Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

FACTORIZED ESTIMATION OF HIGH-DIMENSIONAL NONPARAMETRIC COVARIANCE MODELS

By Jian Zhang[†] and Jie Li[*,†]

University of Kent[†]

Estimation of covariate-dependent conditional covariance matrix in a high-dimensional space poses a challenge to contemporary statistical research. The existing kernel estimators may not be locally adaptive due to using a single bandwidth to explore the smoothness of all entries of the target matrix function. Moreover, the corresponding theory holds only for i.i.d. samples although in most of applications, the samples are dependent. In this paper, we propose a novel estimation scheme to overcome these obstacles by using techniques of factorization, thresholding and optimal shrinkage. Under certain regularity conditions, we show that the proposed estimator is consistent with the underlying matrix even when the sample is dependent. We conduct a set of simulation studies to show that the proposed estimator significantly outperforms its competitors. We apply the proposed procedure to the analysis of an asset return dataset, identifying a number of interesting volatility and co-volatility patterns across different time periods.

1. Introduction. Nonparametric estimation of covariate-dependent conditional covariance matrix $\Sigma(u)$ in covariance models is fundamental to contemporary scientific research including neuroimaging studies in neuroscience, disease mapping in health science, daily ozone concentration analysis in environmental science and asset portfolio risk analysis in finance, among others ([9, 12, 11, 8, 7, 14, 15, 3, 10, 4] and references therein). However, most efforts in nonparametric covariance estimation suffer from a curse of dimensionality [7]. For example, in asset portfolio risk analysis, modeling market-dependent co-volatility of p assets by use of historical return data over n consecutive months involves estimating $p(p+1)/2$ nonparametric curves [5]. The data set we are studying in this paper contains historical returns of 75 assets over three time periods, namely before-financial-crisis, in-financial-crisis and after-financial-crisis with n equal to 84, 36 and 95

[*]Supported in part by the GTA PhD scholarship at University of Kent, the Yunnan Government Scholarship Fund (No.2014070) and the Young Program of Science and Technology Agency of Yunnan Province, China (No.2013FD037).

MSC 2010 subject classifications: Primary 62G08, 62M10; secondary 62P05

Keywords and phrases: High-dimensional nonparametric covariance model, kernel estimator, thresholding, optimal shrinkage, consistency

months respectively. Note that many more assets can be collected for investigation whereas the number of months n in a period is sometimes quite limited. When p is close to or larger than n, the kernel covariance estimator proposed by Yin et al.[11] can be degenerate or ill-conditioned with a high condition number (i.e. with a high ratio of its largest to smallest eigenvalues). Hence, it cannot be reliably inverted to compute the precision matrix which is required in the above risk analysis. To tackle this problem, Chen and Leng [3] proposed a method (called DCM) to regularize the kernel covariance model by thresholding covariance entries. If the resulting estimator is still ill-conditioned, then an eigenvalue-dependent number is added to its diagonals (called ad-hoc shrinkage below). These authors also established a consistency theory for their estimator when the sample is i.i.d. There are three main issues that arise in the application of the DCM. First, in the DCM, the same bandwidth is adopted for estimating entries of varying degrees of smoothness. This may compromise the performance of the DCM. On other hand, letting all the entries have their own bandwidths may overlook the fact that they are constrained in order to form a positive definite matrix. This calls for a factorized estimation with multiple bandwidths adapting to unknown smoothness in factors of the covariance matrix. Secondly, the above ad-hoc shrinkage operation is hard to implement and not optimal from a decision-making point of view. It is desirable to explore an optimal shrinkage procedure. Finally, the existing theory in [3] holds only for i.i.d. samples although, in most of applications, the samples are dependent. For instance, in the above asset portfolio risk analysis, the returns of the market and assets are time series which are serially correlated.

In this paper, we propose a scheme to address these issues. The scheme is based on a factorization of $\Sigma(u)$ in the form of $\Sigma(u) = Q(u)^{1/2}C(u)Q(u)^{1/2}$, where $C(u) = Q(u)^{-1/2}\Sigma(u)Q(u)^{-1/2}$ and $Q(u)$ is an invertible matrix factor of $\Sigma(u)$. For example, if let $Q(u) = \text{diag}(\Sigma(u))$, a diagonal matrix composed by the diagonal entries of $\Sigma(u)$, then $C(u)$ consists of correlation coefficients derived from $\Sigma(u)$. In the scheme, we first estimate $Q(u)$ and $C(u)$ in turn with separate kernel bandwidths. The resulting estimator of $C(u)$ is further enhanced by an entry-wise thresholding. Then , by substituting the resulting estimators of $Q(u)$ and $C(u)$ into the above factorization formula, we can obtain a plug-in estimator of $\Sigma(u)$. Finally, a well-conditioned shrinkage estimator of $\Sigma(u)$ is derived by the principle of minimizing the Frobenius loss. Note that, in practice, $Q(u)$ is often chosen to be less complex than $\Sigma(u)$. For example, let it be much sparse compared to $\Sigma(u)$ when the dimension p is large. Therefore, estimating $Q(u)$ separately from $C(u)$ may help the above procedure circumvent the curse of dimensionality and provide a more accurate estimator for $\Sigma(u)$.

To evaluate the performance of the new proposal, a set of simulation studies are conducted. The results demonstrate that the new proposal substantially outperforms its counterparts in terms of the Frobenius loss. The proposed method is illustrated through an application to the analysis of monthly return data for a group of risky assets mentioned above. The analysis reports the following findings: (1) Some asset returns present a stark nonlinear departure from the linear Capital Asset Pricing Model (CAPM) [5]. (2) Both volatility and co-volatility of these assets are varying with the market. See Figure 1 for more details. These two findings provide an empirical support for building a nonparametric CAPM for risk assessment and portfolio selection. We also establish a theoretical background for the new proposal: we show that under some mixing and regularity conditions, the proposed estimator is asymptotically consistent with the underlying covariance matrix function even when the samples are dependent.

[Put Figure 1 here.]

The rest of the article is organized as follows. In Section 2, the proposed optimal shrinkage estimator is constructed. Then, an algorithm is developed to determine the bandwidth in the kernel smoothing as well as the level of thresholding. In Section 3, the uniform consistency and the convergence rate of the proposed estimator are established with dependent samples. In Section 4, simulation studies are conducted to evaluate the performance of the proposed method and compare it to the method of Chen and Leng [3]. The proposed procedure is employed to analyze asset returns for a group of assets. We conclude with a discussion in Section 5. The technical proofs of asymptotic theory are delayed to the Appendix and the Online Supplementary Material.

Throughout this paper, we let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of a square matrix. For a vector x, let $||x||$ denote its Euclidean norm. For a square matrix $A = (a_{ik})_{p\times p}$, let $||A||_F = \sqrt{\mathrm{tr}(AA^T)/p}$, $||A|| = \lambda_{\max}^{1/2}(AA^T)$, $||A||_{\max} = \max_{1\le i,k\le p} |a_{ik}|$ and $||A||_\infty = \max_{1\le i\le p}\sum_{k=1}^{n} |a_{ik}|$ denote its (size-normalized) Frobenius, spectral, max and ∞-norms. Let $< A, B >= \mathrm{tr}(AB^T)/p$ be the inner product of square matrices A and B. Let $I(\cdot)$ denote an indicator function. Note that these norms satisfy $||A||_F \le ||A|| \le ||A||_\infty \le \max_{1\le i\le p}\sum_{j=1}^{p} I(|a_{ik}| > 0)||A||_{\max}$. Let $\mathrm{diag}(x)$ denote the diagonal matrix with diagonal entries made from the elements of x. Let $c \wedge d$ and $c \vee d$ denote the minimum and maximum of numbers c and d. Let I_p be a p-dimensional identity matrix.

2. Methodology. Let $Y = (Y_1, ..., Y_p)^T \in \mathbb{R}^p$ be a $p-$dimensional random vector and $U \in \mathbb{R}$ be the associated index random variable. We model the conditional mean and covariance matrix of Y given $U = u$ as $\boldsymbol{\mu}(u) = E[Y|U = u]$ and $\text{cov}(Y|U = u) = \Sigma(u)$ respectively whose components are assumed to be an unknown but smooth function of u. Suppose that $(\mathbf{y}_i, u_i)_{i=1}^n$ with $\mathbf{y}_i = (y_{i1}, ..., y_{ip})^T$, are random observations from the population (Y, U), satisfying the equations

$$\mathbf{y}_i = \boldsymbol{\mu}(u_i) + \Sigma(u_i)^{1/2}\varepsilon_i, \quad i = 1, ..., n,$$

where $\boldsymbol{\mu}(u_i) = (\mu_1(u_i), ..., \mu_p(u_i))^T$ and $(u_i)_{i=1}^n$ is a dependent random sample of U. Also, given $(u_i)_{i=1}^n$, ε_i's are dependent on each other and with zero means and unity covariance matrices (i.e., $E[\varepsilon_i|u_i] = 0_p$, $\text{cov}(\varepsilon_i|u_i) = I_p$ and $E[\varepsilon_i\varepsilon_j^T] \neq 0, i \neq j$). Let $K(u)$ be a kernel density function, $K_h(u) = h^{-1}K(u/h)$ (the scaled kernel function with a bandwidth $h > 0$) and $w_{ih}(u) = K_h(u_i - u)/\sum_{k=1}^n K_h(u_k - u)$ (the weighting function). Yin et al. [12] considered the following kernel estimators for $\boldsymbol{\mu}(\cdot)$ and $\Sigma(\cdot)$:

$$\begin{aligned}(2.1)\quad \hat{\boldsymbol{\mu}}(u) &= \sum_{i=1}^n w_{ih_1}(u)\mathbf{y}_i, \\ \hat{\Sigma}(u) &= \sum_{i=1}^n w_{ih}(u)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}(u_i))(\mathbf{y}_i - \hat{\boldsymbol{\mu}}(u_i))^T \hat{=} (\hat{\sigma}_{kj}(u))_{1\leq k,j\leq p},\end{aligned}$$

where h_1 and h are bandwidths for mean and covariance matrix functions respectively.

As the diagonals in a non-negative definite matrix often determine its behavior, we focus on the factorization $\Sigma(u) = Q(u)^{1/2}C(u)Q(u)^{1/2}$, where $Q(u) = \text{diag}(\Sigma(u))$ and $C(u) = Q(u)^{-1/2}\Sigma(u)Q(u)^{-1/2}$. However, the idea can be extended to other banded matrices. Note also that both the conditional number and the estimation accuracy of $\Sigma(u)$ are determined by the corresponding values of $C(u)$ and $Q(u)$. This makes it possible to improve the estimation accuracy of $\Sigma(u)$ by separately enhancing estimation of $Q(u)$ and $C(u)$ and by using the techniques of entry-wise thresholding and optimal shrinkage. The details are as follows.

Construction of a plug-in estimator for $\Sigma(u)$. We start with the kernel estimator $\hat{Q}(u) = \text{diag}(\hat{\sigma}_{kk}(u) : 1 \leq k \leq p)$ with a $Q(u)$-specified bandwidth $h = h_2$. Then, we standardize $\mathbf{y}_i, 1 \leq i \leq n$ by using $\hat{\boldsymbol{\mu}}(u_i)$ and $\hat{Q}(u)$:

$$\tilde{\mathbf{y}}_i = \hat{Q}(u)^{-1/2}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}(u_i)), \quad 1 \leq i \leq n.$$

Based on the standardized observations, we construct the following kernel estimator of $C(u)$:

$$\hat{C}(u) = \sum_{i=1}^{n} w_{ih}(u)\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T \tag{2.2}$$

with a $C(u)$-specified bandwidth h_3. As pointed out before, the dimension p is frequently large than the local sample size nh. This results in a degenerate estimator $\hat{C}(u)$. Following [1], we regularize the above correlation coefficient estimator by thresholding its entries as follows:

$$\hat{C}^{(t)}(u) = \Big(\hat{c}_{jk}(u)I(|\hat{c}_{jk}(u)| > t_0(u)\sqrt{\log(p/h)/(nh)})\Big)_{1\le j,k\le p},$$

where $\hat{c}_{jk}(u)$ is the (j,k)-th entry of $\hat{C}(u)$ and $I(\cdot)$ is an indicator function and $t_0(u)$ is a positive function of u.

Using the above estimators, we construct a plug-in estimator of $\Sigma(u)$ in form

$$\hat{\Sigma}^{(t)}(u) = \hat{Q}(u)^{1/2}\hat{C}^{(t)}(u)\hat{Q}(u)^{1/2}.$$

Shrinkage of $\hat{\Sigma}^{(t)}(u)$. In Section 3 below, under sparsity and regularity conditions, we show that the above thresholded covariance estimator is consistent with the underlying covariance matrix function as n and p tend to infinity. However, for a finite sample, the proposed estimator may still be ill-conditioned. To ameliorate it, we propose to shrink $\hat{\Sigma}^{(t)}(u)$ to the identity matrix I_p, where the amount of shrinkage is optimized in terms of the data-driven Frobenius loss in two steps. Step 1, we find a population version, namely a linear combination of I_p and $\hat{\Sigma}^{(t)}(u)$, $\Sigma^*(u) = \rho a I_p + (1-\rho)\hat{\Sigma}^{(t)}(u)$, whose expected Frobenius loss $E||\Sigma^*(u) - \Sigma(u)||_F^2$ attains the minimum with respect to $0 \le \rho \le 1$ and $a \in \mathbb{R}$. For this purpose, we decompose the above expected quadratic loss as follows:

$$\begin{aligned} E||\Sigma^*(u) - \Sigma(u)||_F^2 &= E||\Sigma^*(u) - E[\Sigma^*(u)] + E[\Sigma^*(u)] - \Sigma(u)||^2 \\ &= (1-\rho)^2 E||\hat{\Sigma}^{(t)}(u) - E[\hat{\Sigma}^{(t)}(u)||_F^2 \\ &\quad + ||\rho(aI_p - E[\Sigma^{(t)}(u)]) + E[\hat{\Sigma}^{(t)}(u)] - \Sigma(u)||_F^2. \end{aligned} \tag{2.3}$$

Differentiating the above loss with respect to a and setting it to zero, we have

$$\begin{aligned} dE||\Sigma^*(u) - \Sigma(u)||_F^2/da &= 2\rho < I_p, \rho(aI_p - E[\hat{\Sigma}^{(t)}(u)]) \\ &\quad + E[\hat{\Sigma}^{(t)}(u)] - \Sigma(u) = 0, \end{aligned}$$

which yields

$$a(u) =< I_p, E[\hat{\Sigma}^{(t)}(u)] > -\rho^{-1} < I_p, E[\hat{\Sigma}^{(t)}(u)] - \Sigma(u) > .$$

Substituting it back to (2.3), we have

$$\begin{aligned} E||\Sigma^*(u) - \Sigma(u)||_F^2 &= (1-\rho)^2 E||\hat{\Sigma}^{(t)}(u) - E[\hat{\Sigma}^{(t)}(u)]||^2 \\ &\quad +||(1-\rho)A_h - A||_F^2 \\ &= (1-\rho)^2 E||\Sigma^*(u) - \Sigma(u)||^2 + \rho^2 ||A_h||_F^2 \\ &\quad +||A_h - A||_F^2 - 2\rho < A_h, A_h - A >, \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} A_h(u) &= E[\hat{\Sigma}^{(t)}(u)] - < I_p, E[\hat{\Sigma}^{(t)}(u)] > I_p. \\ A(u) &= \Sigma(u) - < I_p, \Sigma(u) > I_p. \end{aligned}$$

Differentiating (2.4) with respect to ρ and setting it to zero, we have

$$\begin{aligned} &-2(1-\rho)E||\hat{\Sigma}^{(t)}(u) - \Sigma(u)||^2 + 2\rho||A_h(u)||_F^2 \\ &\quad -2 < A_h(u), A_h(u) - A(u) >= 0. \end{aligned}$$

Solving the above equation, we have the solution

$$\rho_h(u) = \frac{\beta_h^2(u) + Q_h(u)}{\beta_h^2(u) + \alpha_h^2(u)} \wedge 1, \tag{2.5}$$

where

$$\begin{aligned} \alpha_h^2(u) &= ||A_h||_F^2, \quad \beta_h^2(u) = E||\hat{\Sigma}^{(t)}(u) - E[\hat{\Sigma}^{(t)}(u)]||_F^2, \\ Q_h(u) &= < A_h(u), A_h(u) - A(u) > . \end{aligned}$$

It is easy to see that $\alpha_h(u)$ is a Frobenius norm of the residual of $E[\hat{\Sigma}^{(t)}(u)]$ after its projection to the space spanned by the identity matrix I_p while $\beta_h^2(u)$ is a Frobenius-type variance of $\hat{\Sigma}^{(t)}(u)$. And $Q_h(u)$ is a bias effect of the kernel smoothing. If replacing ρ in $a(u)$ by $\rho_h(u)$, then we have the solution

$$a_h(u) =< I_p, E[\hat{\Sigma}^{(t)}(u)] > -\rho_h^{-1}(u) < I_p, E[\hat{\Sigma}^{(t)}(u)] - \Sigma(u) > .$$

Therefore, the optimal solution $\hat{\Sigma}^*(u)$ to the above covariance optimization problem is of the form:

$$\hat{\Sigma}^*(u) = \rho_h(u)a_h(u)I_p + (1-\rho_h)\hat{\Sigma}^{(t)}(u).$$

Note that $\alpha_h^2(u)$, $\beta_h^2(u)$ and $Q_h(u)$ in (2.5) depend on unknown matrices $E[\hat{\Sigma}^{(t)}(u)]$ and $\Sigma(u)$. So, in Step 2, we estimate them by the plug-in estimators,

$$\begin{aligned}
\hat{\alpha}_p^2(u) &= ||\hat{\Sigma}^{(t)}(u) - p^{-1}\mathrm{tr}(\hat{\Sigma}^{(t)}(u))I_p||_F^2. \\
\hat{\beta}_p^2(u) &= \frac{1}{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{i=1}^{n} w_{ih}^2(u)\left((y_{ij} - \hat{\mu}_j(u_i))(y_{ik} - \hat{\mu}_k(u_i)) - \hat{\sigma}_{jk}(u)\right)^2 \\
&\quad \times I(|\hat{\sigma}_{jk}(u)| > t_0(u)\sqrt{\log(p/h)/(nh)}).
\end{aligned}$$

It is easy to see that $\hat{\beta}_p^2(u)$ is the squared Frobenius-norm of the variance estimators of $\hat{\sigma}_{jk}(u)$'s. For simplicity, we shrink $Q_h(u)$ to zero. Combining the above two steps gives the following estimator of $\Sigma(u)$ with a data-driven optimal amount of shrinkage:

$$\hat{\Sigma}^{(st)}(u) = \frac{\hat{\beta}_p^2(u)}{\hat{\alpha}_p^2(u) + \hat{\beta}_p^2(u)} p^{-1}\mathrm{tr}(\hat{\Sigma}^{(t)}(u))I_p + \frac{\hat{\alpha}_p^2(u)}{\hat{\alpha}_p^2(u) + \hat{\beta}_p^2(u)}\hat{\Sigma}^{(t)}(u).$$

3. Theory.

Let $\mathcal{F}_{k_0}$ and $\mathcal{F}_{k_0+k}^{\infty}$ be the σ-algebras generated by $\{(\mathbf{y}_i, u_i) : 1 \le i \le k_0\}$ and $\{(\mathbf{y}_i, u_i) : k_0 + 1 \le k < \infty\}$. Define

$$\alpha(k) = \max_{k_0 \ge 1} \sup_{A \in \mathcal{F}_{k_0}, B \in \mathcal{F}_{k_0+k}^{\infty}} |P(A)P(B) - P(A \cap B)|.$$

We assume the following regularity conditions:

(C1) The symmetric kernel function $K(\cdot)$ on $\mathbb{R}$ with derivative $K'(\cdot)$ satisfies

$$\begin{aligned}
K_0 = \sup_z K(z) &< +\infty, \quad K_1 = \sup_z |K'(z)| < +\infty, \quad \int K(z)dz = 1, \\
\int zK(z)dz &= 0, \quad \int z^2 K(z)dz < +\infty, \quad \int |z|^3 k(z)dz < \infty.
\end{aligned}$$

(C2) The density function of U, $g(u)$, has the second order continuous derivative $g''(\cdot)$ over a compact support $[a, b]$ and $\inf_{u \in [a,b]} g(u) > 0$. For any $i \ne i_1$, the joint density of u_i and u_{i_1}, $\max_{i \ne i_1} \sup_{z, z_1 \in [a,b]} g_{ii_1}(z, z_1)$ is bounded.

(C3) There exist positive constants τ_2 and $\kappa_2 < 1$ such that for $k \ge 1$, $\alpha(k) \le \exp(-\tau_2 k^{\kappa_2})$.

(C4) There exist constants $0 < \kappa_1 \le 1, \tau_1 > 0$ such that

$$\max_{1 \le j \le p} P(|y_{ij}| > v) \le \exp(1 - \tau_1 v^{\kappa_1}).$$

(C5) The second derivatives of $\mu_j(u) = E[y_{1j}|U = u]$, $1 \le j \le p$ are uniformly bounded in the sense that $\max_{1\le j\le p}\sup_{u\in[a,b]}|\mu_j''(u)| < \infty$.

(C6) The conditional variance functions $\sigma_j^2(u) = E[(y_{ij} - \mu_j(u_i))^2|u_i = u]$ are bounded below from zero uniformly for $1 \le j \le p$ and $u \in [a,b]$. Their first order derivatives are also uniformly bounded. The conditional expectations $E[(y_{ij} - \mu_j(u_i))(y_{(i+t)j} - \mu_j(u_{i+k}))|u_i = z, u_{i+t} = z_1]$ with $z, z_1 \in [a,b]$, $1 \le i < \infty$, $1 \le t \le \infty$, $1 \le j \le p$, are uniformly bounded in i, t, z and z_1.

It follows from (C5) that $b_2 \hat{=} \max_{1\le j\le p}\sup_{u\in[a,b]}|\mu_j(u)| < \infty$. Note that these conditions are imposed to facilitate the proofs and thus may not be the weakest possible for establishing the theory below. The above conditions are routinely used in the literature of the kernel smoothing and time series analysis (see [5]).

Let $\hat{g}_{h_1}(u) = \frac{1}{n}\sum_{i=1}^n K_{h_1}(u_i - u)$. be a kernel density estimator of $g(u)$. In the following lemma, we show that $\hat{g}_{h_1}(u)$ is uniformly consistent to $g(u)$.

PROPOSITION 3.1. *Under Conditions (C1)∼(C3), if for a constant* $0 < \zeta_0 < \kappa_2$, *as* $n \to \infty$ *and* $h_1 \to 0$, *the bandwidth* h_1 *satisfies*

$$\frac{\log(nh_1^{-4})}{(nh_1\log(1/h_1))^{\zeta_0/2}} = O(1), \quad \frac{(\log(nh_1\log(1/h_1)))^{\zeta_0}\log(h_1^{-1})}{(nh_1\log(1/h_1))^{\zeta_0(1-\zeta_0)/2}} = O(1),$$

then $\sup_{a\le u\le b}|\hat{g}_{h_1}(u) - g(u)| = O_p\left(\sqrt{\frac{\log(1/h_1)}{nh_1}}\right) + O(h_1^2)$.

Letting $1/\gamma_1 = 1/\kappa_1 + 1/\kappa_2$, we state a uniform consistency result for estimator $\hat{\mu}_j(u)$ in the following theorem.

THEOREM 3.1. *Under Conditions (C1)∼(C6), if as* $n, p \to \infty$ *and* $h_1 \to 0$, *we have* $(\log(p))^{2/\gamma_1 - 1}/n = O(1)$, *and*

$$\frac{\log(h_1^{-4}np)}{(nh_1\log(p/h_1))^{\gamma_1/2}} = O(1), \quad \frac{(\log(nh_1\log(p/h_1))^{\gamma_1}\log(1/h_1)}{(nh_1\log(p/h_1))^{\gamma_1(1-\gamma_1)/2}} = O(1),$$

then

$$\max_{1\le j\le p}\sup_{u\in[a,b]}|\hat{\mu}_j(u) - \mu_j(u)| = O_p\left(\sqrt{\frac{\log(p/h_1)}{nh_1}}\right) + O(h_1^2).$$

Let $1/\gamma_2 = 2/\kappa_1 + 1/\kappa_2$. In the next theorem, we show that the entries of the proposed covariance matrix estimator are consistent with the underlying ones uniformly in u and indices $1 \le j, k \le p$.

THEOREM 3.2. *Under Conditions (C1)∼(C6), if as* $n, p \to \infty$, $h_1, h_3, h \to 0$, $h/h_1 + h_1/h = O(1)$, $h/h_3 + h_3/h = O(1)$, $\log(p)^{2/\gamma_2 - 1}/n = O(1)$ *and*

$$\frac{\log(nph^{-4})}{(nh\log(p/h))^{\gamma_2/2}} = O(1), \quad \frac{(\log(nh\log(p/h)))^{\gamma_2}\log(1/h)}{(nh\log(p/h))^{\gamma_2(1-\gamma_2)/2}} = O(1),$$

then

$$\begin{aligned} \max_{1\le j,k\le p}\sup_{u\in[a,b]} |\hat\sigma_{jk}(u) - \sigma_{jk}(u)| &= O_p\left(\sqrt{\frac{\log(p/h)}{nh}} + h^2\right), \\ \max_{1\le j,k\le p}\sup_{u\in[a,b]} |\hat c_{jk}(u) - c_{jk}(u)| &= O_p\left(\sqrt{\frac{\log(p/h)}{nh}} + h^2\right). \end{aligned}$$

Let $\alpha_p(u) = ||\Sigma(u) - <\Sigma(u), I_p> I_p||_F$ and $\tau_{np} = \sqrt{\log(p/h)/nh}$. Let $\hat t_0(u)$ be an estimator of the thresholding function $t_0(u)$ used in $\hat\Sigma^{(t)}(u)$ and $\hat\Sigma^{(st)}(u)$. Let $m_p(u) = \max_{1\le k\le p}\sum_{j=1}^p I(\sigma_{kj}(u) > 0)$. The smaller $m_p(u)$, the sparser $\Sigma(u)$ is. To state the next theorem, we introduce the following conditions on separability between $\Sigma(u)$ and I_p, sparsity and bounds of $\Sigma(u)$ respectively.

(C7) $\tau_{np}/(\log(p/h)\inf_{u\in[a,b]}\alpha_p^2(u)) = O(1)$, $\sup_{u\in[a,b]} m_p(u)\tau_{np}/\alpha_p(u) = o(1)$.

(C8) There exists a positive constant s_1 such that $\sup_{u\in[a,b]}||\Sigma(u)|| \le s_1$.

(C9) There exists a positive constant s_{0p} such that $s_{0p}/\sqrt{\sup_{u\in[a,b]} m_p(u)\tau_{np}} \to \infty$ and $\inf_{u\in[a,b]}||\Sigma(u)|| \ge s_{0p}$ as $p \to \infty$.

(C10) $\sup_{u\in[a,b]}|\hat t_0(u) - t_0(u)| = o(1)$ and there exist positive constants $t_a < t_b$ such that for $t_a < \inf_{u\in[a,b]} t_0(u) \le \sup_{u\in[a,b]} t_0(u) < t_b$.

Under these conditions, we state a uniform consistent result for $\hat\Sigma^{(st)}(u)$ as follows.

THEOREM 3.3. *Under Conditions (C1)∼(C8), if as* $n, p \to \infty$, $h_1, h, h_3 \to 0$, $h/h_1 + h_1/h = O(1)$, $h/h_3 + h_3/h = O(1)$, $\log(p)^{2/\gamma_3 - 1}/n = O(1)$, $nh^5/\log(p/h) = O(1)$, *and*

$$\frac{\log(nph^{-4})}{(nh\log(p/h))^{\gamma_3/2}} = O(1), \quad \frac{(\log(nh\log(p/h)))^{\gamma_3}\log(1/h)}{(nh\log(p/h))^{\gamma_3(1-\gamma_3)/2}} = O(1),$$

and if $\sup_{u\in[a,b]} m_p(u)\tau_{np} = o(1)$, *then uniformly in* $u \in [a, b]$,

$$||\hat\Sigma^{(st)}(u) - \Sigma(u)|| = O_p(m_p(u)\tau_{np}).$$

In addition to the above conditions, if Condition (C9) holds, then uniformly in $u \in [a,b]$,

$$
\begin{aligned}
||\hat{\Sigma}^{(st)}(u)\Sigma^{-1}(u) - I_p|| &= O_p\left(m_p(u)\tau_{np}s_{0p}^{-1}\right) = o_p\left(\sqrt{m_p(u)\tau_{np}}\right). \\
||\Sigma(u)(\hat{\Sigma}^{(st)}(u))^{-1} - I_p|| &= O_p\left(m_p(u)\tau_{np}s_{0p}^{-1}\right) = o_p\left(\sqrt{m_p(u)\tau_{np}}\right). \\
||(\hat{\Sigma}^{(st)}(u))^{-1} - \Sigma^{-1}(u)|| &= O_p\left(m_p(u)\tau_{np}s_{0p}^{-2}\right) = o_p(1).
\end{aligned}
$$

Finally, in addition to the above conditions, if Condition (C10) holds, then the above results continue to hold after replacing $t_0(u)$ *by* $\hat{t}_0(u)$ *in* $\hat{\Sigma}^{(t)}(u)$ *and* $\hat{\Sigma}^{(st)}(u)$*.*

4. Numerical studies. In this section, to demonstrate the merits of the proposed estimators in finite sample settings, we applied the proposed procedures to both synthetic and real data.

To facility the presentation, let ${}_{\mathrm{t}}\mathrm{NCM}$ and ${}_{\mathrm{st}}\mathrm{NCM}$ denote the proposed estimators $\hat{\Sigma}^{(t)}(u)$ and $\hat{\Sigma}^{(st)}(u)$ respectively. Let DCM_1 and DCM_2 denote two DCM estimators defined by

$$
\begin{aligned}
\mathrm{DCM}_1(u) &= (\hat{\sigma}_{1jk}(u)I(\hat{\sigma}_{1jk}(u) \geq d(u))), \\
\mathrm{DCM}_2(u) &= (\hat{\sigma}_{2jk}(u)I(\hat{\sigma}_{2jk}(u) \geq d(u))),
\end{aligned}
\tag{4.1}
$$

where $d(u)$ is a thresholding constant and

$$
\begin{aligned}
\hat{\Sigma}_1(u) &= \sum_{i=1}^n w_{ih}(u)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}(u_i))(\mathbf{y}_i - \hat{\boldsymbol{\mu}}(u_i))^T \hat{=} (\hat{\sigma}_{1jk}(u))_{1\leq j,k\leq p}, \\
\hat{\Sigma}_2(u) &= \sum_{i=1}^n w_{ih}(u)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}(u))(\mathbf{y}_i - \hat{\boldsymbol{\mu}}(u))^T \hat{=} (\hat{\sigma}_{2jk}(u_i))_{1\leq j,k\leq p}.
\end{aligned}
$$

4.1. *Choice of tuning parameters.* As is common in most smoothing methods, the choice of appropriate tuning parameters plays an important role in the performance of a regularized estimator. Data-driven choice of the tuning parameter is a difficult problem. Here we apply the commonly used practical strategy of choosing the values of tuning parameters in a sequential manner through the cross validation . The details are as follows.

Bandwidth for estimating $\boldsymbol{\mu}(u)$*.* We let $h_1 = \arg\min \mathrm{CV}_{\boldsymbol{\mu}}(h)$ as the optimal bandwidth for the mean kernel estimator in equation (2.1), where

$$
\mathrm{CV}_{\boldsymbol{\mu}}(h) = \frac{1}{n}\sum_{i=1}^n ||\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{h,-i}(u_i)||^2 \omega(u_i).
$$

Here, $\hat{\boldsymbol{\mu}}_{h,-i}(u_i)$ is a kernel mean function estimator after dropping the ith observation from the data. The trimming function $\omega(u) = I(u_{(1)} < u < u_{(n-1)})$ is used for reducing the boundary effects on $\mathrm{CV}_{\boldsymbol{\mu}}(h)$, where $u_{(m)}$ is the mth order statistic of $(u_i)_{i=1}^n$.

Tuning parameters for estimating $Q(u)$*.* To select the bandwidth for $\hat{Q}(u)$, for each h, we calculate $\hat{\sigma}_{kk(-i)} : 1 \le k \le p$ after dropping the ith observation. We choose the optimal bandwidth $h_2 = \arg\min \mathrm{CV}_\sigma(h)$ for $\hat{Q}(u)$, where $\mathrm{CV}_\sigma(u)$ is a Stein-loss-based cross-validation function defined by

$$\mathrm{CV}_\sigma(h) = \sum_{i=1}^n \sum_{k=1}^p \left\{ \frac{(y_{ki} - \hat{\boldsymbol{\mu}}_k(u_i))^2}{s_i(\hat{\sigma}_{kk(-i)}(u_i))} + \log(s_i(\hat{\sigma}_{kk(-i)}(u_i))) \right\}.$$

Bandwidth for estimating $C(u)$*.* There are two existing cross-validation methods for selecting the bandwidth h for estimator in (2.2): One is a Stein-loss-based approach [12] which was however applicable only to low-dimensional data. The other is a subset-based approach [2] for high-dimensional data. In this paper, we opt for an alternative approach by choosing $h_3 = \arg\min \mathrm{CV}_C(h)$ at which the following criterion attains the minimum:

$$\mathrm{CV}_C(h) = \frac{1}{n} \sum_{i=1}^n \|\hat{C}_{(-i)}(u_i) - \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^T\|_F^2,$$

where $\hat{C}_{(-i)}(u_i)$ is the kernel estimator of $C(u)$ based on the leave-one-out dataset $(\tilde{\mathbf{y}}_j, u_i)_{j \ne i}$.

Thresholding level for for $\hat{\Sigma}^{(t)}(u)$*.* Following [1, 3], we split the sample into two sub-samples called trial and testing samples and select the threshold by minimizing the Frobenius norm of the difference between the trial sample based estimator after thresholding and the estimated covariance matrix computed from the testing sample. Specifically, we divide the original sample into two samples at random of size n_1 and n_2, where $n_1 = n(1 - 1/\log(n))$ and $n_2 = n/\log(n)$, and repeat this N_1 times. Here, we set $N_1 = 100$ as the default value according to our numerical experience. Let $\hat{\Sigma}_{1,s}(u)$ and $\hat{\Sigma}_{1,s}(u)$ be the plug-in estimators based on n_1 and n_2 observations respectively with the bandwidth selected by the leave-one-out cross validation. Let $\hat{\Sigma}_{1,s}^{(t)}$ be the thresholded estimator derived from $\hat{\Sigma}_{1,s}(u)$ with the thresholding level $t_0(u)$. Given u, we select $t_0(u)$ by minimizing $\sum_{s=1}^{N_1} ||\hat{\Sigma}_{1,s}^{(t)} - \hat{\Sigma}_{2,s}||_F/N_1$.

Tuning parameters for estimating DCM. The bandwidth h and the level of thresholding of the DCM estimators in (4.1) are determined by the so-called subset and sample-splitting approaches respectively [3].

4.2. *Synthetic data.* In this subsection, we carried out a set of simulation studies. To start, we need a criterion to measure the performance of a nonparametric covariance matrix estimator. There are multiple possible criteria, but one particularly convenient choice is integrated root-squared error (IRSE). For any estimator $\hat{\Psi}(u)$ of $\Sigma(u)$, $u \in [a, b]$ the IRSE is defined as

$$\text{IRSE}(\hat{\Psi}) = \int_a^b ||\hat{\Psi}(u) - \Sigma(u)||_F du \approx \frac{1}{K_0} \sum_{k=1}^{K_0} ||\hat{\Psi}(v_k) - \Sigma(v_k)||_F,$$

where $v_k, 1 \leq k \leq K_0$ be grids evenly distributed over the interval (a, b). In the following, we set $K_0 = 20$ for $(a, b) = (-1, 1)$. We considered three settings for $\boldsymbol{\mu}(u)$ and $\Sigma(u)$ in our simulations.

Setting 1: Following [13, 3], we set $\boldsymbol{\mu}(u)$ and $\Sigma(u)$ as follows. Let $\boldsymbol{\mu}(u) = (\mu_1(u), \cdots, \mu_p(u))^T$ with

$$\mu_j(u) = \sum_{k=1}^{50} \frac{(-1)^{k+1}}{k^2} Z_{jk} cos(k\pi u), \quad 1 \leq j \leq p,$$

where $\{Z_{jk} : 1 \leq j \leq p, 1 \leq k \leq 50\}$ is an independent sample drawn from the uniform distribution over $[-5, 5]$. Let $\Sigma(u) = \{\sigma_{ij}(u)\}_{1 \leq i,j \leq p}$ with $\sigma_{ij}(u) = \exp(u/2)[\{\phi(u) + 0.1\}I(|i-j| = 1) + \phi(u)I(|i-j| = 2) + I(i = j)]$ and $\phi(u)$ is the standard normal density. Note that $\text{diag}(\Sigma(u)) = \exp(u/2)I_p$ is spherical and the correlation matrix $C(u) = (c_{ij}(u))_{1 \leq i,j \leq p}$ with $c_{ij}(u) = I(|i-j| = 1) + \phi(u)I(|i-j| = 2) + I(i = j)$ which is equal to zero when $|i - j| \geq 3$. Therefore, $C(u)$ is sparse as it is banded with bandwidth 2.

Setting 2: Following [14], let $\boldsymbol{\mu}(u) = (\mu_1(u), \cdots, \mu_p(u))^T$ with

$$\mu_j(u) = Z_j \exp\left(\frac{(u - \tau_j)^2}{2^2}\right) \sin(2\pi(u - \tau_j)), \qquad 1 \leq j \leq p,$$

where $Z_j, j = 1, \cdots, p$ are independently drawn from uniform distribution $U(-5, 5)$, $\tau = (\tau_1, \cdots, \tau_p)$ is a row vector of p evenly spaced points between -1 and 1. Set $\Sigma(u) = \{\sigma_{ij}(u)\}_{1 \leq i,j \leq p}$ with $\sigma_{ij}(u) = \exp(u/2)\phi(u)^{|i-j|}$. Note that $\text{diag}(\Sigma(u)) = \exp(u/2)I_p$ is spherical and the correlation matrix $C(u) = (c_{ij}(u))_{1 \leq i,j \leq p}$ with $c_{ij}(u) = \phi(u)^{|i-j|}$. Therefore, $c_{ij}(u)$ is decreasing exponentially fast but $C(u)$ is not sparse.

Setting 3: Let $\boldsymbol{\mu}(u)$ be the same as that in **Setting 1**. Let $\Sigma(u) = A^T(u)A(u)$, where the (i, j)th entry of $A(u)$ equals

$$\begin{aligned} a_{ij}(u) &= \exp\left(\frac{u \sin(ij)}{2}\right) \Big\{ [\sin(\pi u) + 0.1]\, I(|i-j| = 1) \\ &\quad + \sin(\pi u) I(|i-j| = 2) + I(i = j) \Big\}. \end{aligned}$$

Note that $\mathrm{diag}(\Sigma(u)) = \mathrm{diag}(\sum_{j=1}^{p} a_{ij}^2 : 1 \le i \le p)$ is not spherical. $C(u)$ is sparse as it is banded with bandwidth 4.

For each combination of (n, p) with $n = 100, 200, 500$ and $p = 50, 100, 150, 300, 500$, we repeated the experiment 90 times, generating 90 datasets of $(\mathbf{y}_i, u_i)$, $1 \le i \le n$. Each dataset was obtained in two steps. In Step 1, we drew $u_i, i = 1, \cdots, n$ independently from the uniform distribution $U(-1, 1)$. In Step 2, for each given u_i, we drew $\mathbf{y}_i$ from the covariance model $\mathbf{y}_i = \boldsymbol{\mu}(u_i) + \Sigma(u_i)^{1/2}\varepsilon_i$, where ε_i, $i = 1, ..., n$ were iteratively drawn from the vector AR(1) model

$$\varepsilon_0 = \xi_0, \quad \varepsilon_i = \rho\varepsilon_{i-1} + \xi_i, \quad i = 1, \cdots, n$$

with $0 \le \rho < 1$ and $\xi_k, k = 0, 1, ...$ independently sampled from the standard p-dimensional Normal $N(0, I_p)$. We considered $\rho = 0, 0.3, 0.8$.

For each combination of (n, p, ρ), we applied $_{\mathrm{t}}$NCM, $_{\mathrm{st}}$NCM, DCM$_1$ and DCM$_2$ to each of 90 datasets and calculated their IRSE values. The mean and standard error of these values are displayed in Tables 1∼ 3 respectively. As example, the CPU time required by DCM$_1$, DCM$_2$, $_{\mathrm{t}}$NCM and $_{\mathrm{st}}$NCM to estimate the covariance models for 90 datasets simulated in Setting 1 is reported in the Web-Appendix D. The results can be summarized as follows:

- On average, the IRSE loss of each procedure was increasing in the dimension p and in the degree of serial correlation ρ while decreasing in sample size n.
- The degrees of sparsity and diagonal homogeneity in $\Sigma(u)$ had an effect on the performance of these four procedures. For example, when $(n, p, \rho) = (100, 300, 0)$, compared to in Setting 1, the IRSE loss of $_{\mathrm{st}}$NCM in Setting 2 increased by 84%. This is not surprising as the degrees of sparsity and diagonal homogeneity in Setting 2 lead to a higher dimensionality (i.e., the number of effective parameters in the model) than that in Setting 1.
- Among the four procedures, $_{\mathrm{st}}$NCM performed best in all three settings, followed by $_{\mathrm{t}}$NCM, DCM$_1$ and DCM$_2$. In particular, the performance of DCM$_2$ was substantially worse than its competitors. For example, for $(n, p, \rho) = (100, 300, 0)$, in Setting 1, compared to DCM$_1$, on average $_{\mathrm{t}}$NCM and $_{\mathrm{st}}$NCM reduced the IRSE loss by 23% and 25% respectively. Compared to $_{\mathrm{t}}$NCM, on average $_{\mathrm{st}}$NCM reduced the IRSE loss by 3%. Compared to DCM$_2$, on average DCM$_1$ reduced the IRSE loss by 99%. In Setting 2, compared to DCM$_1$, on average $_{\mathrm{t}}$NCM and $_{\mathrm{st}}$NCM reduced the IRSE loss by 12% and 16% respectively. Compared to DCM$_2$, on average DCM$_1$ reduced the IRSE loss by 99%. Compared to $_{\mathrm{t}}$NCM, on average $_{\mathrm{st}}$NCM reduced the IRSE loss by 5%. In Setting 3, compared to

DCM_1, on average ${}_{\text{t}}\text{NCM}$ and ${}_{\text{st}}\text{NCM}$ reduced the IRSE by 14% and 15% respectively. Compared to DCM_2, on average DCM_1 reduced the loss by 94%. Compared to ${}_{\text{t}}\text{NCM}$, on average ${}_{\text{st}}\text{NCM}$ reduced the IRSE loss by 2%. The similar conclusion can be made for dependent samples when $\rho = 0.3$ and 0.8. In particular, the optimal shrinkage can reduce the serial correlation effect on the proposed procedure ${}_{\text{st}}\text{NCM}$.
- The CPU-time costs of ${}_{\text{t}}\text{NCM}$ and ${}_{\text{st}}\text{NCM}$ are much less than those of DCM_1 and DCM_2.

[Put Tables 1∼3 here.]

4.3. *Asset return data.* Capital asset pricing model (CAPM) is a model that describes the relationship between systematic risk and expected return for assets, which is widely used throughout finance for the pricing of risky assets. However, the assumption that asset returns are linearly related to the market return is imposed on the model. The primary goal of this study was to extend the CAPM to the nonlinear setting. In particular, we are interested in how the volatility and co-volatility of a group of asset returns depend on the market return.

For this purpose, from the database of Yahoo Finance, we collected monthly return data of 75 assets across 8 sectors over three time-periods, namely, before-financial-crisis period from 02/2001 to 01/2007, in-financial-crisis period from 02/2007 to 01/2010 and after-financial-crisis period from 02/2010 to 12/2017. The sector distribution of these assets as follows. Technology: AAPL, AMD, HPQ, IBM, IIN, INTC, LNGY, LOGI, MSFT, NTAP, NVDA, SNE, TACT and WDC. Health care: AET, AMGN, AZN, BAX, CVS, GILD, GSK, HUM, IMMU, JNJ, LLY, MRK, NVS, PFE, TECH and UNH. Energy: BP, CVX, OXY, RDS-B, SU and XOM. Financial services: C, GS, HSBC, JPM, MS, PGR, RF and THG. Communication services: SHEN, T and TEO. Consumer defensive: BIG, DLTR, FRED, KO, TGT, TUES, UN and WMT. Consumer cyclical: AMZN, EMMS, KSS, SIRI and TM. Industrials: BA, CAJ, DY, EME, FIX, GE, GVA, IR, MMM, MTZ, PWR, SKYW, UPS, UTX and VMI. We also collected the index return of S&P500 which was treated as the market's return.

We applied the proposed ${}_{\text{st}}\text{NCM}$ to the data for each time-period, obtaining the corresponding estimates for mean $\boldsymbol{\mu}(u)$ and covariance matrix $\Sigma(u)$. Here, the diagonals of estimated $\Sigma(u)$ show the volatility of individual returns while estimated correlation coefficient matrix $C(u)$ captures cross-sectional relationships in these returns.

We plotted the estimated individual mean functions and the estimated volatility functions in Figure 1 and also Figures in the Web-Appendix C, the Online Supplementary Material, revealing a number of assets which had nonlinear relationships to the market return. The degree of this non-linearity significantly decreased after financial crisis, indicating that the CAPM fitted to the market better than before the financial crisis. Figure 1 and Figures in the Web-Appendix C also show that the individual volatility of the assets increased a lot during the financial crisis period but returned to normal after the financial crisis. The pattern of the dependence of the volatility on the market also changed a lot after financial crisis: Changes from non-constant volatility functions before the financial crisis to almost constant volatility functions after the financial crisis. We also investigated effects of the financial crisis on the co-volatility of the selected assets by the estimated non-zero correlation coefficient functions. See Figures in the Web-Appendix B, the Online Supplementary Material for the details. By use of the estimated covariance matrix functions, in each time-period, we identified the associated pairs of assets that were of nonzero market-dependent conditional correlation coefficients (and nonzero conditional co-volatility). We further conducted asymptotic tests for significance of co-volatility for these pairs as follows. For any pair of assets (a, b), let $\text{Corr}_{(a,b)}(u)$ denote its correlation coefficient as a function of u (the market's return) and with estimator $\hat{\text{Corr}}_{(a,b)}(u)$. Let $\hat{F}_{(a,b)}(u) = 0.5 \log(1 + \hat{\text{Corr}}_{(a,b)}(u)) / \log(1 - \hat{\text{Corr}}_{(a,b)}(u))$ be Fisher's Z transformation. To test $H_0 : \text{Corr}_{(a,b)}(u) \not\equiv 0$, we considered the test statistics

$$\text{Avec}_{(a,b)} = \sum_{i=1}^{n} |\hat{F}_{(a,b)}(u_i)|/n \approx N(E[|F_{(a,b)}(U)|], \text{var}(|F_{(a,b)}(U)|)/n)$$

and calculated the approximate P-value $P(\sqrt{n}\text{Avec}_{(a,b)}/\sqrt{\hat{\text{var}}(\text{Corr}_{(a,b)}(U))}|N(0,1))$, where the sample variance of $|\hat{F}_{(a,b)}(u_i)|, 1 \le i \le n$ is denoted by $\hat{\text{var}}(|F_{(a,b)}(U)|)$ and $N(0,1)$ is the standard Normal. Then, even after Bonferroni correction for multiple testing, these P-values were all significant ($< 10^{-2}$) for the above selected pairs of assets. The final list of significant pairs are as follows:

- *Before-financial-crisis.* Within Technology: AET-UNH. Within Energy: BP-CVX, BP-OXY, BP-RDSB, BP-SU, BP-XOM, CVX-OXY, CVX-RDSB, CVX-SU, CVX-XOM, GS-MS,OXY-RDSB, OXY-SU, OXY-XOM and RDSB-XOM. Within Consumer defensive: TGT-WMT.
- *In-financial-crisis.* Within Technology: AET-HUM, AET-MRK, AET-UNH and NVDA-WDC. Within Industrials: EME-MTZ. Within Energy: BP-CVX, BP-OXY, BP-RDSB, CVX-RDSB, CVX-XOM, OXY-RDSB, OXY-SU and RDSB-SU. Within Consumer defensive: TGT-TUES. Within Health care: AMGN-JNJ, AZN-GSK, HUM-UNH and

JNJ-NVS. Within Financial services: C-JPM. Between Industrials and Consumer cyclical: IR-TM. Between Consumer cyclical and Consumer defensive: TM-TUES. Between Financial service and Industrials: RF-UPS.

- *After-financial-crisis.* Within Technology: AET-HUM, AET-UNH and INTC-MSFT. Within Industrials: EME-GVA and IR-UTX. Within Energy: BP-CVX, BP-RDSB, BP-SU, BP-XOM, CVX-RDSB, CVX-SU, CVX-XOM, OXY-SU, RDSB-SU and RDSB-XOM. Within Consumer defensive: KO-UN. Within Health care: AMGN-LLY, AMGN-MRK, AMGN-PFE, AZN-GSK, AZN-LLY, GSK-JNJ, GSK-MRK, GSK-NVS, HUM-UNH, LLY-NVS and MRK-NVS. Within Financial services: C-GS, C-HSBC, C-JPM, C-MS, C-RF, GS-JPM, GS-MS, GS-RF, HSBC-MS, JPM-MS, JPM-RF and MS-RF. Between Financial service and Industrials: JPM-MTZ. Between consumer defensive and Financial services: KO-MS. Between Consumer cyclical and Consumer defensive: KSS-TGT. Between Technology and Industrials: IBM-PWR. Between Health care and Consumer defensive: GSK-UN and NVS-UN.

The results indicate that before financial crisis, there were only 16 significant within-sector co-volatility connections between these assets. In particular, there were no significant cross-sectional co-volatility connections between these assets. The number of co-volatility assets within and across sectors was significantly increasing during and after financial-crisis: The number of within-sector co-volatility connections increased from 16 to 22 during the financial crisis period and to 37 after the financial crisis. The number of between-sector co-volatility connections increased from 0 to 3 during the financial crisis period and to 7 after the financial crisis. This implies that in response to the financial crisis, the financial market has been more closely integrated than before the financial crisis.

5. Discussion and conclusion. Estimating covariate-dependent covariance matrix $\Sigma(u)$ of a high-dimensional response vector poses a big challenge to contemporary statistical research. The existing kernel methods in [11, 2] of Yin et al. [12] might not be flexible enough to capture varying smoothness across key parts of the matrix as they used a single bandwidth for all entries of $\Sigma(u)$. Here, we have proposed a novel estimation procedure to overcome this obstacle, based on a simple factorization of $\Sigma(u)$, namely $\Sigma(u) = Q(u)^{1/2}C(u)Q^{1/2}(u)$, where $Q(u) = \text{diag}(\Sigma(u))$ and $C(u)$ is the correlation coefficient function. The proposal has been implemented in two steps. In Step 1, we estimate $Q(u)$ and $C(u)$ robustly by use of separate bandwidths, followed by substituting these estimators in the above factorization formula to obtain a plug-in estimator. In Step 2, we threshold the entries of the plug-in estimator, followed by an optimal shrinkage from a decision-making point of view. The idea can be extended to other $Q(u)$. For example, we can form $Q(u)$ by diagonal blocks of $\Sigma(u)$ or by selecting a subset of eigenvectors of $\Sigma(u)$ as its blocks. We have conducted a set of simulations to demonstrate that the new proposal outperforms the existing DCM approach in terms of estimation loss. To illustrate our new proposal, we have applied it to a dataset of asset returns. We have developed a nonparametric capital asset pricing model to capture volatility and co-volatility among these risky assets. We have showed that under some regularity conditions, the proposed estimator is consistent with the underlying covariance matrix as both the sample size and dimension tend to infinity. There are a few important topics which are remained to address but beyond the scope of this paper, such as nonparametric nonlinear shrinkage (analogous to those in [2]) and multiple-covariate-dependent covariance models.

APPENDIX A: PROOFS

Proof of Proposition 3.1: It follows from Lemma 0.1, the Online Supplementary Material by letting $\phi_0(u_i) = 1$. The proof is completed.

Proof of Theorem 3.1: It follows from Proposition 3.1 that

$$\sup_{u\in[a,b]} |\hat{g}_{h_1}(u) - g(u)| = O_p\left(\sqrt{\frac{\log(1/h_1)}{nh_1}}\right) + O(h_1^2). \tag{A.1}$$

Therefore, under Condition (C2), for sufficiently large n, we have $\inf_{u\in[a,b]} \hat{g}_{h_1}(u) \geq 0.5 \inf_{u\in[a,b]} g(u) > 0$. Note that

$$\hat{\mu}_j(u) - \mu_j(u) = \frac{1}{\hat{g}_{h_1}(u)}(B_{1jn}(u) + B_{2jn}(u) + B_{3jn}(u)), \tag{A.2}$$

where

$$\begin{aligned}
B_{1jn}(u) &= \frac{1}{nh_1}\sum_{j=1}^{n} K((u_i - u)/h_1)(y_j(u_i) - \mu_j(u_i)) \\
B_{2jn}(u) &= \frac{1}{nh_1}\sum_{i=1}^{n} K((u_i - u)/h_1)\mu_j(u_i) - g(u)\mu_j(u) \\
B_{3jn}(u) &= (g(u) - \hat{g}_{h_1}(u))\mu_j(u).
\end{aligned}$$

We show below that

$$B_{3jn}(u) = O_p\left(\sqrt{\frac{\log(1/h_1)}{nh_1}}\right). \tag{A.3}$$

$$\max_{1\le j\le p}\sup_{u\in[a,b]}|B_{1jn}(u)| = O_p\left(\sqrt{\frac{\log(p/h_1)}{nh_1}}\right) + O(h_1^2). \tag{A.4}$$

$$\max_{1\le j\le p}\sup_{u\in[a,b]}|B_{2jn}(u)| = O_p\left(\sqrt{\frac{\log(p/h_1)}{nh_1}}\right) + O(h_1^2). \tag{A.5}$$

The equation (A.3) directly follows from (A.1) and Condition (C5). We employ Lemma 0.1, the Online Supplementary Material to show that equations (A.4) and (A.5) as follows.

To this end, we verify the conditions in Lemma 0.1, the Online Supplementary Material. First, we bound the tail probability of $y_{ij} - \mu_j(u_i)$:

$$\begin{aligned}\max_{1\le j\le p} P(|y_{ij} - \mu_j(u_i)| > v) &\le \max_{1\le j\le p} P(|y_{ij}| > v - b_2)\\ &\le 1 \wedge (\exp(-\tau_1((v-b_2)\vee 0)^{\kappa_1}).\end{aligned} \tag{A.6}$$

When $v^{\kappa_1} \ge b_2^{\kappa_1} + 1/\tau_1$, we have

$$\begin{aligned}\exp(1-\tau_1(v-b_2)^{\kappa_1}) &\le \exp(1-\tau_1(v^{\kappa_1} - b_2^{\kappa_1}))\\ &\le \exp\left((1+\tau_1 b_2^{\kappa_1})\left(1 - \frac{\tau_1}{1+\tau_1 b_2^{\kappa_1}} v^{\kappa_1}\right)\right)\\ &\le \exp\left(1 - \frac{\tau_1}{1+\tau_1 b_2^{\kappa_1}} v^{\kappa_1}\right).\end{aligned}$$

This together with the inequality in (A.6) implies that when $v^{\kappa_1} \ge b_2^{\kappa_1} + 1/\tau_1$,

$$\max_{1\le j\le p}\sup_u P(|y_{ij} - \mu_j(u_i)| > v) \le \exp\left(1 - \frac{\tau_1}{1+\tau_1 b_2^{\kappa_1}} v^{\kappa_1}\right).$$

On other hand, when $v^{\kappa_1} < b_2^{\kappa_1} + 1/\tau_1$,

$$\max_{1\le j\le p}\sup_u P(|y_{ij} - \mu_j(u_i)| > v) \le 1 \le \exp\left(1 - \frac{\tau_1}{1+\tau_1 b_2^{\kappa_1}} v^{\kappa_1}\right).$$

Therefore, we have

$$\max_{1\le j\le p} P(|y_{ij}-\mu_j(u_i)|>v) \le \exp\left(1-\frac{\tau_1}{1+\tau_1 b_2^{\kappa_1}}v^{\kappa_1}\right). \tag{A.7}$$

Equation (A.4) follows from Lemma 0.1, the Online Supplementary Material by letting $\phi_1(y_{ij},u_i)=y_{ij}-\mu_j(u_i)$. Similarly, the equation (A.5) follows from Lemma 0.1, the Online Supplementary Material by letting $\phi_{0j}=\mu_j(u_i)$. The proof is completed by combining equations (A.2) and (A.3)∼(A.5).

Proof of Theorem 3.2: Note that $h/h_1+h_1/h=O(1)$ and $h/h_3+h_3/h=O(1)$ which imply that h/h_1 and h/h_3 are bounded below from zero and above from infinity. Therefore,

$$\frac{\log(nph^{-4})}{(nh\log(p/h))^{\gamma_2/2}} = O(1), \quad \frac{(\log(nh\log(p/h)))^{\gamma_2}\log(1/h)}{(nh\log(p/h))^{\gamma_2(1-\gamma_2)/2}} = O(1)$$

imply that

$$\frac{\log(nph_1^{-4})}{(nh_1\log(p/h_1))^{\gamma_2/2}} = O(1), \quad \frac{(\log(nh_1\log(p/h_1)))^{\gamma_2}\log(1/h_1)}{(nh_1\log(p/h_1))^{\gamma_2(1-\gamma_2)/2}} = O(1).$$
$$\frac{\log(nph_3^{-4})}{(nh_3\log(p/h_3))^{\gamma_2/2}} = O(1), \quad \frac{(\log(nh_3\log(p/h_3)))^{\gamma_2}\log(1/h_3)}{(nh_3\log(p/h_3))^{\gamma_2(1-\gamma_2)/2}} = O(1).$$

As $0<\gamma_2<\kappa_2$ satisfies the conditions in Proposition 3.1, it follows from Proposition 3.1 that

$$\sup_{u\in[a,b]}|\hat{g}_{h_1}(u)-g(u)| = O_p\left(\sqrt{\frac{\log(1/h_1)}{nh_1}}\right)+O(h_1^2).$$
$$\sup_{u\in[a,b]}|\hat{g}_{h}(u)-g(u)| = O_p\left(\sqrt{\frac{\log(1/h)}{nh}}\right)+O(h^2).$$

This implies that for sufficiently large n, both $\sup_{u\in[a,b]}\hat{g}_{h_1}(u)$ and $\sup_{u\in[a,b]}\hat{g}_h(u)$ are bounded below from zero. Note that

$$\begin{aligned}\hat{\sigma}_{jk}(u) &= \sum_{i=1}^n w_{ih}(u)(y_{ij}-\hat{\mu}_j(u_i))(y_{ik}-\hat{\mu}_k(u_i)) \\ &= \tilde{\sigma}_{jk}(u)+\psi_{1jn}(u)+\psi_{2jn}(u)+\psi_{3jn}(u), \end{aligned}\tag{A.8}$$

where

$$\begin{aligned}
\tilde{\sigma}_{jk}(u) &= \sum_{i=1}^{n} w_{ih}(u)(y_{ij}-\mu_j(u_i))(y_{ik}-\mu_k(u_i)),\\
\psi_{1jkn}(u) &= \sum_{i=1}^{n} w_{ih}(u)(y_{ij}-\mu_j(u_i))(\mu_k(u_i)-\hat{\mu}_k(u_i)),\\
\psi_{2jkn}(u) &= \sum_{i=1}^{n} w_{ih}(u)(\mu_j(u_i)-\hat{\mu}_j(u_i))(y_{ik}-\mu_k(u_i))\\
\psi_{3jkn}(u) &= \sum_{i=1}^{n} w_{ih}(u)(\mu_j(u_i)-\hat{\mu}_j(u_i))(\mu_k(u_i)-\hat{\mu}_k(u_i)).
\end{aligned}$$

By Theorem 3.1, we have

$$\begin{aligned}
|\psi_{1jkn}(u)| &\le \left(\sum_{i=1}^{n} w_{ih}(u)(y_{ij}-\mu_j(u_i))^2\right)^{1/2}\left(\sum_{i=1}^{n} w_{ih}(u)(\mu_k(u_i)-\hat{\mu}_k(u_i))^2\right)^{1/2}\\
&\le \tilde{\sigma}_{jj}^{1/2}\left(O_p(\sqrt{\log(p/h_1)/(nh_1)})+O(h_1^2)\right),\\
|\psi_{2jkn}(u)| &\le \left(\sum_{i=1}^{n} w_{ih}(u)(\mu_j(u_i)-\hat{\mu}_j(u_i))^2\right)^{1/2}\left(\sum_{i=1}^{n} w_{ih}(u)(y_{ik}-\mu_k(u_i))^2\right)^{1/2}\\
&\le \left(O_p\left(\sqrt{\log(p/h_1)/(nh_1)}\right)+O(h_1^2)\right)\tilde{\sigma}_{kk}^{1/2},\\
|\psi_{3jkn}(u)| &\le \left(\sum_{i=1}^{n} w_{ih}(u)(\mu_j(u_i)-\hat{\mu}_j(u_i))^2\right)^{1/2}\left(\sum_{i=1}^{n} w_{ih}(u)(\mu_k(u_i)-\hat{\mu}_k(u_i))^2\right)^{1/2}\\
&\le \left(O_p(\sqrt{\log(p/h_1)/(nh_1)})+O(h_1^2)\right)^2.
\end{aligned}$$

To complete the proof, it suffice to prove

$$\max_{1\le j,k\le p}\sup_{u\in[a,b]}|\hat{\sigma}_{jk}(u)-\sigma_{jk}(u)| = O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right). \tag{A.9}$$

Let

$$y_{ij}^* = y_{ij}-\mu_j(u_i), \quad y_{ik}^* = y_{ik}-\mu_k(u_i).$$

Then, $E[y_{ij}^*|u_i]=0$, $\sigma_{jk}(u_i)=E[y_{ij}^*y_{ik}^*|u_i]$ and

$$\begin{aligned}
\tilde{\sigma}_{jk}(u)-\sigma_{jk}(u) &= \frac{1}{\hat{g}_h(u)}\frac{1}{nh}\sum_{i=1}^{n} K(\frac{u_i-u}{h})y_{ij}^*y_{ik}^* - \sigma_{jk}(u)\\
&= \frac{1}{\hat{g}_h(u)}\left(T_{1jkn}(u)+T_{2jkn}(u)+T_{3jkn}(u)\right), \qquad \text{(A.10)}
\end{aligned}$$

where

$$T_{1jkn}(u) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{u_i - u}{h})\left(y_{ij}^* y_{ik}^* - \sigma_{jk}(u_i)\right)$$
$$T_{2jkn}(u) = \frac{1}{nh}\sum_{i=1}^{n}\left(K(\frac{u_i - u}{h})\sigma_{jk}(u_i) - g(u)\sigma_{jk}(u)\right)$$
$$T_{3jkn}(u) = (g(u) - \hat{g}_h(u))\sigma_{jk}(u).$$

By Proposition 3.1, uniformly in $1 \le j,k \le p$ and $u \in [a,b]$, we have

$$\max_{jk}\sup_{u\in[a,b]}|T_{3jkn}(u)| = O_p\left(\sqrt{\frac{\log(1/h)}{nh}} + h^2\right). \tag{A.11}$$

Note that by using (A.7), we have

$$\max_{1\le j,k\le p} P(|y_{ij}^* y_{ik}^*| > v) \le \max_{j,k}\left(P(|y_{ij}^*| > \sqrt{v}) + P(|y_{ik}^*| > \sqrt{v})\right)$$
$$\le 2\exp\left(1 - \frac{\tau_1}{1+\tau_1 b_2^{\kappa_1}} v^{\kappa_1/2}\right).$$

Let $b_3 = \max_{j,k}\sup_w \sigma_{jk}(w)$. Then, we have

$$\max_{1\le j,k\le p} P\left(|y_{ij}^* y_{ik}^* - \sigma_{jk}(u_i)| > v\right)$$
$$\le \max_{j,k} P\left(|y_{ij}^* y_{ik}^*| > (v - b_3)\vee 0\right)$$
$$\le 1 \wedge \exp\left(1 + \log(2) - \frac{\tau_1}{1+\tau_1 b_2^{\kappa_1}}((v-b_3)\vee 0)^{\kappa_1/2}\right)$$
$$\le 1 \wedge \exp\left(1 + \log(2) + \frac{\tau_1 b_3^{\kappa_1/2}}{1+\tau_1 b_2^{\kappa_1}} - \frac{\tau_1 v^{\kappa_1/2}}{1+\tau_1 b_2^{\kappa_1}}\right)$$
$$\le \exp\left(1 - \frac{v^{\kappa_1/2}}{(\tau_1^{-1} + b_2^{\kappa_1})(1+\log(2)) + b_3^{\kappa_1/2}}\right). \tag{A.12}$$

Letting $\phi_{0jk}(u_i) = \sigma_{jk}(u_i)$ in Lemma 0.1, the Online Supplementary Material and invoking (A.12), we have

$$\max_{1\le j,k\le p}\sup_{u\in[a,b]}|T_{2jkn}(u)| = O_p\left(\sqrt{\frac{\log(p/h)}{nh}}\right) + O(h^2). \tag{A.13}$$

Similarly, letting $\phi_2(y_{ij}, y_{ik}, u_i) = y_{ij}^* y_{ik}^* - \sigma_{jk}(u_i)$ in Lemma 0.1(iii), the Online Supplementary Material, we have

$$\max_{1\le j,k\le p}\sup_{u\in[a,b]}|T_{1jkn}(u)| = O_p\left(\sqrt{\frac{\log(p/h)}{nh}}\right) + O(h^2).$$

This together with (A.8), (A.11) and (A.13) gives

$$\max_{1\le j,k\le p}\sup_{u\in[a,b]}|\tilde\sigma_{jk}(u)-\sigma_{jk}(u)| \;=\; O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right)$$

and

$$\begin{aligned}
|\psi_{1jkn}(u)| &\le \tilde\sigma_{jj}^{1/2}(u)\left(O_p(\sqrt{\log(p/h_1)/(nh_1)})+O(h_1^2)\right)\\
&= O_p(\sqrt{\log(p/h_1)/(nh_1)}+h_1^2).\\
|\psi_{2jkn}(u)| &\le \left(O_p\left(\sqrt{\log(p/h_1)/(nh_1)}\right)+O(h_1^2)\right)\tilde\sigma_{kk}^{1/2}(u)\\
&= O_p\left(\sqrt{\log(p/h)/(nh)}+h^2\right).\\
|\psi_{3jkn}(u)| &= o_p\left(\sqrt{\log(p/h)/(nh)}\right).
\end{aligned}$$

Combining these with (A.8) yields (A.9), namely

$$\begin{aligned}
\max_{1\le j,k\le p}\sup_{u\in[a,b]}|\hat\sigma_{jk}(u)-\sigma_{jk}(u)| &\le \max_{1\le j,k\le p}\sup_{u\in[a,b]}\{|\tilde\sigma_{jk}(u)-\sigma_{jk}(u)|+|\psi_{1jn}(u)|\\
&\quad +|\psi_{2jn}(u)|+|\psi_{3jn}(u)|\}\\
&= O_p\left(\sqrt{\frac{\log(p/h_1)}{nh_1}}+\sqrt{\frac{\log(p/h)}{nh}}+h_1^2+h^2\right)\\
&= O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right),
\end{aligned}$$

where the last equation follows from the condition that $h/h_1+h_1/h=O(1)$.

Let $\tilde y_{ij}=(y_{ij}-\hat\mu_j(u_i))/\sqrt{\hat\sigma_{jj}(u_i)}$, $1\le j\le p$. Let

$$\tilde\sigma_{jk}(u)=\sqrt{\hat\sigma_{jj}(u)\hat\sigma_{kk}(u)}\sum_{i=1}^n w_{h_3}(u_i-u)\tilde y_{ij}\tilde y_{ik}, 1\le j,k\le p.$$

If $h/h_3+h_3/h=O(1)$, then using the similar arguments to the above, we can show that

$$\begin{aligned}
&\sum_{i=1}^n w_{h_3}(u_i-u)(\sqrt{\sigma_{jj}(u)\sigma_{kk}(u)/(\sigma_{jj}(u_i)\sigma_{kk}(u_i))}-1)(y_{ij}-\hat\mu_j(u_i))(y_{ik}-\hat\mu_k(u_i))\\
&\quad = O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right).\\
&\sum_{i=1}^n w_{h_3}(u_i-u)|(y_{ij}-\hat\mu_j(u_i))(y_{ik}-\hat\mu_k(u_i))|=O_p(1).
\end{aligned}$$

Therefore, it follows from the previous arguments that

$$\begin{aligned}|\tilde{\sigma}_{jk}(u)-\hat{\sigma}_{jk}(u)| &= |\sum_{i=1}^{n} w_{h_3}(u_i-u)(\sqrt{\hat{\sigma}_{jj}(u)\hat{\sigma}_{kk}(u)/(\hat{\sigma}_{jj}(u_i)\hat{\sigma}_{kk}(u_i))}-1)\\ &\quad \times(y_{ij}-\hat{\mu}_j(u_i))(y_{ik}-\hat{\mu}_k(u_i))|\\ &= |\sum_{i=1}^{n} w_{h_3}(u_i-u)(\sqrt{\sigma_{jj}(u)\sigma_{kk}(u)/(\sigma_{jj}(u_i)\sigma_{kk}(u_i))}-1)\\ &\quad \times(y_{ij}-\hat{\mu}_j(u_i))(y_{ik}-\hat{\boldsymbol{\mu}}_k(u_i))|+O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right)\\ &= O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right),\end{aligned}$$

which implies

$$\tilde{\sigma}_{jk}(u)=\sigma_{jk}(u)+O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right)$$

uniformly for u. Therefore,

$$\begin{aligned}\hat{c}_{ij}(u) &= \frac{\tilde{\sigma}_{jk}(u)}{\sqrt{\hat{\sigma}_{jj}(u)\hat{\sigma}_{kk}(u)}}+O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right)\\ &= c_{ij}(u)+O_p\left(\sqrt{\frac{\log(p/h)}{nh}}+h^2\right)\end{aligned}$$

uniformly for u. The proof is completed.

Proof of Theorem 3.3. Note that $nh^5/\log(p/h)=O(1)$ implies that $h^2=O(\tau_{np})$. Also $h/h_1+h_1/h=O(1)$ and $h/h_3+h_3/h=O(1)$ imply that h, h_1 and h_3 have the same convergence rate as $h,h_1,h_3\to 0$. By definition and Lemma 0.7, the Online Supplementary Material, we have that uniformly in $u \in [a,b]$,

$$
\begin{aligned}
||\hat{\Sigma}^{(st)}(u) - \Sigma(u)|| &= ||\frac{\hat{\beta}^2(u)}{\hat{\delta}_p^2(u)}(< \hat{\Sigma}^{(t)}(u), I_p > I_p - \Sigma(u)) \\
&\qquad + \frac{\hat{\alpha}^2(u)}{\hat{\delta}_p^2(u)}(\hat{\Sigma}^{(t)}(u) - \Sigma(u))|| \\
&\leq \frac{\hat{\beta}^2(u)}{\hat{\delta}_p^2(u)}||(< \Sigma(u), I_p > + O_p(\tau_{np}))I_p - \Sigma(u)|| + \frac{\hat{\alpha}_p^2(u)}{\hat{\delta}_p^2(u)}||\hat{\Sigma}^{(t)}(u) - \Sigma(u)|| \\
&= \frac{\hat{\beta}^2(u)O(1)}{\alpha_p^2(u)(1 + O_p(m_p(u)\tau_{np}/\alpha_p(u)))^2} + O_p(m_p(u)\tau_{np}) \\
&= \frac{O_p(m_p(u)\tau_{np}^2)}{\alpha_p^2(u)\log(p/h)(1 + O_p(m_p(u)\tau_{np}/\alpha_p(u)))^2} + O_p(m_p(u)\tau_{np}) \\
&= O_p(m_p(u)\tau_{np}) + O_p(m_p(u)\tau_{np}) = O_p(m_p(u)\tau_{np}),
\end{aligned}
$$

since $\sup_{u\in[a,b]} m_p(u)\tau_{np}/\alpha_p(u) = o(1)$ and $\tau_{np}/(\log(p/h)\inf_{u\in[a,b]} \alpha_p^2(u)) = O(1)$.

Furthermore, under Condition (C9), we have that uniformly in $u \in [a,b]$,

$$
\begin{aligned}
||\hat{\Sigma}^{(t)}(u)|| &\geq ||\Sigma(u)|| - ||\hat{\Sigma}^{(t)}(u) - \Sigma(u)|| \\
&= ||\Sigma(u)|| - O_p\left(m_p(u)\tau_{np}\right) \\
&\geq s_{0p} - O_p\left(m_p(u)\tau_{np}\right) \\
&= s_{0p}\left(1 - o_p(1)\right). \\
||\hat{\Sigma}^{(st)}(u)\Sigma^{-1}(u) - I_p|| &\leq ||\hat{\Sigma}^{(t)}(u) - \Sigma(u)||(||\Sigma(u)||)^{-1} \\
&\leq O_p\left(m_p(u)\tau_{np}s_{0p}^{-1}\right). \\
||\Sigma(u)(\hat{\Sigma}^{(st)}(u))^{-1} - I_p|| &\leq ||\hat{\Sigma}^{(t)}(u) - \Sigma(u)||(||\Sigma^{(st)}(u)||)^{-1} \\
&\leq O_p\left(m_p(u)\tau_{np}s_{0p}^{-1}\right). \\
||(\hat{\Sigma}^{(t)}(u))^{-1} - \Sigma(u)^{-1}|| &\leq \left(||\hat{\Sigma}^{(t)}(u)||||\Sigma(u)||\right)^{-1}||\hat{\Sigma}^{(t)}(u) - \Sigma(u)|| \\
&= O_p\left(m_p(u)\tau_{np}s_{0p}^{-2}\right).
\end{aligned}
$$

Finally, under Condition (C10), it follows from Lemmas 0.3 and 0.6, the Online Supplementary Material that the above results continue to hold if we replace the thresholding function $t_0(u)$ by $\hat{t}_0(u)$. The proof is completed.

ACKNOWLEDGEMENTS

We thank Dr. Jaideep Oberoi from Kent Business School for helpful discussions on the Capital Asset Pricing Model.

SUPPLEMENTARY MATERIAL

Supplement A: Online Supplementary Material (http://www.e-publications.org/ims/support/dowload/stncmjzlsuppl.zip). The detailed proofs of the lemmas and some extra information on numerical results.

REFERENCES

[1] Bickel, P. and Levina, E. (2008). Covariance regularization by thresholding, *Ann. Stat.*, **36**, 25772604.

[2] Donoho, D., Gavish, M. and Johnstone, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model, *Ann. Stat.*, **46**, 17421778.

[3] Chen, Z. and Leng, C. (2016). Dynamic covariance models. *Jour. Ameri. Stat. Assoc.*, **111**, 1196-1207.

[4] Engle, R.F., Ledoit, O. and Wolf, M. (2017). Large Dynamic Covariance Matrices. *Jour. Busi. & Econ. Stat.*, DOI: 10.1080/07350015.2017.1345683.

[5] Fama, E. and French, K. (2004). The Capital asset pricing model: theory and evidence". *Journal of Economic Perspectives*, **1**8, 2546.

[6] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

[7] Fox, E. and Dunson, D. (2015). Bayesian nonparametric covariance regression. *Journal of Machine Learning Research*, **1**6, 2501-2542.

[8] Lamusa, C., Hämäläinenc, M.S., Temereancaa,S., Browna,E.N., and Purdona, P.L. (2012). A spatiotemporal dynamic distributed solution to the MEG inverse problem. *Neuroimage*, **6**3, 894-909.

[9] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Jour. Multi. Analy.*, **8**8, 365411.

[10] Li, B. and Solea, E. (2017). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *Jour. Ameri. Stat. Assoc.*, DOI: 10.1080/01621459.2017.1356726.

[11] Reich, B.J., Eidsvik, J., Guindani, M., Nail, A.J. and Schmidt, A.M. (2011). A class of covariate-dependent spatiotemporal covariance functions for the analysis of daily ozone concentration. *Ann. Appl. Stat.* , **5**, 24252447.

[12] Yin, J., Geng, Z., Li, R. and Wang, H.(2010). Nonparametric covariance model. *Statistica Sinica*, **2**0, 469-479.

[13] Yuan, M. and Cai, T. (2010). A reproducing kernel Hilbert space approach to functional linear regression, *Ann. Stat.*, **3**8, 34123444.

[14] Zhang, J. and Liu, C. (2015). On linearly constrained minimum variance beamforming. *Journal of Machine Learning Research*, **1**6, 2099-2145.

[15] Zhang, J. and Su, L. (2015). Temporal autocorrelation-based beamforming With MEG neuroimaging data. *Jour. Ameri. Stat. Assoc.*, **1**10, 1375-1388.

School of Mathematics, Statistics and Actuarial Science
University of Kent, Canterbury
Kent CT2 7FS, U.K.
E-mail: jz79@kent.ac.uk
jl705@kent.ac.uk

Table 1
The Average(standard error) of IRSE for Setting 1

n	p	DCM_2	DCM_1	$_{t}NCM$	$_{st}NCM$
			$\rho = 0$		
100	50	5.1307(0.2504)	0.5807(0.0348)	0.4546(0.0348)	0.4482(0.0361)
	100	16.2142(0.4702)	0.6263(0.0256)	0.4962(0.0254)	0.4878(0.0264)
	150	49.4335(0.7562)	0.6497(0.0212)	0.5199(0.0209)	0.5112(0.0219)
	300	78.0434(0.4833)	0.7045(0.0171)	0.5739(0.0151)	0.5654(0.0156)
	500	102.8816(0.3913)	0.7521(0.0172)	0.6111(0.0133)	0.6021(0.0138)
200	50	2.8919(0.0826)	0.3650(0.0269)	0.2821(0.0242)	0.2834(0.0253)
	100	8.8372(0.1562)	0.3868(0.0179)	0.2976(0.0156)	0.2989(0.0161)
	150	18.5651(0.3007)	0.3915(0.0164)	0.3031(0.0150)	0.3040(0.0155)
	300	70.8479(0.3201)	0.4194(0.0126)	0.3201(0.0100)	0.3204(0.0104)
	500	84.8626(0.2537)	0.4508(0.0108)	0.3315(0.0092)	0.3310(0.0093)
500	50	1.5780(0.0356)	0.2025(0.0128)	0.1814(0.0113)	0.1831(0.0119)
	100	3.3680(0.0440)	0.2071(0.0093)	0.1822(0.0074)	0.1840(0.0078)
	150	6.0579(0.0645)	0.2108(0.0081)	0.1827(0.0061)	0.1845(0.0064)
	300	28.7062(0.2438)	0.2295(0.0052)	0.1838(0.0042)	0.1856(0.0044)
	500	90.9963(0.2061)	0.2519(0.0041)	0.1845(0.0036)	0.1863(0.0038)
			$\rho = 0.3$		
100	50	5.7885(0.2845)	0.6171(0.0345)	0.4934(0.0292)	0.4816(0.0298)
	100	18.6343(0.5494)	0.6686(0.0301)	0.5375(0.0245)	0.5263(0.0248)
	150	55.3409(0.7486)	0.7056(0.0284)	0.5627(0.0212)	0.5512(0.0223)
	300	80.5517(0.4835)	0.7648(0.0206)	0.6157(0.0162)	0.6047(0.0165)
	500	102.5949(0.5096)	0.8198(0.0295)	0.6532(0.0108)	0.6422(0.0110)
200	50	3.0460(0.0937)	0.3980(0.0251)	0.3079(0.0212)	0.3069(0.0218)
	100	8.1745(0.1568)	0.4188(0.0184)	0.3239(0.0151)	0.3229(0.0154)
	150	17.6292(0.2808)	0.4274(0.0180)	0.3310(0.0147)	0.3294(0.0147)
	300	72.9782(0.3237)	0.4608(0.0152)	0.3533(0.0101)	0.3510(0.0103)
	500	93.0913(0.2922)	0.4972(0.0121)	0.3697(0.0083)	0.3671(0.0086)
500	50	1.5613(0.0410)	0.2161(0.0123)	0.1902(0.0097)	0.1915(0.0104)
	100	3.3541(0.0484)	0.2191(0.0102)	0.1904(0.0075)	0.1917(0.0077)
	150	6.4276(0.0660)	0.2250(0.0095)	0.1918(0.0072)	0.1930(0.0073)
	300	27.2019(0.2476)	0.2416(0.0060)	0.1932(0.0048)	0.1946(0.0049)
	500	92.5289(0.2549)	0.2630(0.0043)	0.1932(0.0038)	0.1946(0.0040)
			$\rho = 0.8$		
100	50	5.6548(0.3649)	1.3968(0.1261)	1.3139(0.1178)	1.1319(0.1038)
	100	18.0835(0.5143)	1.8898(0.1264)	1.7843(0.1177)	1.4853(0.1051)
	150	55.7814(0.6103)	2.3268(0.1293)	2.1991(0.1225)	1.8063(0.1094)
	300	80.8167(0.4157)	3.2234(0.1261)	3.0622(0.1199)	2.4755(0.1086)
	500	102.7016(0.4263)	4.1306(0.1081)	3.9262(0.1021)	3.1492(0.0936)
200	50	3.0830(0.1428)	1.1094(0.0783)	1.0305(0.0709)	0.9078(0.0620)
	100	8.1824(0.1964)	1.5004(0.0731)	1.4002(0.0677)	1.1875(0.0595)
	150	17.3394(0.3018)	1.8256(0.0669)	1.7130(0.0620)	1.4254(0.0556)
	300	73.0011(0.3285)	2.5373(0.0588)	2.4022(0.0554)	1.9570(0.0504)
	500	93.1479(0.2667)	3.2475(0.0685)	3.0812(0.0645)	2.4830(0.0583)
500	50	1.6169(0.0621)	0.7379(0.0405)	0.6859(0.0319)	0.6324(0.0270)
	100	3.3779(0.0624)	1.0280(0.0319)	0.9495(0.0279)	0.8410(0.0240)
	150	6.4755(0.0886)	1.2453(0.0295)	1.1566(0.0280)	1.0007(0.0246)
	300	26.5926(0.2880)	1.7210(0.0300)	1.6184(0.0284)	1.3545(0.0253)
	500	92.7346(0.2502)	2.1956(0.0268)	2.0834(0.0254)	1.7137(0.0222)

Table 2

The Average(standard error) of IRSE for Setting 2

n	p	DCM_2	DCM_1	$_{t}NCM$	$_{st}NCM$
		$\rho = 0$			
100	50	11.8865(0.2591)	0.5261(0.0194)	0.4534(0.0229)	0.4338(0.0232)
	100	37.3349(1.2667)	0.5494(0.0144)	0.4834(0.0148)	0.4642(0.0161)
	150	62.9012(0.5970)	0.5609(0.0149)	0.4980(0.0161)	0.4797(0.0168)
	300	87.7651(0.5914)	0.5902(0.0183)	0.5248(0.0107)	0.5077(0.0113)
	500	114.8459(0.5070)	0.6096(0.0171)	0.5449(0.0071)	0.5286(0.0076)
200	50	6.1169(0.2148)	0.3867(0.0186)	0.3177(0.0161)	0.3095(0.0164)
	100	16.9859(0.2304)	0.3995(0.0114)	0.3277(0.0110)	0.3195(0.0111)
	150	35.6152(0.5072)	0.4049(0.0116)	0.3336(0.0096)	0.3253(0.0096)
	300	90.4787(0.4491)	0.4235(0.0105)	0.3447(0.0080)	0.3364(0.0082)
	500	116.3211(0.3969)	0.4442(0.0089)	0.3529(0.0071)	0.3446(0.0071)
500	50	2.8192(0.0788)	0.2626(0.0089)	0.2367(0.0094)	0.2341(0.0102)
	100	7.2953(0.0871)	0.2672(0.0049)	0.2420(0.0051)	0.2395(0.0056)
	150	13.8324(0.0962)	0.2708(0.0044)	0.2453(0.0035)	0.2430(0.0038)
	300	84.4015(0.8099)	0.2825(0.0030)	0.2478(0.0024)	0.2457(0.0025)
	500	117.4588(0.2618)	0.2974(0.0023)	0.2488(0.0017)	0.2469(0.0018)
		$\rho = 0.3$			
100	50	11.9070(0.2632)	0.5488(0.0229)	0.4846(0.0300)	0.4628(0.0280)
	100	37.5512(1.3329)	0.5643(0.0164)	0.5049(0.0165)	0.4850(0.0171)
	150	62.9160(0.5414)	0.5831(0.0159)	0.5190(0.0136)	0.4999(0.0145)
	300	87.6334(0.5364)	0.6114(0.0187)	0.5456(0.0083)	0.5280(0.0093)
	500	114.8919(0.5994)	0.6266(0.0222)	0.5603(0.0071)	0.5433(0.0074)
200	50	6.2194(0.2103)	0.4067(0.0168)	0.3343(0.0171)	0.3243(0.0163)
	100	16.9228(0.2161)	0.4186(0.0161)	0.3465(0.0161)	0.3364(0.0155)
	150	35.6480(0.5216)	0.4275(0.0114)	0.3534(0.0093)	0.3434(0.0089)
	300	90.4008(0.4407)	0.4491(0.0117)	0.3677(0.0080)	0.3575(0.0079)
	500	116.3462(0.3737)	0.4703(0.0121)	0.3787(0.0078)	0.3688(0.0078)
500	50	2.8119(0.0847)	0.2714(0.0083)	0.2456(0.0077)	0.2423(0.0081)
	100	7.3496(0.0772)	0.2751(0.0065)	0.2495(0.0053)	0.2467(0.0054)
	150	13.7107(0.0975)	0.2798(0.0053)	0.2519(0.0040)	0.2492(0.0040)
	300	84.6477(0.5483)	0.2909(0.0037)	0.2536(0.0029)	0.2514(0.0029)
	500	117.4813(0.2674)	0.3046(0.0029)	0.2539(0.0022)	0.2517(0.0022)
		$\rho = 0.8$			
100	50	11.8065(0.2964)	1.3110(0.1046)	1.2273(0.0932)	1.0305(0.0807)
	100	37.5451(1.5225)	1.7824(0.1125)	1.6758(0.1054)	1.3707(0.0942)
	150	62.8919(0.5696)	2.1604(0.1116)	2.0378(0.1050)	1.6487(0.0943)
	300	87.5892(0.5203)	3.0189(0.0912)	2.8635(0.0870)	2.2906(0.0794)
	500	114.9517(0.5658)	3.8808(0.0967)	3.6842(0.0939)	2.9324(0.0865)
200	50	6.0032(0.2641)	1.0720(0.0562)	0.9904(0.0528)	0.8423(0.0479)
	100	16.6911(0.2923)	1.4519(0.0547)	1.3521(0.0527)	1.1169(0.0470)
	150	36.6918(0.6470)	1.7727(0.0591)	1.6604(0.0557)	1.3564(0.0507)
	300	90.4006(0.4564)	2.4649(0.0491)	2.3305(0.0462)	1.8768(0.0421)
	500	116.2670(0.3876)	3.1504(0.0599)	2.9914(0.0571)	2.3935(0.0523)
500	50	2.7977(0.1020)	0.7284(0.0324)	0.6723(0.0267)	0.5950(0.0235)
	100	7.1270(0.1176)	1.0120(0.0268)	0.9328(0.0250)	0.7952(0.0227)
	150	13.3994(0.1379)	1.2240(0.0246)	1.1367(0.0235)	0.9520(0.0215)
	300	82.8582(0.8672)	1.6915(0.0256)	1.5929(0.0242)	1.3054(0.0220)
	500	117.4724(0.2692)	2.1614(0.0218)	2.0525(0.0209)	1.6643(0.0185)

TABLE 3
The Average(standard error) of IRSE for Setting 3

n	p	DCM_2	DCM_1	${}_{t}NCM$	${}_{st}NCM$
			$\rho = 0$		
100	50	7.2002(0.4009)	3.5156(0.1637)	3.0263(0.1351)	2.9973(0.1418)
	100	16.4443(0.4565)	3.8041(0.0726)	3.2530(0.0551)	3.2200(0.0589)
	150	50.7710(0.5789)	3.9366(0.0649)	3.3912(0.0414)	3.3586(0.0432)
	300	78.2423(0.6810)	4.1044(0.0624)	3.5757(0.0297)	3.5416(0.0292)
	500	102.9735(0.6949)	4.2842(0.0793)	3.7076(0.0220)	3.6691(0.0218)
200	50	4.7672(0.2241)	2.7869(0.1292)	2.3683(0.0990)	2.3456(0.1021)
	100	10.3670(0.2726)	2.9617(0.0963)	2.5836(0.0553)	2.5661(0.0574)
	150	16.7314(0.2515)	3.5027(0.0579)	2.7180(0.0450)	2.6976(0.0464)
	300	71.1626(0.4656)	3.7042(0.0368)	2.9883(0.0297)	2.9567(0.0305)
	500	85.0637(0.4323)	3.9768(0.0448)	3.1697(0.0212)	3.1307(0.0217)
500	50	3.0828(0.0990)	2.0581(0.0636)	1.6115(0.0619)	1.6079(0.0634)
	100	5.4027(0.1283)	2.1808(0.0577)	1.8712(0.0435)	1.8677(0.0443)
	150	7.9711(0.0929)	2.2271(0.0473)	2.0339(0.0333)	2.0311(0.0341)
	300	20.0680(0.1489)	2.3369(0.0375)	2.3138(0.0221)	2.3083(0.0226)
	500	90.3515(0.2678)	3.2536(0.0129)	2.5211(0.0175)	2.5125(0.0179)
			$\rho = 0.3$		
100	50	7.9703(0.4326)	3.5611(0.1876)	3.0472(0.1133)	3.0116(0.1188)
	100	18.1734(0.4703)	3.8359(0.0843)	3.2654(0.0516)	3.2310(0.0551)
	150	56.4342(0.6837)	3.9699(0.0778)	3.4091(0.0433)	3.3762(0.0461)
	300	80.4879(0.7218)	4.1319(0.0747)	3.5784(0.0285)	3.5443(0.0284)
	500	102.5341(0.7728)	4.2849(0.0987)	3.6996(0.0233)	3.6641(0.0228)
200	50	4.7921(0.2095)	2.8340(0.1146)	2.3977(0.0845)	2.3708(0.0879)
	100	9.8312(0.2352)	3.0478(0.0996)	2.6016(0.0575)	2.5786(0.0594)
	150	16.2014(0.2387)	3.5476(0.0640)	2.7332(0.0492)	2.7097(0.0505)
	300	73.1892(0.4665)	3.7509(0.0418)	2.9873(0.0307)	2.9555(0.0313)
	500	93.0226(0.4103)	3.9992(0.0487)	3.1546(0.0272)	3.1171(0.0280)
500	50	2.9875(0.1213)	2.0968(0.0805)	1.6410(0.0565)	1.6351(0.0576)
	100	5.3832(0.1591)	2.2157(0.0610)	1.8861(0.0443)	1.8810(0.0458)
	150	7.9702(0.0786)	2.2675(0.0499)	2.0482(0.0352)	2.0437(0.0360)
	300	18.9870(0.1429)	2.3773(0.0436)	2.3161(0.0238)	2.3101(0.0243)
	500	91.6005(0.3058)	3.2760(0.0179)	2.5101(0.0164)	2.5011(0.0167)
			$\rho = 0.8$		
100	50	8.2587(0.5799)	4.6369(0.5098)	4.1809(0.3017)	3.7168(0.1794)
	100	18.1379(0.4944)	5.9809(0.4096)	5.2965(0.3118)	4.4277(0.1898)
	150	50.7099(6.2867)	7.1100(0.5062)	6.2984(0.3498)	5.0698(0.2249)
	300	78.9593(0.6606)	9.3633(0.3683)	8.4308(0.3441)	6.4292(0.2322)
	500	101.1399(0.6726)	11.5664(0.3317)	10.5839(0.2916)	7.8342(0.2232)
200	50	5.5375(0.3775)	3.8373(0.3996)	3.5263(0.2055)	3.1336(0.1301)
	100	10.3672(0.2550)	4.8094(0.4444)	4.5007(0.1761)	3.7962(0.1082)
	150	16.4920(0.2807)	6.2207(0.4457)	5.3075(0.1538)	4.3213(0.0967)
	300	70.7900(0.4641)	8.2310(0.2065)	7.0400(0.1583)	5.4331(0.1133)
	500	91.2056(0.3782)	10.1027(0.2081)	8.8042(0.1784)	6.5836(0.1320)
500	50	3.4523(0.2160)	2.6353(0.2005)	2.4554(0.1428)	2.3046(0.1097)
	100	6.0072(0.1610)	2.9998(0.2064)	3.2643(0.0977)	2.9233(0.0564)
	150	8.2472(0.1143)	3.3002(0.1875)	3.8406(0.0702)	3.3285(0.0437)
	300	19.1704(0.1533)	4.2317(0.3624)	5.0320(0.0766)	4.1276(0.0525)
	500	40.0499(0.3396)	7.2277(0.0840)	6.2086(0.0710)	4.9125(0.0487)

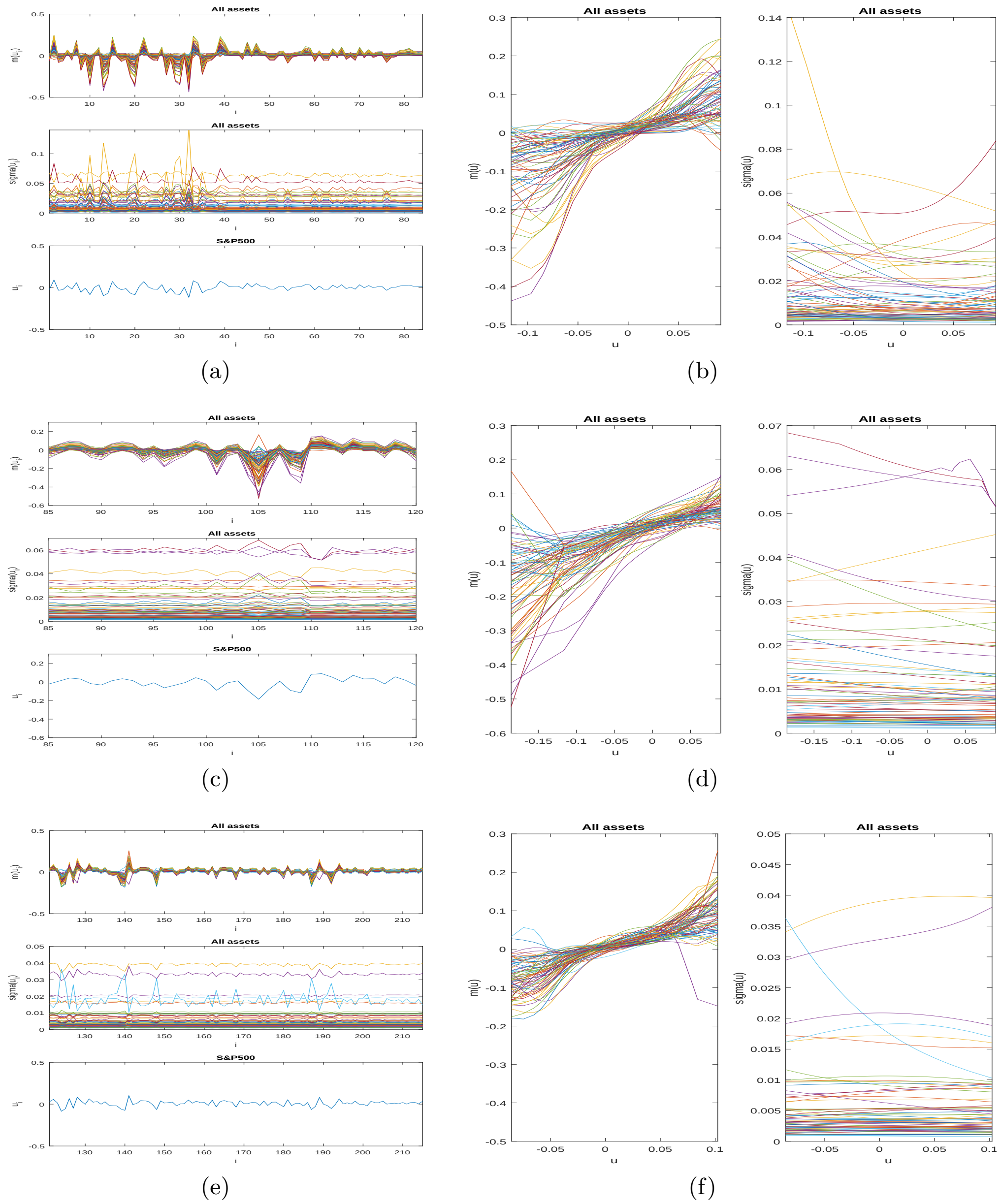

Fig 1: Before-financial-crisis: (a) Plots of estimated means $\hat{\mu}_k(u_i)$ against i (top), estimated individual volatility $\hat{\sigma}_{kk}(u_i)$ against i (middle) and u_i against i (bottom). (b) Plots of estimated $\hat{\mu}_k(u)$ against u (left) and estimated individual volatility $\hat{\sigma}_{kk}(u)$ against u right. Similarly, (c) and (d) for the in-financial-crisis period while (e) and (f) for the after-financial-crisis.