y_{ij}^{b,\omega}$ by following **(C1)** and 4864×4864 by following **(C2)**. Note that the sign of the factor loadings cannot be determined exactly.

3.2.2 Predicting risk attitude

To measure the predictive performance, we need to estimate the the subjects' "oracle" risk attitude β_s , where $s = 1, \dots, 19$ denotes the subject. We follow the approach of Majer et al. (2015) and estimate β_s by the answer given by the subjects to each task with logistic regression. In essence, higher β_s means the subject s is less risk-averse. More on the computation of β_s is provided in Appendices D.1.

In order to use the loadings $\widehat{\mathbf{V}}_\tau^{b,\omega}$ to predict β_s , we apply standard linear regression models. In particular, in the case **(C1)**, we construct a model for β_s using the first two factor loadings

$$\begin{aligned} \beta_s = & \alpha_0^{\omega,\tau} + \alpha_{11}^{\omega,\tau} |(\widehat{\mathbf{V}}_\tau^{1,\omega})_{s1}| + \alpha_{12}^{\omega,\tau} |(\widehat{\mathbf{V}}_\tau^{2,\omega})_{s1}| + \alpha_{13}^{\omega,\tau} |(\widehat{\mathbf{V}}_\tau^{3,\omega})_{s1}| \\ & + \alpha_{21}^{\omega,\tau} |(\widehat{\mathbf{V}}_\tau^{1,\omega})_{s2}| + \alpha_{22}^{\omega,\tau} |(\widehat{\mathbf{V}}_\tau^{2,\omega})_{s2}| + \alpha_{23}^{\omega,\tau} |(\widehat{\mathbf{V}}_\tau^{3,\omega})_{s2}| + \varepsilon_s, \quad s = 1, \dots, 19, \end{aligned} \quad (3.1)$$

where $\{\alpha_0^{\omega,\tau}, \alpha_{11}^{\omega,\tau}, \alpha_{12}^{\omega,\tau}, \alpha_{13}^{\omega,\tau}, \alpha_{21}^{\omega,\tau}, \alpha_{22}^{\omega,\tau}, \alpha_{23}^{\omega,\tau}\} \in \mathbb{R}^7$ are the intercept and the coefficients associated with the regions left and right Anterior insula, and dorsomedial prefrontal cortex. In the case **(C2)**, define the averaged loadings of all tasks for each s

$$\mu_s^{b,\omega,\tau} \stackrel{\text{def}}{=} \frac{1}{256} \sum_{q=1}^{256} |(\widehat{\mathbf{V}}_\tau^{b,\omega})_{256(s-1)+q,1}|.$$

We construct another model for β_s using $\mu_s^{b,\omega,\tau}$:

$$\begin{aligned} \beta_s = & \bar{\alpha}_0^{\omega,\tau} + \bar{\alpha}_{11}^{\omega,\tau} \mu_s^{1,\omega,\tau} + \bar{\alpha}_{12}^{\omega,\tau} \mu_s^{2,\omega,\tau} + \bar{\alpha}_{13}^{\omega,\tau} \mu_s^{3,\omega,\tau} \\ & + \bar{\alpha}_{21}^{\omega,\tau} \mu_s^{1,\omega,\tau} + \bar{\alpha}_{22}^{\omega,\tau} \mu_s^{2,\omega,\tau} + \bar{\alpha}_{23}^{\omega,\tau} \mu_s^{3,\omega,\tau} + \varepsilon_s, \quad s = 1, \dots, 19, \end{aligned} \quad (3.2)$$

where $\{\bar{\alpha}_0^{\omega,\tau}, \bar{\alpha}_{11}^{\omega,\tau}, \bar{\alpha}_{12}^{\omega,\tau}, \bar{\alpha}_{13}^{\omega,\tau}, \bar{\alpha}_{21}^{\omega,\tau}, \bar{\alpha}_{22}^{\omega,\tau}, \bar{\alpha}_{23}^{\omega,\tau}\} \in \mathbb{R}^7$. We take the absolute value of the loadings $\widehat{\mathbf{V}}_\tau^{b,\omega}$ because we are only interested in the magnitude of the loadings.

3.2.3 In-sample and out-of-sample performance

To compare model (3.1) and (3.2), we show their in-sample and out-of-sample performance. For in-sample performance, R^2 of both regression (3.1) and (3.2) is computed. In addition, in order to determine whether (3.1) and (3.2) correctly predict the *order* of risk-aversion of the subjects (rather than the exact value of β_s), we calculate Spearman's and Kendall's rank correlation between the fitted $\hat{\beta}_s$ (in-sample) and β_s .

To measure the out-of-sample performance, we calculate $\{\tilde{\beta}_s\}_{s=1}^{19}$ by leave-one-out algorithm. The steps are as below:

- (1) Fix s , where $s = 1, \dots, 19$. Use the values of the remaining 18 subjects to find the coefficients $\{\alpha_0^{\omega,\tau}, \alpha_{11}^{\omega,\tau}, \alpha_{12}^{\omega,\tau}, \alpha_{13}^{\omega,\tau}, \alpha_{21}^{\omega,\tau}, \alpha_{22}^{\omega,\tau}, \alpha_{23}^{\omega,\tau}\}$ in model (3.1) and $\{\bar{\alpha}_0^{\omega,\tau}, \bar{\alpha}_{11}^{\omega,\tau}, \bar{\alpha}_{12}^{\omega,\tau}, \bar{\alpha}_{13}^{\omega,\tau}, \bar{\alpha}_{21}^{\omega,\tau}, \bar{\alpha}_{22}^{\omega,\tau}, \bar{\alpha}_{23}^{\omega,\tau}\}$ in model (3.2) by standard linear regression.
- (2) Compute $\tilde{\beta}_s$ by the trained models (3.1) and (3.2).
- (3) Repeat steps (1) and (2) for each $s = 1, \dots, 19$.
- (4) Calculate the Spearman's correlation and Kendall's rank correlation between $\{\tilde{\beta}_s\}_{s=1}^{19}$ and $\{\beta_s\}_{s=1}^{19}$.

3.3 Results

In Table 3.1, we present the in-sample fitting and out-of-sample performance for models (3.1) and (3.2) with the constrained model that uses only the 1st factor ($\alpha_{21}^{\omega,\tau} = \alpha_{22}^{\omega,\tau} = \alpha_{23}^{\omega,\tau} = 0$ in (3.1) and $\bar{\alpha}_{21}^{\omega,\tau} = \bar{\alpha}_{22}^{\omega,\tau} = \bar{\alpha}_{23}^{\omega,\tau} = 0$ in (3.2)) and the whole model, under various (τ, ω) combinations.

For the in-sample fitting results, cases with $\omega = 0.1$ and $\omega = 0.9$ perform much better than $\omega = 0.5$. This shows that both negative or positive BOLD can lead to good model fitting, which suggests that negative BOLD may also explain the variation of risk attitude well. In particular, the level τ that are closer to the maximum of the curves of $\omega = 0.9$ and to the minimum of the of the curves of $\omega = 0.9$, which is consistent with our prior belief from Figure 3.1. Moreover, task-wise curves seem to perform better than the whole series.

For the out-of-sample performance, the constrained model (3.2) with the negative BOLD ($\omega = 0.1, \tau = 0.1$) nearly always outperforms all other cases. In contrast, positive BOLD ($\omega = 0.9$) under the same model performs poorly. This provides a new evidence that negative BOLD may be more relevant than the positive BOLD for predicting the risk

attitude. Moreover, the unconstrained model improves the prediction performance in most cases, particularly for the prediction by unconstrained (3.2) under $\omega = 0.9$ and higher τ levels.

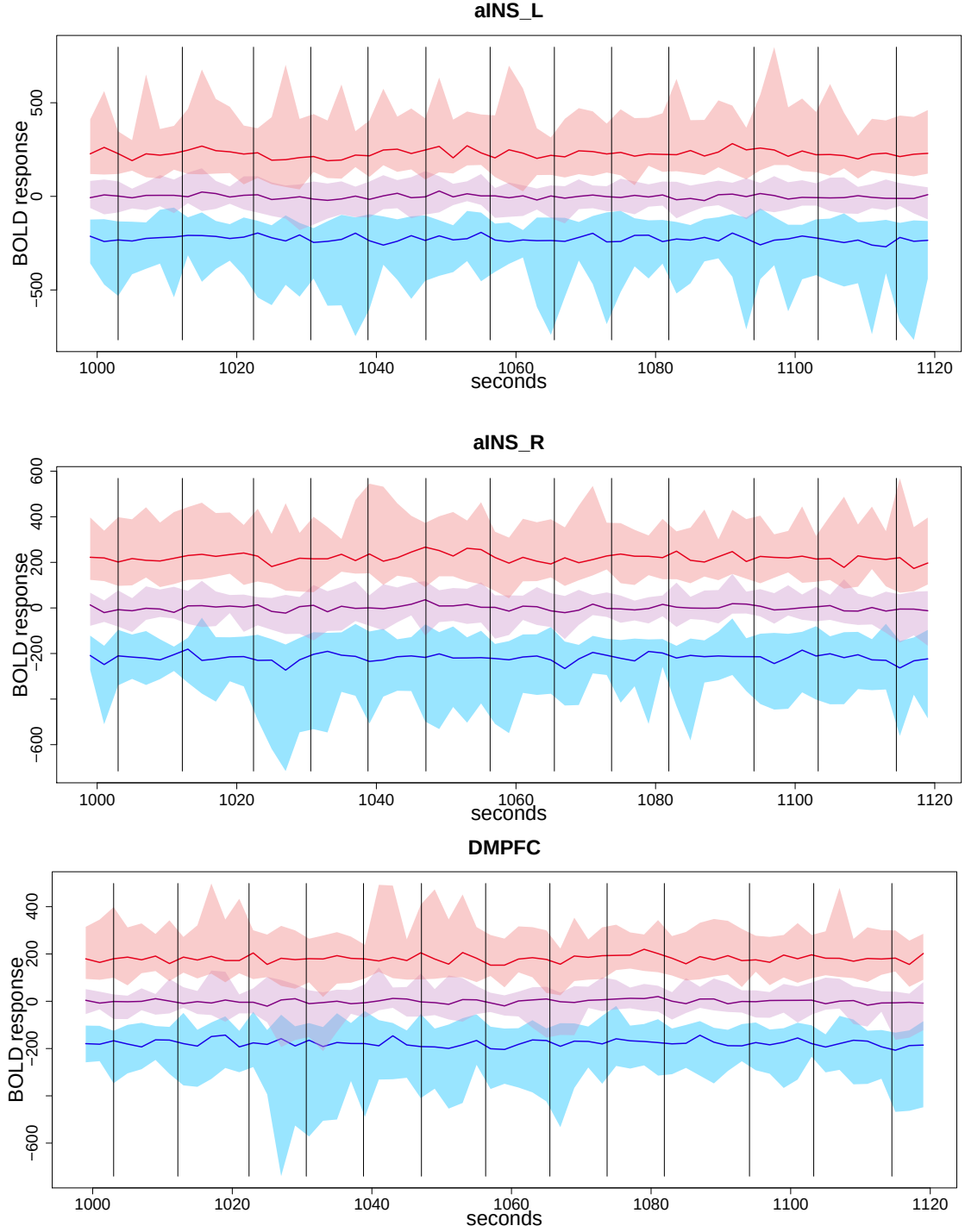


Figure 3.1: In each region, the ω quantiles of the BOLD response over all the voxels between 1000-1120 seconds of the experiment is shown. In each subfigure (region), lowest (resp., middle, highest) solid lines represent the median of $\omega = 0.1$ (resp., $\omega = 0.5, 0.9$) quantiles of all 19 subjects, and the upper and lower boundaries of the bands present the maximum and the minimum of the ω quantiles of the 19 subjects. Vertical lines indicate the occurrence of stimuli.

		Constrained model				Unconstrained model								
		Whole series (3.1)			Task-wise (3.2)			Whole series (3.1)			Task-wise (3.2)			
τ		0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	
		In-sample fitting												
$\omega=0.1$	R^2	0.084	0.158	0.101	0.412	0.412	0.412	0.413	0.312	0.263	0.226	0.455	0.454	0.454
	Spearman's rank corr	0.149	0.377	0.328	0.595	0.595	0.604	0.532	0.526	0.396	0.618	0.618	0.618	
	Kendall's rank corr	0.076	0.263	0.228	0.462	0.462	0.474	0.333	0.357	0.275	0.474	0.474	0.474	
$\omega=0.5$	R^2	0.070	0.043	0.030	0.134	0.136	0.135	0.307	0.260	0.352	0.445	0.440	0.441	
	Spearman's rank corr	0.177	0.140	0.226	0.335	0.316	0.326	0.547	0.528	0.596	0.533	0.544	0.544	
	Kendall's rank corr	0.135	0.088	0.135	0.205	0.193	0.205	0.427	0.333	0.415	0.368	0.380	0.380	
$\omega=0.9$	R^2	0.199	0.238	0.148	0.206	0.205	0.205	0.393	0.367	0.229	0.487	0.496	0.500	
	Spearman's rank corr	0.435	0.540	0.181	0.412	0.412	0.412	0.588	0.628	0.582	0.596	0.637	0.637	
	Kendall's rank corr	0.333	0.391	0.135	0.298	0.298	0.298	0.439	0.439	0.439	0.462	0.497	0.497	
Out-of-sample predicting														
$\omega=0.1$	Spearman's rank corr	-0.453	-0.181	-0.321	0.454	0.451	0.440	-0.079	-0.133	0.072	0.298	0.298	0.298	
	Kendall's rank corr	-0.322	-0.111	-0.240	0.357	0.345	0.345	-0.076	-0.088	0.041	0.216	0.216	0.216	
$\omega=0.5$	Spearman's rank corr	-0.444	-0.700	-0.658	-0.119	-0.119	-0.119	-0.035	-0.196	0.247	0.205	0.204	0.212	
	Kendall's rank corr	-0.275	-0.509	-0.450	-0.064	-0.064	-0.064	-0.006	-0.146	0.135	0.123	0.111	0.123	
$\omega=0.9$	Spearman's rank corr	-0.207	0.204	-0.493	0.023	0.023	0.023	0.161	0.072	-0.447	0.293	0.307	0.307	
	Kendall's rank corr	-0.170	0.135	-0.345	0.006	0.006	0.006	0.076	0.041	-0.298	0.205	0.216	0.216	

Table 3.1: The goodness of fit R^2 , Spearman's and Kendall's rank correlation from the in-sample fitting and out-of-sample prediction by (3.1) or (3.2) with/without constraints, under different τ , ω levels.

References

- Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences* **2**(1): 183–202.
- Bernardo, A. E. and Ledoit, O. (2000). Gain, loss, and asset pricing, *Journal of Political Economy* **108**(1): 144–172.
- Breckling, J. and Chambers, R. (1988). M -quantiles, *Biometrika* **74**(4): 761–772.
- Cabrera, B. L. and Schulz, F. (2016). Forecasting generalized quantiles of electricity demand: A functional data approach, *Journal of the American Statistical Association* (DOI: 10.1080/01621459.2016.1219259).
- Camerer, C. F. (2007). Neuroeconomics: Using neuroscience to make economic predictions, *The Economic Journal* **117**(519): C26–C42.
- Chao, S.-K., Härdle, W. K. and Yuan, M. (2015). Factorisable sparse tail event curves, *SFB 649 Discussion Paper 2015-034*, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. Available at <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2015-034.pdf>.
- Chao, S.-K., Proksch, K., Dette, H. and Härdle, W. K. (2016). Confidence corridors for multivariate generalized quantile regression, *Journal of Business & Economic Statistics* (DOI: 10.1080/07350015.2015.1054493).
- Degras, D. and Lindquist, M. A. (2014). A hierarchical model for simultaneous detection and estimation in multi-subject fMRI studies, *NeuroImage* **98**: 61–72.
- Efron, B. (1991). Regression percentiles using asymmetric squared error loss, *Statistica Sinica* **1**(1): 93–125.
- Fadili, J. M. and Peyré, G. (2011). Total variation projection with first order schemes, *Image Processing, IEEE Transactions on* **20**(3): 657–669.
- Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis, *The Annals of Statistics* **34**(3): 1493–1517.
- Heekeren, H. R., Marrett, S. and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making, *Nature Reviews Neuroscience* **9**(6): 467–479.

- Helmfstein, S. M., Schonberg, T., Congdon, E., Karlsgodt, K. H., Mumford, J. A., Sabb, F. W., Cannon, T. D., London, E. D., Bilder, R. M. and Poldrack, R. A. (2014). Predicting risky choices from brain activity patterns, *Proceedings of the National Academy of Sciences* **111**(7): 2470–2475.
- Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model, *Journal of Multivariate Analysis* **5**(2): 248–264.
- Ji, S. and Ye, J. (2009). An accelerated gradient method for trace norm minimization, *Proceedings of the 26th International Conference on Machine Learning* .
- Koenker, R. and Bassett, G. W. (1978). Regression quantiles, *Econometrica* **46**(1): 33–50.
- Kuan, C.-M., Yen, J.-H. and Hsu, Y.-C. (2009). Assessing value at risk with care, the conditional autoregressive expectile models, *Journal of Econometrics* **150**(2): 261–270.
- Majer, P., Mohr, P. N. C., Heekeren, H. and Härdle, W. K. (2015). Portfolio decisions and brain reactions via the CEAD method, *Psychometrika* **81**(3): 881–903.
- Mullinger, K., Mayhew, S., Bagshaw, A., Bowtell, R. and Francis, S. (2014). Evidence that the negative BOLD response is neuronal in origin: A simultaneous EEG–BOLD–CBF study in humans, *NeuroImage* **94**: 263 – 274.
- Negahban, S. N., Ravikumar, P., Wainwright, M. J. and Yu, B. (2012). A unified framework for high-dimensional analysis of M -estimators with decomposable regularizers, *Statistical Science* **27**(4): 538–557.
- Negahban, S. N. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling, *The Annals of Statistics* **39**(2): 1069–1097.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing, *Econometrica* **55**(4): 819–847.
- Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: methods and case studies*, Vol. 77, Springer, New York.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, Series in Statistics, 2nd edn, Springer, New York.
- Reinsel, G. C. and Velu, R. P. (1998). *Multivariate Reduced-Rank Regression*, Springer, New York.
- Schultz, W. (2015). Neuronal reward and decision signals: From theories to data, *Physiological Reviews* **95**(3): 853–951.

- Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications, *Statistica Neerlandica* **56**(2): 214–232.
- Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics* **6**(2): 231–252.
- Tran, N. M., Burdejova, P., Osipenko, M. and Härdle, W. K. (2016). Principal component analysis in an asymmetric norm, *SFB 649 Discussion Paper 2016-040*, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. Available at <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2016-040.pdf>.
- van Bömmel, A., Song, S., Majer, P., Mohr, P. N. C., Heekeren, H. R. and Härdle, W. K. (2014). Risk patterns and correlated brain activities. multidimensional statistical analysis of fmri data in economic decision making study, *Psychometrika* **79**(3): 489–514.
- Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices, in Y. Eldar and G. Kutyniok (eds), *Compressed Sensing, Theory and Applications*, Cambridge University Press, chapter 5, pp. 210–268.
- Xu, X., Mihoci, A. and Härdle, W. K. (2015). ICARE - localizing conditional autoregressive expectiles, *SFB 649 Discussion Paper 2015-052*, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. Available at <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2015-052.pdf>.
- Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A. and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate, *Biometrics* **59**(3): 676–685.
- Yuan, M., Ekici, A., Lu, Z. and Monteiro, R. (2007). Dimension reduction and coefficient estimation in multivariate linear regression, *Journal of the Royal Statistical Society: Series B* **69**(3): 329–346.
- Ziegel, J. F. (2016). Coherence and elicibility, *Mathematical Finance* **26**(4): 901–918.

APPENDIX

APPENDIX A: Proofs for Section 2.2

A.1 Proof for Theorem 2.1

Theorem 4.4 in Beck and Teboulle (2009) gives the upper bound of the loss difference in the t -th step of the iteration by

$$|F(\mathbf{\Gamma}_t) - F(\widehat{\mathbf{\Gamma}})| \leq \frac{2L_{\nabla g}\|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2}{(t+1)^2}, \quad (\text{A.1})$$

where $L_{\nabla g}$ is the Lipschitz constant of $\nabla g(\mathbf{\Gamma})$ defined in (2.8).

We note that

$$\rho'_\tau(u) = \begin{cases} 2\tau u & \text{for } u \geq 0; \\ 2(1-\tau)u & \text{for } u < 0. \end{cases} \quad (\text{A.2})$$

Hence, the gradient is

$$\nabla g(\mathbf{\Gamma}) = -(mn)^{-1}\mathbf{X}^\top \{\mathbf{W} \circ (\mathbf{Y} - \mathbf{X}\mathbf{\Gamma})\}, \quad (\text{A.3})$$

where $\mathbf{W}(\mathbf{\Gamma}) = (w_{ij}) \in \mathbb{R}^{n \times m}$, $w_{ij} \stackrel{\text{def}}{=} 2\{\tau + \mathbf{1}(Y_{ij} \leq \mathbf{X}_i^\top \mathbf{\Gamma}_{\cdot j})(1-2\tau)\}$, " $\circ$ " represents the Hadamard product.

To simplify the notations, define $\mathbf{U}(\mathbf{\Gamma}) = (Y_{ij} - \mathbf{X}_i^\top \mathbf{\Gamma}_{\cdot j}) \in \mathbb{R}^{n \times m}$. For all $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2 \in \mathbb{R}^{p \times m}$, let $\mathbf{U}_1 = \mathbf{U}(\mathbf{\Gamma}_1)$, $\mathbf{U}_2 = \mathbf{U}(\mathbf{\Gamma}_2)$, $\mathbf{W}_1 = \mathbf{W}(\mathbf{\Gamma}_1)$ and $\mathbf{W}_2 = \mathbf{W}(\mathbf{\Gamma}_2)$.

$$\begin{aligned} \|\nabla g(\mathbf{\Gamma}_1) - \nabla g(\mathbf{\Gamma}_2)\|_{\text{F}} &= (mn)^{-1}\|\mathbf{X}^\top (\mathbf{W}_1 \circ \mathbf{U}_1) - \mathbf{X}^\top (\mathbf{W}_2 \circ \mathbf{U}_2)\|_{\text{F}} \\ &\leq (mn)^{-1}\|\mathbf{X}\|_{\text{F}}\|\mathbf{W}_1 \circ \mathbf{U}_1 - \mathbf{W}_2 \circ \mathbf{U}_2\|_{\text{F}} \quad (\text{by submultiplicity}) \\ &= (mn)^{-1}\|\mathbf{X}\|_{\text{F}}\left[\sum_{i=1}^n \sum_{j=1}^m \{\rho'_\tau(u_{1,ij}) - \rho'_\tau(u_{2,ij})\}^2\right]^{1/2} \\ &\leq (mn)^{-1}\|\mathbf{X}\|_{\text{F}}\left[\sum_{i=1}^n \sum_{j=1}^m \{2\max(\tau, 1-\tau)\}^2(u_{1,ij} - u_{2,ij})^2\right]^{1/2} \\ &= 2(mn)^{-1}\max(\tau, 1-\tau)\|\mathbf{X}\|_{\text{F}}\|\mathbf{Y} - \mathbf{X}\mathbf{\Gamma}_1 - (\mathbf{Y} - \mathbf{X}\mathbf{\Gamma}_2)\|_{\text{F}} \\ &\leq 2(mn)^{-1}\max(\tau, 1-\tau)\|\mathbf{X}\|_{\text{F}}^2\|\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2\|_{\text{F}} \quad (\text{by submultiplicity}), \end{aligned} \quad (\text{A.4})$$

where the fourth line makes use of the fact that $\rho'_\tau(u)$ is Lipschitz continuous with Lipschitz constant $2\max(\tau, 1-\tau)$, see Chao et al. (2016).

Plug $L_{\nabla g} = 2(mn)^{-1} \max(\tau, 1 - \tau) \|\mathbf{X}\|_F^2$ into (A.1) yields

$$|F(\mathbf{\Gamma}_t) - F(\widehat{\mathbf{\Gamma}})| \leq \frac{4(mn)^{-1} \max(\tau, 1 - \tau) \|\mathbf{X}\|_F^2 \|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}\|_F^2}{(t + 1)^2}. \quad (\text{A.5})$$

Moreover, setting the right hand side of (A.5) to be ϵ ($\forall \epsilon > 0$) and solving for t gives

$$t \geq \frac{2\sqrt{\max(\tau, 1 - \tau) \|\mathbf{X}\|_F \|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}\|_F}}{\sqrt{mn\epsilon}} - 1. \quad (\text{A.6})$$

□

A.2 Proof for Theorem 2.2

Following the proof of Theorem 1 in Fadili and Peyré (2011), define

$$I(\mathbf{\Gamma}_t) \stackrel{\text{def}}{=} g(\mathbf{\Gamma}_t) - g(\widehat{\mathbf{\Gamma}}) - \langle \nabla g(\mathbf{\Gamma}_t), \mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}} \rangle, \quad (\text{A.7})$$

$$J(\mathbf{\Gamma}_t) \stackrel{\text{def}}{=} h(\mathbf{\Gamma}_t) - h(\widehat{\mathbf{\Gamma}}) + \langle \nabla g(\mathbf{\Gamma}_t), \mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}} \rangle, \quad (\text{A.8})$$

the sum of them gives

$$I(\mathbf{\Gamma}_t) + J(\mathbf{\Gamma}_t) = F(\mathbf{\Gamma}_t) - F(\widehat{\mathbf{\Gamma}}). \quad (\text{A.9})$$

According to Lemma C.2, we have

$$\begin{aligned} I(\mathbf{\Gamma}_t) &\geq \kappa \|\mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}}\|_F^2 \\ &= \frac{1}{9} m^{-1} \min(\tau, 1 - \tau) \sigma_{\min}(\mathbf{\Sigma}) \|\mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}}\|_F^2 \end{aligned} \quad (\text{A.10})$$

where the second line holds with probability greater than $1 - a_n$ under (A1).

Since $\widehat{\mathbf{\Gamma}}$ is the optimizer of (2.5), therefore,

$$\mathbf{0} \in \nabla g(\widehat{\mathbf{\Gamma}}) + \nabla h(\widehat{\mathbf{\Gamma}}), \quad (\text{A.11})$$

which implies

$$-\nabla g(\widehat{\mathbf{\Gamma}}) \in \nabla h(\widehat{\mathbf{\Gamma}}). \quad (\text{A.12})$$

As a result, we have

$$h(\mathbf{\Gamma}_t) - h(\widehat{\mathbf{\Gamma}}) \geq \langle -\nabla g(\mathbf{\Gamma}_t), \mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}} \rangle, \quad (\text{A.13})$$

i.e., $J(\mathbf{\Gamma}_t) \geq 0$.

Plugging (A.10) and (A.13) into (A.9) yields,

$$\begin{aligned}\|\mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2 &\leq \frac{9m}{\min(\tau, 1 - \tau)\sigma_{\min}(\mathbf{\Sigma})} \{F(\mathbf{\Gamma}_t) - F(\widehat{\mathbf{\Gamma}})\} \\ &\leq \frac{36}{n(t+1)^2} \frac{\max(\tau, 1 - \tau)}{\min(\tau, 1 - \tau)} \frac{\|\mathbf{X}\|_{\text{F}}^2}{\sigma_{\min}(\mathbf{\Sigma})} \|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2,\end{aligned}\quad (\text{A.14})$$

with probability greater than $1 - a_n$. The second line comes from the result of Theorem 2.1. $\square$

APPENDIX B: Proofs for Theorem 2.3

By triangle inequality, we have

$$\|\mathbf{\Gamma}_t - \mathbf{\Gamma}\|_{\text{F}}^2 = \|\mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}} + \widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\text{F}}^2 \leq 2\|\mathbf{\Gamma}_t - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2 + 2\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\text{F}}^2. \quad (\text{B.1})$$

Combining the results of Lemma B.2 and Theorem 2.2, it follows that

$$\begin{aligned}\|\mathbf{\Gamma}_t - \mathbf{\Gamma}\|_{\text{F}}^2 &\leq 18^3 c^2 \frac{p+m}{n} \frac{\max(\tau, 1 - \tau)^2}{\min(\tau, 1 - \tau)^2} \frac{\|\mathbf{\Sigma}\|}{\sigma_{\min}(\mathbf{\Sigma})^2} K_u^2 \dim(\overline{\mathcal{M}}) \\ &\quad + 144c \sqrt{\frac{p+m}{n}} \frac{\max(\tau, 1 - \tau)}{\min(\tau, 1 - \tau)} \frac{\sqrt{\|\mathbf{\Sigma}\|}}{\sigma_{\min}(\mathbf{\Sigma})} K_u \|\mathbf{\Gamma}_{\mathcal{M}^\perp}\|_* \\ &\quad + \frac{72}{n(t+1)^2} \frac{\max(\tau, 1 - \tau)}{\min(\tau, 1 - \tau)} \frac{\|\mathbf{X}\|_{\text{F}}^2}{\sigma_{\min}(\mathbf{\Sigma})} \|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2,\end{aligned}\quad (\text{B.2})$$

holds with probability greater than $1 - 3 \times 8^{-(p+m)} - a_n$.

Furthermore, given

$$\|\mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2 = \|\mathbf{\Gamma}_0 - \mathbf{\Gamma} + \mathbf{\Gamma} - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2 \leq 2\|\mathbf{\Gamma}_0 - \mathbf{\Gamma}\|_{\text{F}}^2 + 2\|\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}}\|_{\text{F}}^2, \quad (\text{B.3})$$

and applying Lemma B.2 again we complete the proof of Theorem 2.3. $\square$

Now we show auxiliary results used in the proof of Theorem 2.3. The next theorem is an application of Theorem 1 of Negahban et al. (2012).

Theorem B.1 (Error bounds for the estimator). *Under (A1), any optimal solution $\widehat{\mathbf{\Gamma}}$ in the problem (2.5) with $\lambda \geq 2\|\nabla g(\mathbf{\Gamma})\|$ satisfies the bound*

$$\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\text{F}}^2 \leq \frac{9m^2\lambda^2}{\{c_1 \min(\tau, 1 - \tau)\sigma_{\min}(\mathbf{\Sigma})\}^2} \dim(\overline{\mathcal{M}}) + \frac{36m\lambda}{\min(\tau, 1 - \tau)\sigma_{\min}(\mathbf{\Sigma})} \|\mathbf{\Gamma}_{\mathcal{M}^\perp}\|_*, \quad (\text{B.4})$$

with probability greater than $1 - a_n$, where $\mathbf{\Gamma}_{\mathcal{M}^\perp} = \arg \min_{\mathbf{Z} \in \mathcal{M}^\perp} \|\mathbf{Z} - \mathbf{\Gamma}\|_{\text{F}}$.

Proof for Theorem B.1. The proof is an application of Theorem 1 of Negahban et al. (2012). We will verify its conditions (G1) and (G2). For condition (G1), we note that the nuclear norm $\|\cdot\|_*$ is decomposable with respect to $(\mathcal{M}, \overline{\mathcal{M}}^\perp)$ defined in (2.17). For condition (G2), note that on the event

$$\Omega_1 \stackrel{\text{def}}{=} \left\{ \sigma_{\min}\left(\frac{\mathbf{X}^\top \mathbf{X}}{n}\right) \geq c_1 \sigma_{\min}(\boldsymbol{\Sigma}), \sigma_{\max}\left(\frac{\mathbf{X}^\top \mathbf{X}}{n}\right) \leq c_2 \sigma_{\max}(\boldsymbol{\Sigma}) \right\}, \quad (\text{B.5})$$

the loss function g is restrictive strongly convex with coefficients κ and $\xi = 0$ (we replace $\tau_{\mathcal{L}}$ in Negahban et al. (2012) by ξ) shown in Lemma C.2. We note that the nuclear norm and the spectral norm of a matrix are dual, and their subspace compatibility constant $\Psi(\overline{\mathcal{M}}) \leq \dim(\overline{\mathcal{M}})$. $\square$

Lemma B.1. Under (A1)-(A3),

$$\mathbb{P} \left(\|\nabla g(\boldsymbol{\Gamma})\| \leq cm^{-1} \max(\tau, 1 - \tau) K_u \sqrt{\|\boldsymbol{\Sigma}\|} \sqrt{\frac{p+m}{n}} \right) \geq 1 - 3 \times 8^{-(p+m)} - a_n, \quad (\text{B.6})$$

where $c > 0$ is an absolute constant.

Proof for Lemma B.1. Throughout the proof, we restrict on the event Ω_1 in (B.5). Recall the expression from (A.3) that

$$\nabla g(\boldsymbol{\Gamma}) = -(mn)^{-1} \mathbf{X}^\top \{\mathbf{W} \circ (\mathbf{Y} - \mathbf{X}\boldsymbol{\Gamma})\}.$$

and the matrix $\mathbf{U}(\boldsymbol{\Gamma}) = (u_{ij}) = (Y_{ij} - \mathbf{X}_i^\top \boldsymbol{\Gamma}_{\cdot j}) \in \mathbb{R}^{n \times m}$. Following the proof of Lemma 3 in Negahban and Wainwright (2011), we have

$$\begin{aligned} \mathbb{P} \left(n^{-1} \|\mathbf{X}^\top (\mathbf{W} \circ \mathbf{U})\| \geq 4s \right) &= \mathbb{P} \left(\sup_{\substack{\boldsymbol{\beta} \in \mathcal{S}^{p-1}, \\ \boldsymbol{\alpha} \in \mathcal{S}^{m-1}}} n^{-1} |\boldsymbol{\beta}^\top \mathbf{X}^\top (\mathbf{W} \circ \mathbf{U}) \boldsymbol{\alpha}| \geq 4s \right) \\ &\leq 8^{p+m} \sup_{\substack{\boldsymbol{\beta} \in \mathcal{S}^{p-1}, \\ \boldsymbol{\alpha} \in \mathcal{S}^{m-1}}} \mathbb{P} \left(n^{-1} |\langle \mathbf{X}\boldsymbol{\beta}, (\mathbf{W} \circ \mathbf{U})\boldsymbol{\alpha} \rangle| \geq s \right) \\ &\leq 8^{p+m} \sup_{\substack{\boldsymbol{\beta} \in \mathcal{S}^{p-1}, \\ \boldsymbol{\alpha} \in \mathcal{S}^{m-1}}} \mathbb{P} \left(n^{-1} \sum_{i=1}^n \langle \boldsymbol{\beta}, \mathbf{X}_i \rangle \langle \boldsymbol{\alpha}, (\mathbf{W} \circ \mathbf{U})_i \rangle \geq s \right), \end{aligned} \quad (\text{B.7})$$

where $\mathcal{S}^{m-1} \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in \mathbb{R}^m : \|\boldsymbol{\alpha}\|_2 = 1\}$ is the Euclidean sphere in m -dimensions. $\forall s \geq 0$, there exists $C > 0$ such that $\mathbb{P}(|u_{ij}| > s) \leq \exp(1 - s^2/C^2)$. Since $|w_{ij}| \leq \max(\tau, 1 - \tau)$,

we have

$$\begin{aligned}
\mathbb{P} \left(|w_{ij}u_{ij}| > s \right) &\leq \mathbb{P} \left(\max(\tau, 1 - \tau) |u_{ij}| > s \right) \\
&= \mathbb{P} \left(|u_{ij}| > \frac{s}{\max(\tau, 1 - \tau)} \right) \\
&\leq \exp \left(1 - \frac{s^2}{\max(\tau, 1 - \tau)^2 C^2} \right).
\end{aligned} \tag{B.8}$$

It means for each $j \in \{1, \dots, m\}$, $w_{ij}u_{ij}$ are sub-gaussian. Moreover, the maximal sub-gaussian norm is bounded by

$$\begin{aligned}
\max_{1 \leq j \leq m} \|w_{ij}u_{ij}\|_{\psi_2} &= \max_{1 \leq j \leq m} \sup_{p \geq 1} p^{-1/2} \left(\mathbb{E} |w_{ij}u_{ij}|^p \right)^{1/p} \\
&\leq \max(\tau, 1 - \tau) \max_{1 \leq j \leq m} \sup_{p \geq 1} p^{-1/2} \left(\mathbb{E} |u_{ij}|^p \right)^{1/p} \\
&= \max(\tau, 1 - \tau) K_u.
\end{aligned} \tag{B.9}$$

Then by Hoeffding's inequality (Proposition 5.10 of Vershynin, 2012), we can conclude that $\langle \boldsymbol{\alpha}, (\mathbf{W} \circ \mathbf{U})_i \rangle$ is also sub-gaussian,

$$\begin{aligned}
\mathbb{P} \left(\langle \boldsymbol{\alpha}, (\mathbf{W} \circ \mathbf{U})_i \rangle \geq s \right) &= \mathbb{P} \left(\left| \sum_{j=1}^m \alpha_j w_{ij} u_{ij} \right| \geq s \right) \\
&\leq \exp \left(1 - \frac{C' s^2}{\max(\tau, 1 - \tau)^2 K_u^2 \|\boldsymbol{\alpha}\|_2^2} \right) \\
&= \exp \left(1 - \frac{C' s^2}{\max(\tau, 1 - \tau)^2 K_u^2} \right),
\end{aligned} \tag{B.10}$$

where $C' > 0$ is an absolute constant. Furthermore, its sub-gaussian norm is bounded by

$$\begin{aligned}
\|\langle \boldsymbol{\alpha}, (\mathbf{W} \circ \mathbf{U})_i \rangle\|_{\psi_2} &= \sup_{p \geq 1} p^{-1/2} \left\{ \mathbb{E} |\langle \boldsymbol{\alpha}, (\mathbf{W} \circ \mathbf{U})_i \rangle|^p \right\}^{1/p} \\
&= \sup_{p \geq 1} p^{-1/2} \left(\mathbb{E} \left| \sum_{j=1}^m \alpha_j w_{ij} u_{ij} \right|^p \right)^{1/p} \\
&\leq \max(\tau, 1 - \tau) \sup_{p \geq 1} p^{-1/2} \left(\mathbb{E} \left| \sum_{j=1}^m \alpha_j u_{ij} \right|^p \right)^{1/p} \\
&\leq \max(\tau, 1 - \tau) M K_u,
\end{aligned} \tag{B.11}$$

where $M > 0$ is an absolute constant. The last line comes from Khintchine inequality (Corollary 5.12 of Vershynin, 2012) and recall that $\|\boldsymbol{\alpha}\|_2 = 1$. Applying Hoeffding's

inequality again we can obtain

$$\begin{aligned}
\mathbb{P} \left(n^{-1} \sum_{i=1}^n \langle \boldsymbol{\beta}, \mathbf{X}_i \rangle \langle \boldsymbol{\alpha}, (\mathbf{W} \circ \mathbf{U})_i \rangle \geq s \right) &\leq \exp \left(1 - \frac{C'' s^2 n}{\max(\tau, 1 - \tau)^2 M^2 K_u^2 n^{-1} \sum_{i=1}^n \langle \boldsymbol{\beta}, \mathbf{X}_i \rangle^2} \right) \\
&\leq \exp \left(1 - \frac{C'' s^2 n}{\max(\tau, 1 - \tau)^2 M^2 K_u^2 n^{-1} \|\mathbf{X} \boldsymbol{\beta}\|_2^2} \right), \\
&\leq \exp \left(1 - \frac{C'' s^2 n}{c_2 \max(\tau, 1 - \tau)^2 M^2 K_u^2 \|\boldsymbol{\Sigma}\|} \right),
\end{aligned} \tag{B.12}$$

where C'' is an absolute constant. Combining (B.7) and (B.12) gives

$$\mathbb{P} \left(n^{-1} \|\mathbf{X}^\top (\mathbf{W} \circ \mathbf{U})\| \geq 4s \right) \leq \exp \left(1 - \frac{C'' s^2 n}{9 \max(\tau, 1 - \tau)^2 M^2 K_u^2 \|\boldsymbol{\Sigma}\|} + (p + m) \log 8 \right). \tag{B.13}$$

Set $s = \frac{1}{4} c \max(\tau, 1 - \tau) K_u \sqrt{\|\boldsymbol{\Sigma}\|} \sqrt{\frac{p+m}{n}}$, where $c \stackrel{\text{def}}{=} 4 \cdot \sqrt{2 \log 8 \frac{9M^2}{C''}}$, then we can conclude from the fact $P(\Omega_1) \geq 1 - a_n$,

$$\begin{aligned}
&\mathbb{P} \left(n^{-1} \|\mathbf{X}^\top (\mathbf{W} \circ \mathbf{U})\| \leq c \max(\tau, 1 - \tau) K_u \sqrt{\|\boldsymbol{\Sigma}\|} \sqrt{\frac{p+m}{n}} \right) \\
&\geq [1 - \exp(1 - (p + m) \log 8)] \times (1 - a_n) \\
&\geq [1 - 3 \times 8^{-(p+m)}] \times (1 - a_n) \\
&\geq 1 - 3 \times 8^{-(p+m)} - a_n \quad (\text{as } p + m > 1).
\end{aligned} \tag{B.14}$$

This finishes the proof. $\square$

Lemma B.2. Under (A1)-(A3), selecting $\lambda = 2cm^{-1} \max(\tau, 1 - \tau) K_u \sqrt{\|\boldsymbol{\Sigma}\|} \sqrt{\frac{p+m}{n}}$, for $n \geq 2 \min(m, p)$, any optimal solution $\hat{\boldsymbol{\Gamma}}$ in the problem (2.5) satisfies the bound

$$\begin{aligned}
\|\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_{\text{F}}^2 &\leq c' \frac{p+m}{n} \frac{\max(\tau, 1 - \tau)^2}{\min(\tau, 1 - \tau)^2} \frac{\|\boldsymbol{\Sigma}\|}{\sigma_{\min}(\boldsymbol{\Sigma})^2} K_u^2 \dim(\overline{\mathcal{M}}) \\
&\quad + c' \sqrt{\frac{p+m}{n}} \frac{\max(\tau, 1 - \tau)}{\min(\tau, 1 - \tau)} \frac{\sqrt{\|\boldsymbol{\Sigma}\|}}{\sigma_{\min}(\boldsymbol{\Sigma})} K_u \|\boldsymbol{\Gamma}_{\mathcal{M}^\perp}\|_*,
\end{aligned} \tag{B.15}$$

with probability greater than $1 - 3 \times 8^{-(p+m)} - a_n$, where $c, c' > 0$ are absolute constants.

Proof of Lemma B.2. Recall that Ω_1 is defined as (B.5), and let the event that (B.6) holds as Ω_2 . On event $\Omega_1 \cap \Omega_2$, (B.15) can be achieved by simply plugging $\lambda = 2cm^{-1} \max(\tau, 1 - \tau) K_u \sqrt{\|\boldsymbol{\Sigma}\|} \sqrt{\frac{p+m}{n}}$ into (B.4). We note that

$$\begin{aligned}
\mathbb{P}(\Omega_2 \cap \Omega_1) &= \mathbb{P}(\Omega_2 | \Omega_1) \mathbb{P}(\Omega_1) \geq [1 - 3 \times 8^{-(p+m)}] \times (1 - a_n) \\
&\geq 1 - 3 \times 8^{-(p+m)} - a_n \quad (\text{as } p + m > 1).
\end{aligned} \tag{B.16}$$

□

APPENDIX C: Auxiliary Results

Lemma C.1. For any $u, \delta \in \mathbb{R}$ and $\tau \in (0, 1)$,

$$\rho_\tau(u + \delta) - \rho_\tau(u) - \rho'_\tau(u)\delta \geq \min(\tau, 1 - \tau)\delta^2. \quad (\text{C.1})$$

Proof of Lemma C.1. When $u = 0$, we have $\rho_\tau(u) = \rho'_\tau(u) = 0$, therefore

$$\rho_\tau(\delta) = |\tau - \mathbf{1}\{\delta < 0\}|\delta^2 \geq \min(\tau, 1 - \tau)\delta^2.$$

If $u > 0$, $u + \delta < 0$ ($\delta < 0$), we have

$$\rho_\tau(u + \delta) - \rho_\tau(u) - \rho'_\tau(u)\delta - \min(\tau, 1 - \tau)\delta^2 = \begin{cases} (1 - 2\tau)(\delta + u)^2 \geq 0 & \text{for } \tau \leq 1 - \tau; \\ (1 - 2\tau)(u + 2\delta)u > 0 & \text{for } \tau > 1 - \tau. \end{cases}$$

If $u > 0$, $u + \delta > 0$ ($\delta > 0$), we have

$$\rho_\tau(u + \delta) - \rho_\tau(u) - \rho'_\tau(u)\delta - \min(\tau, 1 - \tau)\delta^2 = \begin{cases} (2\tau - 1)(u + 2\delta)u \geq 0 & \text{for } \tau \leq 1 - \tau; \\ (2\tau - 1)(u + \delta)^2 u > 0 & \text{for } \tau > 1 - \tau. \end{cases}$$

In the other two cases,

$$\rho_\tau(u + \delta) - \rho_\tau(u) - \rho'_\tau(u)\delta = \begin{cases} \tau\delta^2 \geq \min(\tau, 1 - \tau)\delta^2 & \text{for } u > 0, u + \delta \geq 0; \\ (1 - \tau)\delta^2 \geq \min(\tau, 1 - \tau)\delta^2 & \text{for } u < 0, u + \delta \leq 0. \end{cases}$$

Therefore, we can conclude that

$$\rho_\tau(u + \delta) - \rho_\tau(u) - \rho'_\tau(u)\delta \geq \min(\tau, 1 - \tau)\delta^2.$$

□

Lemma C.2. $g(\mathbf{\Gamma})$ defined in (2.9) is κ -strongly convex and differentiable with $\kappa = m^{-1} \min(\tau, 1 - \tau) \sigma_{\min}(\frac{\mathbf{X}^\top \mathbf{X}}{n})$.

Proof of Lemma C.2. Denote $\tilde{u}_{ij} \stackrel{\text{def}}{=} Y_{ij} - \mathbf{X}_i^\top (\mathbf{\Gamma}_{\cdot j} + \mathbf{\Delta}_{\cdot j})$ and $u_{ij} \stackrel{\text{def}}{=} Y_{ij} - \mathbf{X}_i^\top \mathbf{\Gamma}_{\cdot j}$, for

$i = 1, \dots, n, j = 1, \dots, m$, we have

$$\begin{aligned}
\langle\langle \nabla g(\Gamma), \Delta \rangle\rangle &= \text{tr}(\nabla g(\Gamma)^\top \Delta) \\
&= -(mn)^{-1} \sum_{j=1}^m \sum_{l=1}^p \Delta_{lj} \sum_{i=1}^n \rho'(u_{ij}) X_{il} \\
&= -(mn)^{-1} \sum_{i=1}^n \sum_{j=1}^m \left\{ \sum_{l=1}^p \Delta_{lj} \rho'(u_{ij}) X_{il} \right\} \\
&= -(mn)^{-1} \sum_{i=1}^n \sum_{j=1}^m \left\{ \rho'(u_{ij}) \mathbf{X}_i^\top \Delta_{\cdot j} \right\}. \tag{C.2}
\end{aligned}$$

Therefore,

$$\begin{aligned}
g(\Gamma + \Delta) - g(\Gamma) - \langle\langle \nabla g(\Gamma), \Delta \rangle\rangle &= (mn)^{-1} \sum_{i=1}^n \sum_{j=1}^m \left\{ \rho(\tilde{u}_{ij}) - \rho(u_{ij}) + \rho'(u_{ij}) \mathbf{X}_i^\top \Delta_{\cdot j} \right\} \\
&\geq (mn)^{-1} \min(\tau, 1 - \tau) \sum_{i=1}^n \sum_{j=1}^m (\mathbf{X}_i^\top \Delta_{\cdot j})^2 \text{ (by Lemma C.1)} \\
&= (mn)^{-1} \min(\tau, 1 - \tau) \|\mathbf{X}\Delta\|_F^2 \\
&= (mn)^{-1} \min(\tau, 1 - \tau) \text{tr}(\Delta^\top \mathbf{X}^\top \mathbf{X} \Delta) \\
&\geq m^{-1} \min(\tau, 1 - \tau) \sigma_{\min}\left(\frac{\mathbf{X}^\top \mathbf{X}}{n}\right) \|\Delta\|_F^2. \tag{C.3}
\end{aligned}$$

□

APPENDIX D: Additional Details for Section 3

D.1 Risk Attitude Parameter

The risk attitude parameter β is estimated by logistic model via maximum likelihood estimation (MLE)

$$\begin{aligned}
P\{\text{risky choice}|x\} &= \frac{1}{1 + \exp[-\sigma\{\bar{x} - \beta S(x) - 5\}]} \\
P\{\text{sure choice}|x\} &= 1 - \frac{1}{1 + \exp[-\sigma\{\bar{x} - \beta S(x) - 5\}]} \tag{D.1}
\end{aligned}$$

where x is the return stream displayed to the individual, its mean and standard deviation are $\bar{x}$ and $S(x)$.

The estimated risk attitude parameters for 19 subjects in order are plotted in Figure 3.2, also see Majer et al. (2015). Negative parameters imply risk-seeking behaviours; while

positive parameters indicate averse risk patterns. We can see most of the individuals are risk-averse and the two extremes #1 and #19 are the most risk-averse and most risk-seeking persons respectively.

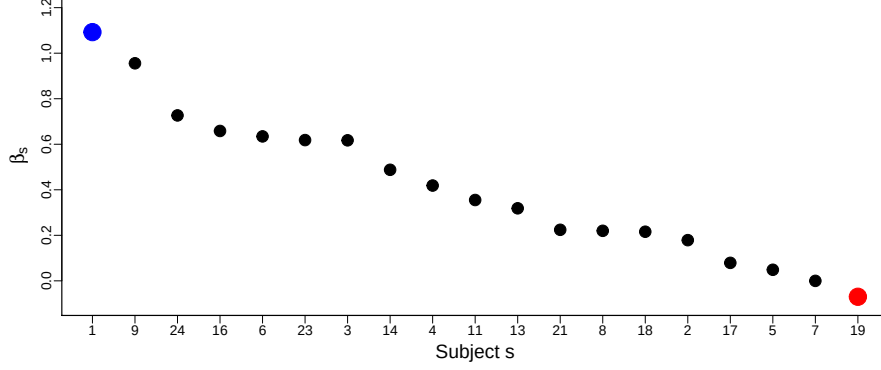


Figure 3.2: Estimated risk attitude for 19 subjects.

D.2 Tuning Parameters by Cross-Validation

Choosing $\omega = 0.1$, $b = 1$ (aINS_L cluster) in (C1) case as an example, Figure 3.3 illustrates the cross-validation error function in terms of λ under different τ levels. The optimal tuning parameters determined by 5-fold cross-validation under all cases are reported in Table 3.2.

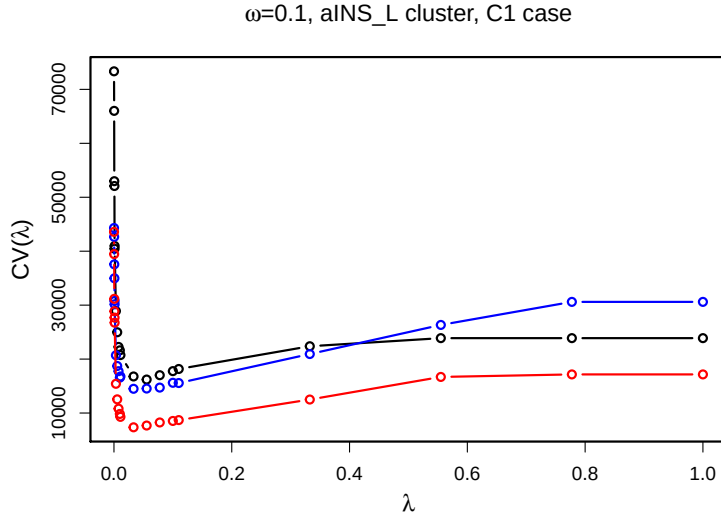


Figure 3.3: The cross-validation error function in terms of tuning parameter λ , with $\tau=0.1$, 0.5, and 0.9, respectively.

		Whole series (C1)			Task-wise (C2)		
τ		0.1	0.5	0.9	0.1	0.5	0.9
$\omega=0.1$	aINS _L	0.0442	0.0552	0.0383	0.0008	0.0006	0.0008
	aINS _R	0.0303	0.0421	0.0293	0.0004	0.0008	0.0004
	DMPFC	0.0348	0.0504	0.0198	0.0004	0.0007	0.0006
$\omega=0.5$	aINS _L	0.0181	0.0403	0.0153	0.0004	0.0006	0.0003
	aINS _R	0.0137	0.0393	0.0157	0.0006	0.0004	0.0005
	DMPFC	0.0195	0.0391	0.0143	0.0006	0.0002	0.0007
$\omega=0.9$	aINS _L	0.0253	0.0408	0.0275	0.0006	0.0004	0.0004
	aINS _R	0.0243	0.0442	0.0200	0.0008	0.0002	0.0006
	DMPFC	0.0193	0.0474	0.0206	0.0005	0.0008	0.0008

Table 3.2: Tuning parameters by 5-fold cross validation.

SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Downside risk and stock returns: An empirical analysis of the long-run and short-run dynamics from the G-7 Countries" by Cathy Yi-Hsuan Chen, Thomas C. Chiang and Wolfgang Karl Härdle, January 2016.
- 002 "Uncertainty and Employment Dynamics in the Euro Area and the US" by Aleksei Netsunajev and Katharina Glass, January 2016.
- 003 "College Admissions with Entrance Exams: Centralized versus Decentralized" by Isa E. Hafalir, Rustamdjan Hakimov, Dorothea Kübler and Morimitsu Kurino, January 2016.
- 004 "Leveraged ETF options implied volatility paradox: a statistical study" by Wolfgang Karl Härdle, Sergey Nasekin and Zhiwu Hong, February 2016.
- 005 "The German Labor Market Miracle, 2003 -2015: An Assessment" by Michael C. Burda, February 2016.
- 006 "What Derives the Bond Portfolio Value-at-Risk: Information Roles of Macroeconomic and Financial Stress Factors" by Anthony H. Tu and Cathy Yi-Hsuan Chen, February 2016.
- 007 "Budget-neutral fiscal rules targeting inflation differentials" by Maren Brede, February 2016.
- 008 "Measuring the benefit from reducing income inequality in terms of GDP" by Simon Voigts, February 2016.
- 009 "Solving DSGE Portfolio Choice Models with Asymmetric Countries" by Grzegorz R. Dlugoszek, February 2016.
- 010 "No Role for the Hartz Reforms? Demand and Supply Factors in the German Labor Market, 1993-2014" by Michael C. Burda and Stefanie Seele, February 2016.
- 011 "Cognitive Load Increases Risk Aversion" by Holger Gerhardt, Guido P. Biele, Hauke R. Heekeren, and Harald Uhlig, March 2016.
- 012 "Neighborhood Effects in Wind Farm Performance: An Econometric Approach" by Matthias Ritter, Simone Pieralli and Martin Odening, March 2016.
- 013 "The importance of time-varying parameters in new Keynesian models with zero lower bound" by Julien Albertini and Hong Lan, March 2016.
- 014 "Aggregate Employment, Job Polarization and Inequalities: A Transatlantic Perspective" by Julien Albertini and Jean Olivier Hairault, March 2016.
- 015 "The Anchoring of Inflation Expectations in the Short and in the Long Run" by Dieter Nautz, Aleksei Netsunajev and Till Strohsal, March 2016.
- 016 "Irrational Exuberance and Herding in Financial Markets" by Christopher Boortz, March 2016.
- 017 "Calculating Joint Confidence Bands for Impulse Response Functions using Highest Density Regions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, March 2016.
- 018 "Factorisable Sparse Tail Event Curves with Expectiles" by Wolfgang K. Härdle, Chen Huang and Shih-Kang Chao, March 2016.
- 019 "International dynamics of inflation expectations" by Aleksei Netšunajev and Lars Winkelmann, May 2016.
- 020 "Academic Ranking Scales in Economics: Prediction and Imputation" by Alona Zharova, Andrija Mihoci and Wolfgang Karl Härdle, May 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 021 "CRIX or evaluating blockchain based currencies" by Simon Trimborn and Wolfgang Karl Härdle, May 2016.
- 022 "Towards a national indicator for urban green space provision and environmental inequalities in Germany: Method and findings" by Henry Wüstemann, Dennis Kalisch, June 2016.
- 023 "A Mortality Model for Multi-populations: A Semi-Parametric Approach" by Lei Fang, Wolfgang K. Härdle and Juhyun Park, June 2016.
- 024 "Simultaneous Inference for the Partially Linear Model with a Multivariate Unknown Function when the Covariates are Measured with Errors" by Kun Ho Kim, Shih-Kang Chao and Wolfgang K. Härdle, August 2016.
- 025 "Forecasting Limit Order Book Liquidity Supply-Demand Curves with Functional AutoRegressive Dynamics" by Ying Chen, Wee Song Chua and Wolfgang K. Härdle, August 2016.
- 026 "VAT multipliers and pass-through dynamics" by Simon Voigts, August 2016.
- 027 "Can a Bonus Overcome Moral Hazard? An Experiment on Voluntary Payments, Competition, and Reputation in Markets for Expert Services" by Vera Angelova and Tobias Regner, August 2016.
- 028 "Relative Performance of Liability Rules: Experimental Evidence" by Vera Angelova, Giuseppe Attanasi, Yolande Hiriart, August 2016.
- 029 "What renders financial advisors less treacherous? On commissions and reciprocity" by Vera Angelova, August 2016.
- 030 "Do voluntary payments to advisors improve the quality of financial advice? An experimental sender-receiver game" by Vera Angelova and Tobias Regner, August 2016.
- 031 "A first econometric analysis of the CRIX family" by Shi Chen, Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle, TM Lee and Bobby Ong, August 2016.
- 032 "Specification Testing in Nonparametric Instrumental Quantile Regression" by Christoph Breunig, August 2016.
- 033 "Functional Principal Component Analysis for Derivatives of Multivariate Curves" by Maria Grith, Wolfgang K. Härdle, Alois Kneip and Heiko Wagner, August 2016.
- 034 "Blooming Landscapes in the West? - German reunification and the price of land." by Raphael Schoettler and Nikolaus Wolf, September 2016.
- 035 "Time-Adaptive Probabilistic Forecasts of Electricity Spot Prices with Application to Risk Management." by Brenda López Cabrera , Franziska Schulz, September 2016.
- 036 "Protecting Unsophisticated Applicants in School Choice through Information Disclosure" by Christian Basteck and Marco Mantovani, September 2016.
- 037 "Cognitive Ability and Games of School Choice" by Christian Basteck and Marco Mantovani, Oktober 2016.
- 038 "The Cross-Section of Crypto-Currencies as Financial Assets: An Overview" by Hermann Elendner, Simon Trimborn, Bobby Ong and Teik Ming Lee, Oktober 2016.
- 039 "Disinflation and the Phillips Curve: Israel 1986-2015" by Rafi Melnick and Till Strohsal, Oktober 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 040 "Principal Component Analysis in an Asymmetric Norm" by Ngoc M. Tran, Petra Burdejová, Maria Osipenko and Wolfgang K. Härdle, October 2016.
- 041 "Forward Guidance under Disagreement - Evidence from the Fed's Dot Projections" by Gunda-Alexandra Detmers, October 2016.
- 042 "The Impact of a Negative Labor Demand Shock on Fertility - Evidence from the Fall of the Berlin Wall" by Hannah Liepmann, October 2016.
- 043 "Implications of Shadow Bank Regulation for Monetary Policy at the Zero Lower Bound" by Falk Mazelis, October 2016.
- 044 "Dynamic Contracting with Long-Term Consequences: Optimal CEO Compensation and Turnover" by Suvi Vasama, October 2016.
- 045 "Information Acquisition and Liquidity Dry-Ups" by Philipp Koenig and David Pothier, October 2016.
- 046 "Credit Rating Score Analysis" by Wolfgang Karl Härdle, Phoon Kok Fai and David Lee Kuo Chuen, November 2016.
- 047 "Time Varying Quantile Lasso" by Lenka Zbonakova, Wolfgang Karl Härdle, Phoon Kok Fai and Weining Wang, November 2016.
- 048 "Unraveling of Cooperation in Dynamic Collaboration" by Suvi Vasama, November 2016.
- 049 "Q3-D3-LSA" by Lukas Borke and Wolfgang K. Härdle, November 2016.
- 050 "Network Quantile Autoregression" by Xuening Zhu, Weining Wang, Hangsheng Wang and Wolfgang Karl Härdle, November 2016.
- 051 "Dynamic Topic Modelling for Cryptocurrency Community Forums" by Marco Linton, Ernie Gin Swee Teo, Elisabeth Bommers, Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, November 2016.
- 052 "Beta-boosted ensemble for big credit scoring data" by Maciej Zieba and Wolfgang Karl Härdle, November 2016.
- 053 "Central Bank Reputation, Cheap Talk and Transparency as Substitutes for Commitment: Experimental Evidence" by John Duffy and Frank Heinemann, December 2016.
- 054 "Labor Market Frictions and Monetary Policy Design" by Anna Almosova, December 2016.
- 055 "Effect of Particulate Air Pollution on Coronary Heart Disease in China: Evidence from Threshold GAM and Bayesian Hierarchical Model" by Xiaoyu Chen, December 2016.
- 056 "The Effect of House Price on Stock Market Participation in China: Evidence from the CHFS Micro-Datal" by Xiaoyu Chen and Xiaohao Ji, December 2016.
- 057 "Factorisable Multi-Task Quantile Regression" by Shih-Kang Chao, Wolfgang K. Härdle and Ming Yuan, December 2016.
- 058 "Multivariate Factorisable Sparse Asymmetric Least Squares Regression" by Shih-Kang Chao, Wolfgang K. Härdle and Chen Huang, December 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

