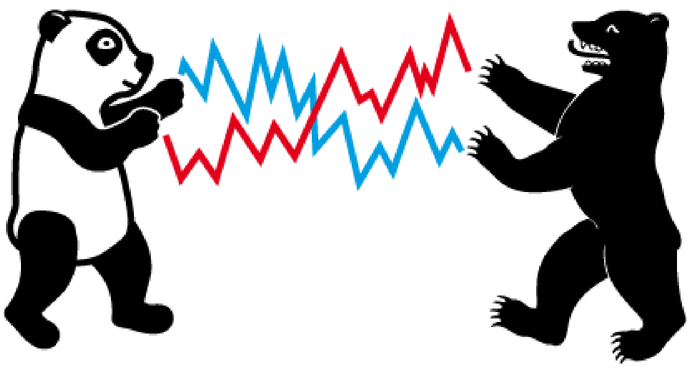
IRTG 1792 Discussion Paper 2020-024

Dynamic Spatial Network Quantile Autoregression

Xiu Xu [*]
Weining Wang [*2]
Yongcheol Shin [*2]



[*] Soochow University, China
[*2] University of York, UK

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Dynamic Spatial Network Quantile Autoregression [*]

Xiu Xu [†] Weining Wang [‡] Yongcheol Shin [§]

Abstract

This paper proposes a dynamic spatial autoregressive quantile model. Using predetermined network information, we study dynamic tail event driven risk using a system of conditional quantile equations. Extending Zhu, Wang, Wang and Härdle (2019), we allow the contemporaneous dependency of nodal responses by incorporating a spatial lag in our model. For example, this is to allow a firm's tail behavior to be connected with a weighted aggregation of the simultaneous returns of the other firms. In addition, we control for the common factor effects. The instrumental variable quantile regressive method is used for our model estimation, and the associated asymptotic theory for estimation is also provided. Simulation results show that our model performs well at various quantile levels with different network structures, especially when the node size increases. Finally, we illustrate our method with an empirical study. We uncover significant network effects in the spatial lag among financial institutions.

JEL classification: C32, C51, G17

Keywords: Network, Quantile autoregression, Instrumental variables, Dynamic models.

[*]Xiu Xu acknowledges the support of the Natural Science Foundation of China (Grant No. 71803140).

[†]Dongwu Business School, Soochow University, 50 Donghuan Road, Suzhou, Jiangsu 215021, PR China. Email: xiux@suda.edu.cn

[‡]Department of Economics, University of York. Email: weining.wang@york.ac.uk

[§]Department of Economics, University of York. Email: yongcheol.shin@york.ac.uk

1 Introduction

Quantifying network effect and tail dependence is important for studying financial contagion and systemic risk, see (Acemoglu et al., 2015; Fafchamps and Gubert, 2007; Härdle et al., 2016; Betz et al., 2016; Hautsch et al., 2015). Quantile regression is a powerful tool for characterizing the heterogeneous impact in policy analysis and measuring dynamic tail risk. However, in most complex financial systems, multiple entities are often intertwined and interact with each other, which is represented as networks (Hautsch et al., 2015; Härdle et al., 2016; Chen et al., 2019). On the other hand, in practice, the endogeneity issue commonly leads to inconsistent estimates in conventional quantile regression (Wüthrich, 2019). Our interest lies in studying the complex tail dependency structure in the context of endogeneity and dynamic networks with a large number of nodes.

In this paper we propose a dynamic spatial quantile regression model to study financial markets. By extending the instrumental variable quantile regression (IVQR) model (Chernozhukov and Hansen, 2006), we focus on capturing simultaneous effects for network nodes, along with lagged network effects and exogenous common shocks. Specifically, in order to cope with the endogeneity problem arising from incorporating simultaneous network effects, we extend IVQR estimation for modeling dynamic spatial data. Our empirical application concerning the financial contagions in the US stock market finds that stock returns are overwhelmingly affected by its "peers" (returns of other stocks in the same period) controlling for a few other factors. Meanwhile, the effects of common factors on all response variables in a network structure are non-negligible.

Quantile regression in time series has been of particular interest in the existing literature since a seminal work by Bassett and Koenker (1978). Engle and Manganelli (2004) propose a conditional autoregressive value at risk (CAViaR) model, which specifies the evolution of value-at-risk (VaR) over time using an autoregressive process. Koenker and Xiao (2006) consider a quantile autoregressive method to model the conditional quantile function, which allows an examination of asymptotic properties of the underlying process. Galvao Jr et al. (2013) study quantile regression in an autoregressive dynamic framework with exogenous stationary covariates. Our methodology features a dynamic tail dependency graph within a network framework, which is important in modeling financial networks, see e.g. Diebold and Yilmaz (2014).

Furthermore, in the network and spatial literature, Zhu et al. (2017) develop a network autoregression modeling framework at the mean level. There are many extensions of network models in this direction. For example, Zhu et al. (2018) consider a multivariate spatial autoregression model; Zhu, Chang, Li and Wang (2019) investigate a screening method to select influential nodes; and Zhu et al. (2020) study nonconvex penalized estimation methods in high-dimensional vector autoregression models. Zhu and Pan (2018) extend the network vector autoregression model to allow for group-specific parameters and enhance the model's flexibility. Zhu, Wang, Wang and Härdle (2019) propose a network quantile autoregression (NQAR) model to study conditional quantiles. In this research we extend the NQAR model to characterize the simultaneous effects and cross-sectional financial tail risk dependence in complex financial networks. We utilize the IVQR approach by Chernozhukov and Hansen (2006) to cope with the endogeneity issue due to simultaneous spatial items, motivated by Su and Yang (2011), who extend the IVQR model to non-iid data with a correctly specified linear spatial autoregressive model and nonstochastic regressors. Our work can be seen as an extension of Su and Yang (2011) to a dynamic model framework. However, the technical assumptions are substantially different as we assume near epoch dependence of the underlying processes.

There are also many studies that cope with the endogeneity issue by using IVQR estimation, see e.g. Wüthrich (2019, 2020); Machado and Silva (2019). Frölich and Melly (2013) use an instrumental variable (IV) to solve for the endogeneity of the binary treatment variable. Su and Hoshino (2016) consider sieve instrumental variable quantile regression estimation of functional coefficient models. On the other hand, our paper is also closely related to literature on tail dependence in a complex financial system, see e.g. Hautsch et al. (2014); Härdle et al. (2016). In a multivariate quantile context, White et al. (2015) propose an innovative vector autoregressive model for quantile dynamics. Compared with the existing literature, our approach is different mainly in the following three aspects: First, the proposed model admits cross-section dependence in quantile dynamics, which facilitates investigating the simultaneous effects. Second, it embeds an observed network structure, which provides a parametric estimation framework for modeling a large number of nodes along with controlling for the observed nodal heterogeneity.

Third, the model allows for exogenous common covariates to affect the tail dependence of the process, which accounts for the effects under various economic environments.

Finally, our main contributions are summarized as follows: First, we propose a dynamic network quantile model to characterize the cross-sectional and temporal tail dependence, which incorporates valuable predetermined network information. Second, we study the asymptotic properties for both the underlying process and the estimated IVQR parameters. Moreover, detailed conditions on the network structures are derived to ensure the consistency and the asymptotic normality of the estimator. Lastly, when applying our model in the US financial market to investigate financial risk contagion, we find that the simultaneous network effects are dominated for different network structures. This implies that one cannot neglect the contemporaneous effects in tail dependence.

This paper is structured as follows: In Section 2 we introduce the model and discuss the stationarity of the data- generating processes as well as the asymptotic stationary distribution of the average. Section 3 introduces the IVQR estimation and the corresponding asymptotic properties are presented in Section 3.2. Simulation results with different network scenarios are illustrated in Section 4. Finally, we implement our model in a US financial market risk contagion application in Section 5. Section 6 concludes our paper. All the related proofs can be found in Appendix 7.

Notations: For a constant $k > 0$ and a vector $v = (v_1, \dots, v_d)^\top \in \mathbb{R}^d$, we denote $|v|_k = (\sum_{i=1}^d |v_i|^k)^{1/k}$, $|v| = |v|_2$ and $|v|_\infty = \max_{i \le d} |v_i|$. For a matrix $A = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$, we define the spectral norm $|A|_2 = \max_{|v|=1} |Av|$ and the max norm $|A|_{\max} = \max_{i,j} |a_{i,j}|$. We write $a_n = O(b_n)$ or $a_n \lesssim b_n$ if there exists a positive constant C such that $a_n/b_n \leqslant C$ for all large n, and we denote $a_n = o(b_n)$ (resp. $a_n \sim b_n$), if $a_n/b_n \to 0$ (resp. $a_n/b_n \to 1$). For two sequences of random variables (X_n) and (Y_n), we write $X_n = o_p(Y_n)$, if $X_n/Y_n \to 0$ in probability. Let I_N or I be the $N \times N$ identity matrix. $\mathbf{I}(\cdot)$ is the indicator function. Denote $|.|_2$ or $||.||$ as the L_2 norm of a matrix or vector. $\mathbf{1}_N$ denotes a $N \times 1$ vector with each element as one. $\mathbb{N}$ represents the integer set. For any $n \times m$ matrix A with element a_{ij}, define the column- and the row-sum by $\|A\|_1 = \max\limits_{1 \leqslant j \leqslant m} \sum\limits_{i=1}^n |a_{ij}|$ and $\|A\|_\infty = \max\limits_{1 \leqslant i \leqslant n} \sum\limits_{j=1}^m |a_{ij}|$. For any random vector X, denote $\|X\|_p = (\mathsf{E}\,|X|^p)^{1/p}$ as its L_p-norm if the absolute pth moment exists.

2 The model

Consider the network quantile autoregressive distributed lag model (NQADL):

$$Y_{it} = \gamma_0(U_{it}) + \sum_{l=1}^{q} Z_{il}\alpha_l(U_{it}) + \gamma_1(U_{it})n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{jt} \\ + \gamma_2(U_{it})n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{j,t-1} + \gamma_3(U_{it})Y_{i,t-1} + \sum_{k=0}^{p} F_{t-k}^{\top}\beta_k(U_{it}) \tag{1}$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where $\{U_{it}\}_{i,t}$ is a sequence of i.i.d. uniform random variables on the set $[0, 1]$ such that $\gamma_j(.)s(j = 0, 1, 2, 3)$, $\alpha_l(.)s(l = 1, 2, \cdots, q)$, and $\beta_k(.)s(k = 0, 1, \cdots, p)$ are unknown parameter functions from $[0, 1]$ to $\mathbb{R}$. In particular, $\gamma_1(.)$ measures the contemporaneous network effect, and $\gamma_1(.)$ measures the lagged network effect. $Z_i = (Z_{i1}, \cdots, Z_{iq})^{\top} \in \mathbb{R}^q$ is a $q \times 1$ vector of time-invariant node-specific covariates, and $F_t \in \mathbb{R}^m$ is an $m \times 1$ vector of common covariates that captures the systemic influences on response variables. a_{ij} is the (i, j) element of the adjacency matrix $W \in \mathbb{R}^{N\times N}$ with $a_{ii} = 0$ (not self-connection), where $a_{ij} = 1$ if there is an edge from node i to j while $a_{ij} = 0$ otherwise, and $n_i = \sum_j a_{ij}$ is the out-degree for the ith node.

Assuming that the right hand side of (1) is monotonically increasing in U_{it}, we can write the τ-th conditional quantile function of Y_{it} given the information set $\mathcal{F}_t = \{Z_i, \mathbb{Y}_{t-1}, \mathbb{Y}_t, F_t\}$ as

$$Q_{Y_{it}}(\tau|\mathcal{F}_t) = \gamma_0(\tau) + \sum_{l=1}^{q} Z_{il}\alpha_l(\tau) + \gamma_1(\tau)n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{jt} \\ + \gamma_2(\tau)n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{j,t-1} + \gamma_3(\tau)Y_{i,t-1} + \sum_{k=0}^{p} F_{t-k}^{\top}\beta_k(\tau). \tag{2}$$

First, $\gamma_0(\tau) + \sum_{l=1}^{q} Z_{il}\alpha_l(\tau)$ is the (time-invariant) nodal impact of node i, where $\gamma_0(\tau)$ is the baseline function and Z_{il}s is assumed to be independent of $U_{it}s$. Next, network interactions between nodes (e.g., firms with common shareholders), both contemporaneously and with lag, are captured via the network variables, $n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{jt}$ and $n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{j,t-1}$. Then, $\gamma_1(\tau)$ captures the (simultaneous) structural network effects while $\gamma_2(\tau)$ measures the lagged network function. If the momentum function $\gamma_3(\tau)$ is statistically significant, this points to the usual temporal dynamics for the same node. Furthermore, we control for the impacts of the observed common factors, i.e. F_t with lags which can mitigate any remaining common shock effect in the data. For instance, if Y_{it} denotes stock returns for a large number of firms, we accommodate common macroeconomic and financial variables such as interest rate, inflation rate, the market index return, book-to-market ratio, price-earnings ratio, market volatility and so on, see e.g. Zhu, Wang, Wang and Härdle (2019); Galvao Jr et al. (2013).

To facilitate further developments we introduce convenient notations. Let $\mathbb{Y}_t = (Y_{1t}, \cdots, Y_{Nt})^\top \in \mathbb{R}^N$, $\mathbb{Z} = (Z_1, \cdots, Z_N)^\top \in \mathbb{R}^{N\times q}$, $\mathbb{F}_t = (F_t^\top, \cdots, F_{t-p}^\top)^\top \in \mathbb{R}^{(p+1)m}$. Define $\mathbf{A}_{0t} = (\gamma_0(U_{it}) + \sum_{l=1}^q Z_{il}\alpha_l(U_{it}), 1 \leqslant i \leqslant N)^\top \in \mathbb{R}^N$, $\mathbf{A}_{1t} = diag\{\gamma_1(U_{it}), 1 \leqslant i \leqslant N\}^\top \in \mathbb{R}^{N\times N}$, $\mathbf{A}_{2t} = diag\{\gamma_2(U_{it}), 1 \leqslant i \leqslant N\}^\top \in \mathbb{R}^{N\times N}$, $\mathbf{A}_{3t} = diag\{\gamma_3(U_{it}), 1 \leqslant i \leqslant N\}^\top \in \mathbb{R}^{N\times N}$, $\mathbf{B}_t = (\beta_0^\top(U_{it}), \cdots, \beta_p^\top(U_{it})) \in \mathbb{R}^{N\times(p+1)m}$, and $\Gamma = (\mathbf{1}_N\gamma_0(\tau) + \sum_{l=1}^q Z_{il}\alpha_l(\tau), 1 \leqslant i \leqslant N)^\top \in \mathbb{R}^N$.

Then, the NQADL model (1) can be written in the matrix form as

$$\mathbb{Y}_t = \Gamma + \mathbf{A}_{1t}W\mathbb{Y}_t + \mathbf{H}_t\mathbb{Y}_{t-1} + \mathbf{B}_t\mathbb{F}_t + V_t \tag{3}$$

where $\Gamma = \mathsf{E}(\mathbf{A}_{0t})$, $W = (w_{ij}) = (n_i^{-1}a_{ij}) \in \mathbb{R}^{N\times N}$ is the row-normalized adjacency matrix, $\mathbf{H}_t = \mathbf{A}_{2t}W + \mathbf{A}_{3t} \in \mathbb{R}^{N\times N}$, and $V_t = \mathbf{A}_{0t} - \Gamma \in \mathbb{R}^N$ is independent identically distributed (iid) over t with mean $\mathbf{0}$ and covariance $\Sigma_V = \sigma_V^2 I_N \in \mathbb{R}^{N\times N}$.

The nontrivial issue for the estimation is that the endogeneity caused by contemporaneous network spillovers across nodes renders the standard ordinary least squares (OLS) quantile estimator to be inconsistent. In the conditional mean regression, Quasi-Maximum Likelihood (QML) techniques, based upon a data transformation removing the endogeneity, have been developed by Cliff and Ord (1981), and the asymptotic properties are studied rigorously in Lee (2004). The evaluation of the QML requires calculation of a Jacobian matrix which grows with the cross-section dimension. For applications in which the number of network nodes is large, however, the computational cost can be prohibitive. This problem is exacerbated by heterogeneity across quantiles. In this regard Chernozhukov and Hansen (2005) propose the instrumental variable (IV) approach to estimating quantile treatment effects in the presence of endogeneity. Chernozhukov and Hansen (2006) and Chernozhukov and Hansen (2008) develop the robust inference. Here, we follow the IV approach to cope with the endogeneity.

2.1 Stationarity

In this section we derive the stationarity conditions for the dependent variable in (3), and present the asymptotic distribution of $\mathbb{Y}_t$. Recall that $|A|_2 = \sup_{\{v\in\mathbb{R}^d, v\neq 0\}} |Av|_2/|v|_2$, where $|.|_2$ is the two norm of a vector or matrix. Then, we make the following assumptions:

Assumption 2.1. *(1) Let* $|W|_2 \leqslant 1$ *and* $\Upsilon = \sup_u |\gamma_1(u)| \leqslant c_1 < 1$, *where* W *is a row normalized network matrix. Assume that* $U_{it}s$ *are iid over* i *and* t, *and* F_ts *are iid.* $c_{23}(2)\max_i |\gamma_2(U_{it})| + \max_i |\gamma_3(U_{it})| \leqslant c_{23} < 1$, *and* $c_1 + c_{23} < 1$, *where* c_1, c_{23} *are positive constants.*

(3) $\max_i |\gamma_0(U_{it})| + \max_i \sum_{l=1}^q |Z_{il}||\alpha_l(U_{it})| \leqslant d_z$. $|B_t|_\infty |F_t|_1 \leqslant d_f$, *where* d_z *and* d_f *are assumed to be random variables with bounded moments. Let* $S_t = I - \mathbf{A}_{1t}W$ *and* $\mathbb{D}_t = S_t^{-1}(B_t\mathbb{F}_t + \mathbf{A}_{0t})$ *with* $\mathbb{D} = \mathsf{E}\,\mathbb{D}_t$ *and the elementwise maximum value* $\mathbb{D}_{max} < \infty$. *Then,* $|\mathsf{E}\{vec(\mathbb{D}_{t-l_1}\mathbb{D}_{t-l_1}^\top)\}|_\infty \leqslant \sigma_{d\max} < \infty$, *where* $l_1 = (0, 1, \cdots, t-1)$.

(4) The right hand side of the model (1) *is monotonically increasing in* U_{it}.

Assumption 2.1(1) assures the invertibility of the random matrix $S_t = I - \mathbf{A}_{1t}W$. Then, the model (3) has a unique solution if and only if every principal minor of $I - \gamma_1(U_{it})W$ is positive, which is met by Assumption 2.1(1). But, it is a sufficient but not necessary condition for the existence of a unique solution. Assumption 2.1(2) is necessary to obtain the strict stationarity of $\{\mathbb{Y}_t\}_t$. Under Assumptions 2.1(2) and (3), the covariance stationarity can be achieved. We can relax the factor process to be weakly dependent, however, studying factor processes is not our main theoretical focus. Then we have the following lemma concerning the stationarity.

Lemma 2.1. *(i) Under Assumption 2.1, the process* $\{\mathbb{Y}_t\}_t$ *is strictly stationary.*

(ii)Under Assumption 2.1, $\{\mathbb{Y}_t\}_t$ *is also covariance stationary.*

The detailed discussions of strict stationarity are found in Appendix 7.1. Once $\mathbb{Y}_t$ is strictly stationary, additionally if $\mathsf{Var}(\mathbb{Y}_t)$ and $\Gamma_l = \mathrm{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-l})$ exist, then $\mathbb{Y}_t$ is covariance stationary. The model (3) can also be written as

$$\begin{aligned}\mathbb{Y}_t &= S_t^{-1}\left(H_t\mathbb{Y}_{t-1} + B_t\mathbb{F}_t + \mathbf{A}_{0t}\right) \\ &= S_t^{-1}H_t\mathbb{Y}_{t-1} + S_t^{-1}B_t\mathbb{F}_t + S_t^{-1}\mathbf{A}_{0t}\end{aligned} \tag{4}$$

where $S_t = I - \mathbf{A}_{1t}W$. Then, we have the following covariance stationary solution:

$$\mathbb{Y}_t = \sum_{l=0}^{\infty}\Pi_l\mathbb{D}_{t-l} = \sum_{l=0}^{\infty}\Pi_l S_{t-l}^{-1}B_{t-l}\mathbb{F}_{t-l} + \sum_{l=0}^{\infty}\Pi_l S_{t-l}^{-1}\mathbf{A}_{0t}, \tag{5}$$

where $\mathbb{D}_t = S_t^{-1}(B_t\mathbb{F}_t + \mathbf{A}_{0t})$, $M_t = S_t^{-1}H_t$ and $\Pi_l = M_t \times \cdots \times M_{t-l+1}$ for $l > 1$ with $\Pi_0 = I$ and $\Pi_1 = M_t$. In Appendix 7.1 we prove that the variance and covariance of $\mathbb{Y}_t$ exist under Assumption 2.1.

2.2 Asymptotic stationary distribution

We define a as any vector $a \in \mathbb{R}^N$ with $|a|_2 = 1$ and fixed d number of non zero elements. We then show that the averaged response is asymptotically normal distributed. Let $\tilde{\mathbb{Y}}_t = \mathbb{Y}_t - \mu_{\mathbb{Y}}$, $L_T = \sum_{t=1}^{T} a^\top\tilde{\mathbb{Y}}_t$, and $L_t = L_{\lfloor t\rfloor} + (t - \lfloor t\rfloor)a^\top\mathbb{Y}_{\lfloor t\rfloor+1}, t \geqslant 1$, where $\lfloor t\rfloor = \max\{k \in \mathbb{Z} : k \geqslant t\}$ is the floor function. Then, we have the following thoerem.

Theorem 1. *Consider any vector $a \in \mathbb{R}^N$ with $|a|_2 = 1$ and fixed $d < N$ number of non zero elements. Under Assumption 2.1, then*

$$\frac{L_{Tu}}{\sqrt{T}} \Rightarrow \sigma_{a\mathbb{Y}}\mathbf{B}(u), \quad 0 \leqslant u \leqslant 1 \tag{6}$$

where $\sigma_{a\mathbb{Y}}^2 \stackrel{\text{def}}{=} \sum_{l\geqslant 0} a^\top\Gamma_l a$ is the long run variance of $a^\top\tilde{\mathbb{Y}}_t$ and $\mathbf{B}(u)$ $(0 \leqslant u \leqslant 1)$ is a Brownian motion.

Remark For $u = 1$, Theorem 1 also implies:

$$\sqrt{T}(a^\top(\overline{\mathbb{Y}} - \mu_{\mathbb{Y}})) \xrightarrow{\mathcal{L}} \mathsf{N}(0, \sigma_{a\mathbb{Y}}^2), \quad \text{as } T \to \infty. \tag{7}$$

where $\overline{\mathbb{Y}} = T^{-1}\sum_{t=1}^{T}\mathbb{Y}_t$. Thus, the mean of $\mathbb{Y}_t$ converges in law to a normal distribution.

3 The IVQR estimation

In this section, we briefly discuss our estimator and the associated assumptions. We also show the estimation steps adopted for simulations and applications.

3.1 IVQR estimator

Suppose that there exists an $N \times \ell$ instrumental variable matrix, which is denoted as $\mathbf{R}_t = (R_{1t}, \cdots, R_{Nt})^\top \in \mathbb{R}^{N\times\ell}$. We have the following quantile conditions:

$$\mathrm{P}\left(Y_{it} \leqslant \gamma_1(\tau)\overline{Y}_{it} + X_{it}^\top \phi(\tau) | X_{it}, R_{it}\right) = \tau \ a.s. \tag{8}$$

We denote $\overline{Y}_{it} = n_i^{-1}\sum_{j=1}^N a_{ij}Y_{jt}$ and $X_{it} = \left(1, Z_i^\top, \overline{Y}_{i,t-1}, Y_{i,t-1}, F_t^\top, ..., F_{t-k}^\top\right)^\top$ with $\phi(\tau) = [\gamma_0(\tau), \alpha^\top(\tau), \gamma_2(\tau), \gamma_3(\tau), \beta_0^\top(\tau), ..., \beta_k^\top(\tau)]^\top \in \mathbb{R}^{q+3+(p+1)m}$.

In order to solve (8) we need to find $(\gamma_1(\tau), \phi(\tau))$ such that 0 is a solution to the standard quantile regression (QR) of $Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - X_{it}^\top\phi(\tau)$ on (X_{it}, R_{it}):

$$0 \in \arg\min_{g\in\mathscr{G}} \mathsf{E}\left[\rho_\tau\left\{Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - X_{it}^\top\phi(\tau) - g\left(X_{it}, R_{it}\right)\right\}\right], \tag{9}$$

where $\rho_\tau(u) = u\{\tau - \mathbf{I}(u < 0)\}$ is the check function with $\mathbf{I}(\cdot)$ as the indicator function, and $\mathscr{G}$ is the class of measurable functions of (X_{it}, R_{it}). This is referred to as the instrumental variable quantile regression estimator. Following Chernozhukov and Hansen (2005) and Su and Yang (2011), we restrict $\mathscr{G}$ to the class of linear functions, say

$$\mathscr{G} = \{g\left(X_{it}, R_{it}\right) = R_{it}^\top\lambda(\tau) : \lambda \in \Lambda\}, \tag{10}$$

where Λ is a compact set in $\mathbb{R}^\ell$. Here, we consider some transformed instrumental variables Φ_{it}, which are obtained by the least squares projection of $\overline{Y}_{it}$ on (X_{it}, R_{it}). Then, we obtain the finite sample analogue of the quantile regression objective function as

$$Q(\gamma_1(\tau), \phi(\tau), \lambda(\tau)) = \sum_{i=1}^N\sum_{t=1}^T\left[\rho_\tau\left\{Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - X_{it}^\top\phi(\tau) - \Phi_{it}^\top\lambda(\tau)\right\}\right]. \tag{11}$$

To simplify the presentation, we define $\theta(\tau) = (\gamma_1(\tau), \phi(\tau))$ and $\eta(\tau) = (\phi(\tau), \lambda(\tau))$. The main idea behind the IVQR estimation lies in that the estimator $(\widehat{\gamma}_1(\tau), \widehat{\phi}(\tau), \widehat{\lambda}(\tau))$ based on (11) can approximate the target true parameter set $(\gamma_1(\tau), \phi(\tau), 0)$. For a given value of endogenous parameter $\gamma_1(\tau)$ over a fine grid of a compact subset of the interval $(-1, 1)$, we first run the ordinary QR of $Y_{it} - \gamma_1(\tau)\overline{Y}_{it}$ on (X_{it}, Φ_{it}) and obtain the corresponding estimator. The estimator is denoted as $\widehat{\eta}(\gamma_1(\tau), \tau) = \left[\widehat{\phi}(\gamma_1(\tau), \tau), \widehat{\lambda}(\gamma_1(\tau), \tau)\right]$. Next, we select $\gamma_1(\tau)$ which minimizes $\widehat{\lambda}(\gamma_1(\tau), \tau)$ over the interval $(-1, 1)$. The IVQR estimator of $\theta(\tau) = (\gamma_1(\tau), \phi(\tau))$ is then obtained by $(\widehat{\gamma}_1(\tau), \widehat{\phi}(\widehat{\gamma}_1(\tau), \tau))$.

For a given quantile index τ, the IVQR estimation can proceed as follows:

(i) For a given value of $\gamma_1(\tau)$, run the quantile regression of $Y_{it} - \gamma_1(\tau)\overline{Y}_{it}$ against (X_{it}, Φ_{it}) to obtain

$$\widehat{\eta}(\gamma_1(\tau), \tau) = \arg\min_{(\phi,\lambda)} Q(\gamma_1(\tau), \phi(\tau), \lambda(\tau)). \tag{12}$$

(ii) Minimize a weighted norm of $\widehat{\lambda}(\gamma_1(\tau), \tau)$ over $\gamma_1(\tau)$ to obtain the IVQR estimator of $\gamma_1(\tau)$:

$$\widehat{\gamma}_1(\tau) = \arg\min_{\gamma_1 \in (-1,1)} \widehat{\lambda}^\top(\gamma_1(\tau), \tau)\; A\; \widehat{\lambda}(\gamma_1(\tau), \tau), \tag{13}$$

where A is some positive definite matrix. Without loss of generality we shall set $A = I$ throughout the paper.

(iii) Run the quantile regression of $Y_{it} - \widehat{\gamma}_1(\tau)\overline{Y}_{it}$ on X_{it} to obtain the estimator of $\phi(\tau)$. The estimator is denoted as $\widehat{\phi}(\tau) = \widehat{\phi}(\widehat{\gamma}_1(\tau), \tau)$. Then, we finally obtain the IVQR estimator by

$$\widehat{\theta}(\tau) = (\widehat{\gamma}_1(\tau), \widehat{\phi}(\tau)) = (\widehat{\gamma}_1(\tau), \widehat{\phi}(\widehat{\gamma}_1(\tau), \tau)). \tag{14}$$

3.2 Asymptotic theory

In order to develop the relevant asymptotic theory for the IVQR estimator, we need to carefully deal with some topological properties of the dependent variable $\mathbb{Y}_t$ that are spatially and temporally dependent. We utilize NED to address the spatial dependence of the statistics involved. The derivation of the asymptotic property follows from the standard M-estimation with quantile loss function as a special case. The complication is the detailed assumptions imposed on the adjacency matrix and we study how this translates to the dependence of the statistics involved. First, conditioning on the common factor process, we show that the elements of $\{\mathbb{Y}_t\}_t$ follows a near-epoch dependent (NED) process after introducing the basic definitions as in 3.2.1. Then, we derive the asymptotic distribution of the IVQR estimator under certain regularity assumptions in subsection 3.2.2.

3.2.1 Definition and notations

Jenish and Prucha (2009, 2012) extend the notion of near-epoch dependent (NED) processes used in the time series literature to random fields. This class of NED processes can accommodate a wide range of models with spatial dependence. They derive the central limit theorem and the law of large numbers for NED random fields. Accordingly, we consider $\mathbb{R}^d, d \geqslant 1$. The space $\mathbb{R}^d$ is endowed with the metric $\rho(i,j) = \max_{1 \leqslant l \leqslant d} |j_l - i_l|$ with the corresponding norm $|i| = \max_{1 \leqslant l \leqslant d} |i_l|$, where i_l is the l-th element of i. The distance between any subsets $U, V \subseteq D$ is defined as $\rho(U,V) = \inf\{\rho(i,j) : i \in U \text{ and } j \in V\}$. Let $|U|$ denote the cardinality of a finite subset U. We set in our two dimensional data set t as a special dimensional in space, therefore $\rho((i,t),(i',t')) = \max(|i-i'|, |t-t'|)$.

Assumption 3.1. *Let the lattice $D \subseteq \mathbb{R}^d, d \geqslant 1$, be countably infinite. Then, $\rho(i,j) \geqslant \rho_0, \forall i,j \in D$. We set $\rho_0 > 1$ w.l.o.g.*

Let $Z = \{X_{it}, U_{it}, (i,t) \in D_{NT}, NT \geqslant 1\}$ be triangular arrays of random fields defined on a probability space $(\Omega, \mathscr{F}, P)$ with $D_{NT} \subseteq D$. The cardinality of D_{NT} satisfying $\lim_{N\to\infty, T\to\infty} |D_{NT}| \to \infty$. Let $\mathcal{C} \overset{\text{def}}{=} \{\mathbb{F}_t\}_t$, and define $\mathscr{F}_{it}(s) = \sigma(X_{i't'}, U_{i't'}, \mathcal{C} : (i',t') \in D_{NT}, \rho((i',t'),(i,t)) \leqslant s)$ as the σ- field generated by random vectors $X_{i't'}, U_{i't'}$ located within distance s from (i,t).

Definition 3.1. *Let $Z = \{Z_{it}, (i,t) \in D_{NT}, NT \geqslant 1\}$ and $\varepsilon = \{\varepsilon_{it}, (i,t) \in D_{NT}, NT \geqslant 1\}$ be random fields with $\|Z_{it}\|_p < \infty$ for $p \geqslant 1$, where $D_{NT} \subseteq D$ and its cardinality is given by $|D_{NT}| = NT$. Let $\{d_{it}, (i,t) \in D_{NT}, NT \geqslant 1\}$ be an array of positive constants. Then,* *the random field Z is L_p-near-epoch dependent on the random field ε if*

$$\|Z_{it} - \mathsf{E}(Z_{it}|\mathscr{F}_{it}(s))\|_p < d_{it}\varphi(s)$$

for some sequence $\varphi(s) \geqslant 0$ with $\lim_{s\to\infty}\varphi(s) = 0$. $\{\varphi(s)\}$ are the NED coefficients and d_{it}s are the NED scaling factors. If $\sup_{NT}\sup_{(i,t)\in D_{NT}} d_{it} < \infty$, then Z is uniformly L_p-NED on ε.

Next, we present the L_2-NED properties of random field Z on some α-mixing random field ε. The definition of the α-mixing coefficient is stated as follows:

Definition 3.2. *Let $\mathscr{A}$ and $\mathscr{B}$ be two σ-algebras of $\mathscr{F}$, and define*

$$\alpha(\mathscr{A},\mathscr{B}) = \sup(|P(AB) - P(A)P(B)|, A \in \mathscr{A}, B \in \mathscr{B}),$$

For $U \subseteq D_{NT}$ and $V \subseteq D_{NT}$, let $\sigma_{NT}(U) = \sigma(\varepsilon_{it}, (i,t) \in U)$ and $\alpha_{NT}(U,V) = \alpha(\sigma_{NT}(U), \sigma_{NT}(V))$. Then, the α-mixing coefficients for the random field ε is defined as:

$$\overline{\alpha}(u,v,h) = \sup_{N,T}\sup_{U,V}(\alpha_{NT}(U,V), |U| \leqslant u, |V| \leqslant v, \rho(U,V) \geqslant h)$$

with $u, v, h \in \mathbb{N}$.

Unlike the standard mixing time-series processes, the mixing coefficients for random fields depend not only on the distance between two sets but also on their sizes. We further assume that $\overline{\alpha}(u,v,h) \leqslant \varphi(u,v)\widehat{\alpha}(h)$, where the function $\varphi(u,v)$ is nondecreasing with u and v, and $\widehat{\alpha}(h) \to 0$ as $h \to \infty$ (see Assumption 3.4(ii) below). This suggests that we explicate the two different sources of dependence, separately: (i) the decay of dependence with the distance, and (ii) the accumulation of dependence as the sample region expands. In the random field literature, $\varphi(u,v)$ can be commonly selected such that $\varphi(u,v) = (u+v)^a, a \geqslant 0$ or $\varphi(u,v) = \min(u,v)$, see Jenish and Prucha (2012).

Following Xu and Lee (2015), we outline some conditions for NED properties for the dependent variable $\mathbb{Y}_t$ in Assumptions 3.2 and 3.3.

Assumption 3.2. *The network matrix W is non-stochastic one with zero diagonals and uniformly bounded for all N with absolute row and column sums such that the matrix $S_t = \mathbf{I} - \mathbf{A}_{1t}W$ is nonsingular. We consider two cases for $w_{ij} \geqslant 0$ for any i,j.*

(1) Case 1: $|w_{ij}| \leqslant \pi_0 \rho(i,j)^{-c_w}$ with constants $\pi_0 \geqslant 0$ and $c_w > d$. In addition, there exists at most the $K(\geqslant 1)$ number of columns in W, with $\min_u |\gamma_1(u)| \sum_{i=1}^n |w_{ij}| > \Upsilon$.

(2) Case 2: Two nodes influence each other only if they are located sufficiently close; namely, $w_{ij} \neq 0$ if $\rho(i,j) \leqslant \bar{\rho}_0$ and $w_{ij} = 0$ otherwise. We set $\bar{\rho}_0 > 1$ w.l.o.g.

Assumption 3.2 (1) allows two individuals to have direct interaction even though their locations are far away from each other, with the requirement that the strength of interaction w_{ij} declines with the distance $\rho(i,j)$ in the power of c_w. The assumption is in line with Xu and Lee (2015). Moreover, by excluding a limited number of nodes $K(\geqslant 1)$, the total effects on other units from each node should be bounded, i.e., we assume that $\sup|\gamma_1(u)| \sup_j \sum_{i=1}^n |w_{ij}| < \Upsilon$ or $\sup|\gamma_1(u)| \sup_j \sum_{i=1}^N |w_{ij}| < 1$ w.l.o.g. This corresponds to the existence of a narrow number of units with large aggregate effects on others even as the total number of nodes rises. Assumption 3.2 (2) allows two individuals to have direct interaction only if they are located within a specific distance. Assumption 3.2 is mainly used to restrict the NED coefficients $\varphi(s) \to 0$ as $s \to \infty$. We will discuss the NED properties of $\{Y_{it}\}_{i,t}$ in Proposition 1 under these two scenarios.

Assumption 3.3. *(1) $\{X_{it}, U_{it}\}_{i,t}$ is an α-mixing random field with an α-mixing coefficient, $\overline{\alpha}(u,v,h) \leqslant (u+v)^{\varsigma} \widehat{\alpha}(h)$, $\varsigma \geqslant 0$, where $\sum_{h=1}^{\infty} h^{2(\varsigma_0+1)-1} \widehat{\alpha}^{\delta/(4+2\delta)}(h) < \infty$ with $\varsigma_0 = \delta\varsigma/(4+2\delta)$ and some constants $\delta, \xi > 0$.*
(2) $\sup_{i,t} \|(X_{it}, U_{it})\|_{2+\delta} < \infty$ for some $\delta > 0$.

The α-mixing coefficient of $\{X_{it}, U_{it}\}_{i,t}$ in Assumption 3.3(1) is related to the properties of NED process $\{Y_{it}\}_{i,t}$. Assumption 3.3(2) is required to constrain the bound property of the NED scaling factors d_{it} in $\{Y_{it}\}_{i,t}$.

Define $u_{it} = u_{it}(\gamma_1, \phi, \lambda, \tau) = Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - X_{it}^\top \phi(\tau) - \Phi_{it}^\top \lambda(\tau)$ with its transformations, the check function $\rho_\tau(u) = (\tau - \mathbf{1}(u \leqslant 0))u$ and $\psi_\tau(u) = \tau - \mathbf{1}(u \leqslant 0)$ which is the (directional) derivative of $\rho_\tau(u)$. Proposition 3.1 provides the NED properties of $\{Y_{it}\}_{i,t}$, $\{\rho_\tau(u_{it})\}_{i,t}$ and $\{\psi_\tau(u_{it})\}_{i,t}$ on the base $\{X_{it}, U_{it}\}_{i,t}$.

Proposition 3.1. *(1) Under Assumptions 3.1-3.2(1) and 3.3(ii), $\{Y_{it}\}_{i,t}$ is geometrically L_2-NED on $\{X_{it}, U_{it}\}_{i,t}$ such that $\|Y_{it} - \mathsf{E}(Y_{it}|\mathscr{F}_{it}(s))\|_2 < Cs^{-(c_w-d)}$ ($c_w > d$) for some*

$C > 0$ that does not depend on i and t. Similarly, $\{u_{it}\}_{i,t}$ hold the same conclusions. The transformations $\{\psi_\tau(u_{it})\}_{i,t}$ and $\{\rho_\tau(u_{it})\}_{i,t}$ are also L_2-NED on $\{X_{it}, U_{it}\}_{i,t}$.

(2) Under Assumptions 3.1-3.2(2) and 3.3(ii), $\{Y_{it}\}_{i=1}^n$ is geometrically L_2-NED on $\{X_{it}, U_{it}\}_{it}$ such that $\|Y_{it} - \mathsf{E}(Y_{it}|\mathscr{F}_{it}(s))\|_2 < C\Upsilon^{s/\bar{\rho}_0}$ $(\Upsilon < 1)$ for some $C > 0$ that does not depend on i and t. Similarly, $\{u_{it}\}_{i,t}$ hold the same conclusions. The transformations $\{\psi_\tau(u_{it})\}_{i,t}$ and $\{\rho_\tau(u_{it})\}_{i,t}$ are also L_2-NED on $\{X_{it}, U_{it}\}_{i,t}$.

Define $s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau) = \psi_\tau \left\{Y_{it} - \gamma_1^0(\tau)\overline{Y}_{it} - \Psi_{it}^\top \eta^0(\gamma_1^0, \tau)\right\} \Psi_{it}$ with $\Psi_{it} = (X_{it}^\top, \Phi_{it}^\top)^\top$ and $\check{s}_{it} = \check{s}_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau) = s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau)$. Then, conditioning on $\mathcal{C}$, it is straightforward to show that the process $\{\check{s}_{it}\}_{i,t}$ is NED. To derive the central limit theorem of $G_{NT}^0 = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau) - \mathsf{E}\, s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau)] = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \check{s}_{it}$, where the variance of G_{NT}^0 is given by $\Omega_0 = \tau(1-\tau) \lim_{N\to\infty, T\to\infty} \mathsf{E}(\Psi_{it}\Psi_{it}^\top)$, we make the following assumptions:

Assumption 3.4. *(i) There exist nonrandom positive constants $\{c_{it}, (i,t) \in D_{NT}, NT \geqslant 1\}$ such that $\rho_\tau(u_{it})/c_{it}$ is uniformly L_p-bounded for $p > 1$, i.e., $\sup_{N,T} \sup_{(i,t)\in D_{NT}} \mathsf{E}\, |\rho_\tau(u_{it})/c_{it}|^p < \infty$.*

(ii) The α-mixing coefficient of the random field $\{X_{it}, U_{it}\}_{i,t}$ satisfies: $\overline{\alpha}(u, v, h) \leqslant \varphi(u, v)\widehat{\alpha}(h)$ where the function $\varphi(u, v)$ is non-decreasing with u and v, and $\sum_{h=1}^\infty h^{d-1}\widehat{\alpha}(h) < \infty$.

Assumption 3.5. *(Uniform $L_{2+\delta}$ Integrability)*

(i) There exists an array of positive constants $\{c_{it}, (i,t) \in D_{NT}, NT \geqslant 1\}$ such that $\lim_{k\to\infty} \sup_{N,T} \sup_{(i,t)\in D_{NT}} \mathsf{E}\{|\check{s}_{it}/c_{it}|^{2+\delta} 1(|\check{s}_{it}/c_{it}| > k)\} = 0$ for $\delta > 0$.

(ii) $\inf_{N,T} |D_{NT}|^{-1} M_{NT}^{-2} \Omega_0 > 0$, where $M_{NT} = \max_{(i,t)\in D_{NT}} c_{it}$.

(iii) NED coefficients satisfy: $\sum_{h=1}^\infty h^{d-1}\varphi(h) < \infty$, and NED scaling factors satisfy: $\sup_{NT} \sup_{(i,t)\in D_{NT}} c_{it}^{-1} d_{it} \leqslant C < \infty$.

3.2.2 Asymptotic distribution of the IVQR estimator

In this subsection, we show the asymptotic normality of our estimator.

Assumption 3.6 (Conditions for identification and estimation)**.** ***R1****. (Compactness and Convexity) For all τ, $(\gamma_1(\tau), \phi(\tau)) \in \mathcal{A} \times \mathcal{B}$, $\mathcal{A} \times \mathcal{B}$ is compact and convex.*

R2*. (Full Rank and Continuity) $\mathbb{Y}_t$ has bounded conditional density, a.s. $\sup_{\mathbb{Y}_t \in \mathbb{R}^N} f_{\mathbb{Y}_t|\mathcal{F}_t}(y) < K$, where $\mathcal{F}_t = \{Z_i, \mathbb{Y}_{t-1}, \mathbb{Y}_t, F_t\}$ is the information set. Define*

$$\begin{aligned}
S_{NT}(\pi, \tau) &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\psi_\tau \left\{ Y_{it} - \gamma_1(\tau) \overline{Y}_{it} - X_{it}^\top \phi(\tau) - \Phi_{it}^\top \lambda(\tau) \right\} \Psi_{it} \right], && (15) \\
S_\infty(\pi, \tau) &= \lim_{N \to \infty, T \to \infty} \mathsf{E}\left[S_{NT}(\pi, \tau) | \mathcal{C} \right], \quad S_\infty^*(\pi, \tau) = \lim_{N \to \infty, T \to \infty} \mathsf{E}\left[S_{NT}(\pi, \tau) \right], && (16) \\
S_{NT}(\theta, \tau) &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\psi_\tau \left\{ Y_{it} - \gamma_1(\tau) \overline{Y}_{it} - X_{it}^\top \phi(\tau) \right\} \Psi_{it} \right], && (17) \\
S_\infty(\theta, \tau) &= \lim_{N \to \infty, T \to \infty} \mathsf{E}\left[S_{NT}(\theta, \tau) | \mathcal{C} \right], \quad S_\infty^*(\theta, \tau) = \lim_{N \to \infty, T \to \infty} \mathsf{E}\left[S_{NT}(\theta, \tau) \right]. && (18)
\end{aligned}$$

where $\pi \equiv (\gamma_1, \phi^\top, \lambda^\top)^\top$, $\theta \equiv (\gamma_1, \phi^\top)^\top$, $\psi_\tau(u) = \tau - \mathbf{I}(u < 0)$, and $\Psi_{it} = (X_{it}^\top, \Phi_{it}^\top)^\top$. Then, Jacobian matrices $\dfrac{\partial S_\infty(\theta, \tau)}{\partial(\gamma_1, \phi)}$ and $\dfrac{\partial S_\infty(\pi, \tau)}{\partial(\phi, \lambda)}$ are continuous and have full rank, uniformly over $\mathcal{A} \times \mathcal{B} \times \mathcal{G} \times \mathcal{T}$, where $\mathcal{G}$ is a compact set with $\lambda(\tau) \in \mathcal{G}$, and the image of $\mathcal{A} \times \mathcal{B}$ under the mapping $(\gamma_1, \phi) \mapsto S_\infty(\theta, \tau)$ is simply connected.

R3*. For a given fixed $\tau \in \mathcal{T}$, the unknown true parameter $\theta^0(\tau) = (\gamma_1^0(\tau), \phi^0(\tau))$ uniquely solves $S_\infty(\theta, \tau) = 0$ over $\mathcal{A} \times \mathcal{B}$.*

Remark: Condition R1 assumes compactness of the parameter space, which is needed for $\gamma_1(\tau)$ due to the non-convex objective function with respect to γ_1. The condition R2 implies the global identification, and the continuity condition is required for deriving the asymptotic normality. R3 requires that $(\gamma_1^0(\tau), \phi^0(\tau))$ be the unique solution to $S_\infty(\theta, \tau) = 0$, which is necessary for the consistency of the estimator.

Denote $\widehat{\theta}(\tau) = (\widehat{\gamma}_1(\tau), \widehat{\phi}(\tau))$ as the IVQR estimator of $\theta(\tau) = (\gamma_1(\tau), \phi(\tau))$, where $\widehat{\phi}(\tau) = \widehat{\phi}(\widehat{\gamma}_1(\tau), \tau)$. Denote the unknown true parameter as $\theta^0(\tau) = (\gamma_1^0(\tau), \phi^0(\tau))$. Define the $(q + 4 + (p + 1)m) \times (q + 4 + (p + 1)m)$ matrices:

$$J(\tau) = \left. \frac{\partial S_\infty(\pi, \tau)}{\partial(\gamma_1, \phi)} \right|_{\gamma_1 = \gamma_1^0, \phi = \phi^0, \lambda = 0}, \quad J^*(\tau) = \left. \frac{\partial S_\infty^*(\pi, \tau)}{\partial(\gamma_1, \phi)} \right|_{\gamma_1 = \gamma_1^0, \phi = \phi^0, \lambda = 0}. \qquad (19)$$

Theorem 2 (Linearization)**.** *Under Assumption 3.1 - 3.4, 3.6, as min$\{N, T\} \to \infty$,*

$$\sqrt{NT} \left\{ \widehat{\theta}(\tau) - \theta^0(\tau) \right\} = -\, J^{-1}(\tau) G_{NT}^0(\theta^0, \tau) + o_p(1). \qquad (20)$$

Further, under Assumption 3.5, the NED process $\{\check{s}_{it}\}_{i,t}$ satisfies the central limit theorem such that $G_{NT}^0(\theta^0, \tau) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \check{s}_{it}$ follows a mean zero Gaussian process with covariance function $\Omega_0 = \tau(1-\tau) \,\mathsf{E}(\Psi_{it}\Psi_{it}^\top | \mathcal{C})$. Define $\Omega_0^* = \tau(1-\tau)\,\mathsf{E}(\Psi_{it}\Psi_{it}^\top)$, one can prove that $\Omega_0^{-1}\Omega_0^* \to_p I$. Then, we establish the asymptotic normality of $\widehat{\theta}(\tau)$ in the theorem below.

Theorem 3 (Estimation)**.** *Under Assumption 3.1 - 3.6, we have* $\Omega_0^{-1}\Omega_0^* \to_p I$, *and* $J^{-1}(\tau)J^*(\tau) \to_p I$, *then*

$$\sqrt{NT}\left\{\widehat{\theta}(\tau) - \theta(\tau)\right\} \xrightarrow{d} \mathsf{N}(0, \Sigma_\theta), \tag{21}$$

where $\Sigma_\theta = J^*(\tau)^{-1}\Omega_0^* J^*(\tau)^{-1}$, *and* $\Omega_0^* = \tau(1-\tau)\,\mathsf{E}(\Psi_{it}\Psi_{it}^\top)$.

Remark: We estimate the variance covariance matrix $\widehat{\Omega}_0^*$ as $(NT)^{-1}\tau(1-\tau)\sum_{i=1}^N\sum_{t=1}^T \Psi_{it}\Psi_{it}^\top$. Moreover, to estimate $J^*(\tau)$, define $\tilde{\xi}_{it} = (\overline{Y}_{it}, \Phi_{it}^\top)^\top$, then

$$\widehat{J}^*(\tau) = (2NTh)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathbf{I}(\widehat{u}_{it} \leqslant h)\Psi_{it}\tilde{\xi}_{it}^\top. \tag{22}$$

4 Monte Carlo simulations

We examine the finite sample properties of the IVQR estimator via a Monte Carlo simulation using a data generating process based on the model in equation (1) using the three different network structures.

4.1 The setup

We construct the DGP from the model (1) as follows: First, we generate the five nodal covariates, $Z_i = (Z_{i1}, \cdots, Z_{i5})^\top \in \mathbb{R}^5$ from a multivariate normal distribution $N(\mathbf{0}, \Sigma_z)$, where $\Sigma_z = (\sigma_{j_1j_2})$ and $\sigma_{j_1j_2} = 0.5^{|j_1-j_2|}$. Then, we construct the two common covariates, $F_t = (F_{1t}, F_{2t})^\top \in \mathbb{R}^2$ from the iid standard normal distribution. Let $\gamma_{j,it} = \gamma_j(U_{it})$ for $j = 0, 1, 2, 3$, $\alpha_{j,it} = \alpha_j(U_{it})$ for $j = 1, ..., 5$, and $\beta_{mj,it} = \beta_{mj}(U_{it})$ for $m = 1, 2$ and $j = 0, 1$, where we set the lag of 2 common covariates to 1 ($p = 1$). We then generate the random coefficients as follows:

$$
\begin{aligned}
\gamma_{0,it} =& u_{it},\ \gamma_{1,it} = 0.1\Phi(u_{it}),\ \gamma_{2,it} = 0.4\{1+\exp(u_{it})\}^{-1}\exp(u_{it}),\ \gamma_{3,it} = 0.4\Phi(u_{it}),\\
\alpha_{1,it} =& 0.5\Phi(u_{it}),\ \alpha_{2,it} = 0.3\mathbf{G}(u_{it},1,2),\ \alpha_{3,it} = 0.2\mathbf{G}(u_{it},2,2),\\
\alpha_{4,it} =& 0.25\mathbf{G}(u_{it},3,2), \alpha_{5,it} = 0.2\mathbf{G}(u_{it},2,1),\\
& \beta_{10,it} = 0.1\Phi(u_{it}),\ \beta_{11,it} = 0.3\mathbf{G}(u_{it},2,2), \beta_{20,it} = 0.2\mathbf{G}(u_{it},1,2),\ \beta_{21,it} = 0.3\mathbf{G}(u_{it},2,1),
\end{aligned}
$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\mathbf{G}(\cdot,a,b)$ is the Gamma distribution function with shape parameter a and scale parameter b, and u_{it}s are iid random variables, generated either from (a) the standard normal distribution or from (b) the t-distribution with 5 degrees of freedom. Finally, $\mathbb{Y}_t$s are generated by (1).

To check the robustness of the finite sample performance of the IVQR estimator, we consider the three different adjacency matrices that have been popular in the literature. The adjacency matrix setup follows for example from Zhu, Wang, Wang and Härdle (2019).

Type 1. (Stochastic Block Model) We first consider the Stochastic Block Model (Wang and Wong, 1987; Nowicki and Snijders, 2001) with an important application in community detection (Zhao et al., 2012). We follow Nowicki and Snijders (2001) and randomly assign each node a block label indexed from 1 to K, where $K \in \{5, 10, 20\}$. We then set $P(a_{ij} = 1) = 0.3N^{-0.3}$ if i and j are in the same block, and $P(a_{ij} = 1) = 0.3N^{-1}$ otherwise. This indicates that the nodes within the same block have higher probability of connecting with each other than the nodes between blocks.

Type 2. (Dyad Independence Model) Holland and Leinhardt (1981) introduce a Dyad Independence Model with a Dyad $D_{ij} = (a_{ij}, a_{ji})$ for $1 \leqslant i < j \leqslant N$, where D_{ij}s are assumed to be independent. We set the probability of dyads being mutually connected to $P(D_{ij} = (1,1)) = 20N^{-1}$ to ensure the network sparsity. Then, we set $P(D_{ij} = (1,0)) = P(D_{ij} = (0,1)) = 0.5N^{-0.8}$, which implies that the expected degree for each node is $\mathcal{O}(N^{0.2})$. Accordingly, we have $P(D_{ij} = (0,0)) = 1 - 20N^{-1} - N^{-0.8}$, which tends to 1 as $N \to \infty$.

Type 3. (Power-law Distribution Network) It is a common observation that the majority of nodes in the network have small links while a small number of nodes have a large number of links, see Barabási and Albert (1999). In this case the degrees of nodes can be characterized by the power-law distribution. We simulate the adjacency matrix as follows: For each node, we generate the in-degree such that $d_i = \sum_j a_{ji}$ according to the discrete power-law distribution such as $P(d_i = k) = ck^{-\beta}$, where c is a normalizing constant and the exponent parameter β is set at 2.5 as in Clauset et al. (2009). Finally, for the ith node, we randomly select d_i nodes as its followers.

In order to perform the IVQR estimation, we should obtain the valid IVs internally, which is denoted $\mathbb{D}_t = (D_{1t}, \cdots, D_{Nt})^\top \in \mathbb{R}^{N\times\ell}$. Here we suggest constructing the IVs by the higher network orders of lagged dependent variables, say[1]

$$IV = [W^2\mathbb{Y}_{t-1}, W^3\mathbb{Y}_{t-1}] \tag{23}$$

where W is the row-sum normalized network matrix selected.

4.2 Simulation Results

We first consider the case where N is larger than $T\,(= N/10)$, and set the size of nodes to $N = 100, 500, 1000$. Using $R = 100$ replications, we evaluate the finite sample performance of the IVQR estimator at $\tau = 0.1, 0.5, 0.9$, respectively. In Tables 1, 2 and 3 we report the simulation results in terms of bias and RMSE. Overall, we find that the RMSEs of all the parameters are monotonically decreasing as the sample size increases for the three different network structures, across the different quantiles ($\tau = 0.1, 0.5, 0.9$) and disregarding the distributions of u_{it}s. In particular, the RMSEs of γ_1 are somewhat larger, especially in a small sample ($N = 100$), which may reflect uncertainty associated with the selection of the IV variables. But, when the sample size grows, i.e. $N = 1000$, all the RMSEs decline sharply, implying that the parameters converge at a proper rate as the sample size grows sufficiently.

Next, we consider the cases with $N = T$ and $T > N$. Setting $N = 100$, we examine

[1] In principle, we can also select the higher network orders such as $IV = [W^2\mathbb{Y}_{t-1}, W^3\mathbb{Y}_{t-1}, W^2\mathbb{Y}_{t-2}, W^3\mathbb{Y}_{t-2}, ...]$. But, we find that two instruments in (23) were often the best choice.

Table 1: RMSE with W as Type 1 network (Dyad independence model)

Dist.	τ	γ_0	γ_1	γ_2	γ_3	α_1	α_2	α_3	α_4	α_5	β_1	β_2	β_3	β_4
								$N = 100$						
Z	0.1	0.072	0.250	0.083	0.057	0.225	0.079	0.071	0.071	0.057	0.091	0.139	0.074	0.078
	0.5	0.061	0.287	0.055	0.042	0.100	0.077	0.047	0.045	0.050	0.053	0.077	0.051	0.053
	0.9	0.078	0.352	0.078	0.049	0.072	0.074	0.071	0.070	0.064	0.080	0.074	0.073	0.079
T	0.1	0.099	0.097	0.086	0.054	0.088	0.098	0.095	0.107	0.071	0.103	0.106	0.102	0.111
	0.5	0.056	0.214	0.046	0.030	0.049	0.056	0.045	0.049	0.042	0.054	0.062	0.061	0.057
	0.9	0.101	0.402	0.080	0.055	0.084	0.100	0.101	0.104	0.086	0.106	0.099	0.099	0.106
								$N = 500$						
Z	0.1	0.012	0.037	0.014	0.010	0.221	0.020	0.014	0.013	0.010	0.047	0.132	0.011	0.011
	0.5	0.009	0.130	0.022	0.008	0.094	0.062	0.012	0.009	0.017	0.024	0.066	0.009	0.011
	0.9	0.014	0.249	0.035	0.010	0.041	0.032	0.017	0.015	0.027	0.011	0.011	0.013	0.019
T	0.1	0.016	0.026	0.014	0.011	0.017	0.021	0.018	0.020	0.017	0.015	0.013	0.015	0.012
	0.5	0.009	0.140	0.017	0.005	0.009	0.011	0.009	0.008	0.008	0.007	0.011	0.008	0.008
	0.9	0.018	0.272	0.030	0.012	0.017	0.018	0.017	0.018	0.017	0.015	0.022	0.016	0.015
								$N = 1000$						
Z	0.1	0.003	0.088	0.005	0.003	0.097	0.007	0.003	0.003	0.003	0.020	0.061	0.002	0.002
	0.5	0.002	0.088	0.012	0.002	0.042	0.027	0.003	0.002	0.008	0.010	0.030	0.003	0.004
	0.9	0.005	0.088	0.019	0.003	0.018	0.013	0.007	0.003	0.012	0.003	0.003	0.005	0.007
T	0.1	0.003	0.009	0.002	0.004	0.004	0.005	0.005	0.004	0.005	0.004	0.003	0.002	0.003
	0.5	0.002	0.058	0.010	0.001	0.002	0.004	0.003	0.002	0.001	0.002	0.004	0.001	0.001
	0.9	0.006	0.117	0.015	0.005	0.006	0.005	0.005	0.003	0.004	0.005	0.010	0.004	0.004

Note: The size of the time period is $T = N/10$. u_{it} are generated from a standard normal distribution (i.e. Z) and t-distribution with 5 degrees of freedom (i.e., T). The simulation results are reported with 100 replications.

Table 2: RMSE with W as Type 2 network (stochastic block model)

Dist.	τ	γ_0	γ_1	γ_2	γ_3	α_1	α_2	α_3	α_4	α_5	β_1	β_2	β_3	β_4
								$N = 100$						
Z	0.1	0.077	0.579	0.231	0.049	0.221	0.072	0.078	0.075	0.070	0.090	0.110	0.094	0.100
	0.5	0.072	0.589	0.187	0.046	0.097	0.070	0.056	0.050	0.050	0.059	0.089	0.069	0.064
	0.9	0.092	1.014	0.277	0.057	0.068	0.078	0.064	0.068	0.067	0.089	0.131	0.104	0.097
T	0.1	0.147	0.881	0.307	0.050	0.078	0.089	0.087	0.089	0.078	0.154	0.162	0.130	0.133
	0.5	0.070	0.628	0.170	0.033	0.046	0.055	0.057	0.053	0.047	0.068	0.078	0.066	0.063
	0.9	0.173	1.118	0.298	0.064	0.085	0.095	0.091	0.101	0.069	0.156	0.174	0.137	0.128
								$N = 500$						
Z	0.1	0.012	0.331	0.066	0.011	0.219	0.021	0.015	0.013	0.012	0.039	0.118	0.012	0.011
	0.5	0.009	0.290	0.166	0.008	0.093	0.060	0.011	0.009	0.021	0.046	0.137	0.010	0.015
	0.9	0.017	0.625	0.192	0.012	0.038	0.032	0.017	0.014	0.029	0.041	0.118	0.010	0.012
T	0.1	0.019	0.457	0.131	0.012	0.016	0.018	0.019	0.020	0.016	0.027	0.063	0.018	0.019
	0.5	0.009	0.237	0.129	0.005	0.009	0.011	0.010	0.009	0.007	0.022	0.061	0.007	0.007
	0.9	0.018	0.341	0.186	0.014	0.019	0.018	0.016	0.018	0.016	0.035	0.100	0.014	0.017
								$N = 1000$						
Z	0.1	0.003	0.050	0.024	0.002	0.097	0.007	0.002	0.004	0.003	0.023	0.066	0.002	0.003
	0.5	0.003	0.063	0.015	0.001	0.041	0.027	0.003	0.002	0.008	0.028	0.082	0.005	0.008
	0.9	0.007	0.112	0.022	0.002	0.017	0.014	0.005	0.003	0.012	0.029	0.085	0.002	0.002
T	0.1	0.004	0.030	0.017	0.004	0.004	0.004	0.004	0.003	0.004	0.004	0.005	0.003	0.003
	0.5	0.003	0.062	0.108	0.001	0.002	0.004	0.002	0.002	0.002	0.018	0.052	0.004	0.006
	0.9	0.008	0.123	0.180	0.003	0.005	0.006	0.004	0.004	0.004	0.031	0.089	0.007	0.009

Note: The size of the time period is $T = N/10$. u_{it} are generated from a standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The simulation results are reported with 100 replications.

Table 3: RMSE with W as Type 3 network (power-law distribution network)

Dist.	τ	γ_0	γ_1	γ_2	γ_3	α_1	α_2	α_3	α_4	α_5	β_1	β_2	β_3	β_4
								$N = 100$						
Z	0.1	0.095	0.359	0.337	0.056	0.227	0.068	0.066	0.069	0.068	0.102	0.189	0.083	0.082
	0.5	0.072	0.475	0.255	0.039	0.104	0.072	0.053	0.052	0.048	0.073	0.157	0.064	0.059
	0.9	0.108	0.721	0.424	0.056	0.074	0.065	0.069	0.075	0.062	0.093	0.167	0.110	0.093
T	0.1	0.121	0.392	0.419	0.070	0.087	0.102	0.103	0.100	0.085	0.092	0.101	0.097	0.099
	0.5	0.063	0.394	0.247	0.030	0.052	0.056	0.049	0.046	0.047	0.069	0.101	0.057	0.058
	0.9	0.137	0.762	0.504	0.058	0.075	0.086	0.080	0.089	0.086	0.130	0.218	0.125	0.117
								$N = 500$						
Z	0.1	0.014	0.125	0.034	0.011	0.219	0.019	0.014	0.014	0.013	0.049	0.147	0.011	0.010
	0.5	0.013	0.159	0.140	0.009	0.094	0.048	0.011	0.010	0.017	0.041	0.122	0.010	0.013
	0.9	0.264	0.387	0.204	0.028	0.021	0.023	0.017	0.021	0.019	0.034	0.087	0.017	0.019
T	0.1	0.017	0.066	0.025	0.011	0.015	0.021	0.019	0.019	0.016	0.013	0.017	0.016	0.016
	0.5	0.010	0.187	0.116	0.006	0.010	0.024	0.011	0.010	0.013	0.020	0.059	0.008	0.008
	0.9	0.356	0.310	0.175	0.036	0.055	0.039	0.022	0.022	0.022	0.043	0.112	0.025	0.021
								$N = 1000$						
Z	0.1	0.004	0.050	0.016	0.001	0.097	0.007	0.003	0.004	0.004	0.022	0.066	0.002	0.003
	0.5	0.002	0.067	0.064	0.002	0.043	0.022	0.003	0.003	0.006	0.018	0.054	0.004	0.004
	0.9	0.099	0.165	0.083	0.010	0.006	0.006	0.005	0.004	0.008	0.012	0.034	0.003	0.004
T	0.1	0.004	0.018	0.010	0.004	0.004	0.004	0.004	0.005	0.003	0.003	0.004	0.003	0.003
	0.5	0.003	0.073	0.049	0.001	0.003	0.010	0.001	0.002	0.003	0.008	0.025	0.001	0.002
	0.9	0.122	0.118	0.070	0.013	0.020	0.010	0.005	0.006	0.007	0.013	0.043	0.005	0.005

Note: The size of the time period is $T = N/10$. u_{it}s are generated from a standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The simulation results are reported with 100 replications.

the two scenarios with $T = N = 100$ and $T = 10, N = 1000$. Overall, we find that the simulation results reported in Table 4 are qualitatively similar to the previous ones with $N > T$. Again, the RMSEs for γ_1 seem to be somewhat higher than those of other parameters, especially for small T. But the RMSEs of all the parameters decline sufficiently fast as the sizes of the time period increase for the three different network structures and across different quantiles ($\tau = 0.1, 0.5, 0.9$)

Table 4: RMSE with T = N and T > N

Dist.	τ	γ_0	γ_1	γ_2	γ_3	α_1	α_2	α_3	α_4	α_5	β_1	β_2	β_3	β_4
							$W1$, $T = 100$							
Z	0.1	0.019	0.323	0.021	0.017	0.229	0.027	0.021	0.029	0.012	0.053	0.139	0.019	0.022
	0.5	0.020	0.325	0.026	0.008	0.107	0.057	0.014	0.011	0.023	0.022	0.061	0.017	0.014
	0.9	0.013	0.364	0.031	0.015	0.046	0.038	0.026	0.032	0.029	0.017	0.024	0.023	0.030
T	0.1	0.023	0.042	0.017	0.017	0.014	0.035	0.034	0.024	0.018	0.032	0.027	0.025	0.018
	0.5	0.010	0.275	0.020	0.008	0.015	0.017	0.015	0.011	0.009	0.014	0.012	0.009	0.007
	0.9	0.031	0.447	0.018	0.011	0.017	0.038	0.038	0.036	0.029	0.027	0.020	0.008	0.015
							$W1$, $T = 1000$							
Z	0.1	0.004	0.209	0.006	0.006	0.216	0.017	0.008	0.007	0.005	0.044	0.134	0.004	0.004
	0.5	0.003	0.203	0.027	0.005	0.093	0.061	0.008	0.005	0.018	0.023	0.068	0.006	0.009
	0.9	0.011	0.211	0.039	0.004	0.041	0.031	0.013	0.008	0.029	0.005	0.007	0.010	0.016
T	0.1	0.009	0.047	0.009	0.008	0.014	0.011	0.010	0.012	0.008	0.010	0.009	0.005	0.008
	0.5	0.006	0.155	0.012	0.003	0.004	0.005	0.003	0.005	0.003	0.004	0.007	0.005	0.003
	0.9	0.012	0.302	0.025	0.007	0.011	0.010	0.009	0.007	0.012	0.008	0.011	0.007	0.007
							$W2$, $T = 100$							
Z	0.1	0.020	0.524	0.114	0.021	0.223	0.028	0.023	0.025	0.014	0.036	0.093	0.020	0.023
	0.5	0.014	0.202	0.064	0.007	0.094	0.065	0.019	0.013	0.019	0.027	0.076	0.014	0.014
	0.9	0.027	0.262	0.088	0.010	0.043	0.031	0.023	0.023	0.029	0.029	0.038	0.016	0.017
T	0.1	0.050	0.529	0.177	0.022	0.028	0.027	0.021	0.025	0.027	0.039	0.072	0.026	0.040
	0.5	0.013	0.218	0.045	0.011	0.011	0.013	0.019	0.021	0.012	0.010	0.016	0.012	0.012
	0.9	0.018	0.514	0.100	0.013	0.029	0.030	0.021	0.029	0.032	0.017	0.031	0.023	0.027
							$W2$, $T = 1000$							
Z	0.1	0.008	0.132	0.070	0.005	0.220	0.017	0.007	0.007	0.006	0.033	0.100	0.006	0.006
	0.5	0.007	0.134	0.060	0.004	0.094	0.062	0.008	0.006	0.018	0.028	0.084	0.007	0.010
	0.9	0.013	0.137	0.104	0.006	0.038	0.033	0.015	0.009	0.028	0.010	0.029	0.011	0.015
T	0.1	0.018	0.144	0.028	0.009	0.010	0.010	0.010	0.010	0.009	0.023	0.063	0.010	0.010
	0.5	0.006	0.145	0.034	0.003	0.005	0.006	0.005	0.005	0.005	0.006	0.017	0.003	0.003
	0.9	0.009	0.374	0.047	0.010	0.008	0.011	0.010	0.009	0.009	0.011	0.023	0.007	0.008
							$W3$, $T = 100$							
Z	0.1	0.016	0.065	0.135	0.020	0.212	0.024	0.020	0.021	0.020	0.049	0.171	0.014	0.008
	0.5	0.032	0.397	0.209	0.010	0.092	0.066	0.012	0.013	0.023	0.056	0.155	0.010	0.020
	0.9	0.051	0.784	0.337	0.018	0.038	0.039	0.024	0.027	0.037	0.049	0.133	0.027	0.018
T	0.1	0.028	0.229	0.114	0.013	0.017	0.024	0.016	0.033	0.024	0.026	0.043	0.019	0.019
	0.5	0.030	0.401	0.163	0.011	0.018	0.017	0.011	0.013	0.019	0.030	0.092	0.009	0.015
	0.9	0.063	0.724	0.371	0.017	0.023	0.034	0.021	0.034	0.032	0.064	0.149	0.018	0.035
							$W3$, $T = 1000$							
Z	0.1	0.021	0.072	0.087	0.005	0.218	0.018	0.008	0.005	0.005	0.056	0.175	0.006	0.006
	0.5	0.044	0.114	0.110	0.005	0.095	0.055	0.005	0.006	0.017	0.054	0.161	0.009	0.017
	0.9	0.069	0.278	0.128	0.005	0.036	0.030	0.014	0.011	0.024	0.052	0.140	0.006	0.012
T	0.1	0.016	0.127	0.106	0.007	0.010	0.008	0.009	0.009	0.005	0.017	0.056	0.007	0.009
	0.5	0.023	0.192	0.105	0.003	0.006	0.011	0.004	0.004	0.005	0.035	0.102	0.005	0.010
	0.9	0.058	0.207	0.042	0.014	0.010	0.012	0.013	0.011	0.006	0.057	0.183	0.013	0.025

Note: RMSE with the size of sample agent $N = 100$, at the time period $T = N = 100$ and $T = 10 * N = 1000$ for the three network structures, respectively. u_{it}s are generated from a standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The simulation results are reported with 100 replications.

5 Application

We now explore the financial quantile network effects by analyzing the quantile connectedness among the stock returns. Anton and Polk (2014) find that stock returns tend to display significant comovements due to common active mutual fund owners. In addition, Pirinsky and Wang (2006) document strong comovements in the stock returns of firms headquartered in the same geographic area. Further, Garcia and Norli (2012) point out that the firms headquartered in the same geographic area have uniformly excessive returns compared to geographically dispersed firms, a phenomenon which is called the return local bias.

In this study, we investigate the two different financial network structures. First, we construct the common shareholder network (W_1), using information from the common mutual fund ownership. In particular, we let the stocks be connected if they are invested in by at least five common mutual fund owners. In addition, we construct the uniform headquarter location network (W_2) by checking whether the headquarters of companies are located in the same state or city. In particular, we treat the companies with headquarters located in the same state or city as connected.

We examine all the stocks traded in NYSE and NASDAQ in 2016. We collect the addresses associated with firms' headquarters from COMPUSTAT. The dataset on mutual fund holdings is downloaded from Thomson Reuters. After merging these data from the databases according to the unique trading code and moving the stocks with missing values, we finally obtain 928 stocks. These stock return data are downloaded from Datastream. We also select market capitalization, book value per share, cash flow and price-earning ratio as the individual firm-specific variables, which are then standardized. Finally, we consider VIX, Fama French three factors (excess market return, SMB, HML) as the common covariates to analyze the performance of stock returns under various market environments (different quantile levels).

We display the topology of two networks for the top 100 market-value stocks in Figure 1, where the two panels present the same stocks with the different network structures. The larger nodes imply the higher market capitalization while the darker nodes present the higher connectedness especially for the network structure with W1. There is a large connected group in the left panel, presenting the stocks that are more centralized connected by common investors, while the right panel reveals more small groups, implying that the stocks are more locally connected when measuring by uniform headquarter location. This is consistent with our expectation and the literature. Crucially, Figure 1 depicts totally different network structures for the same stocks.

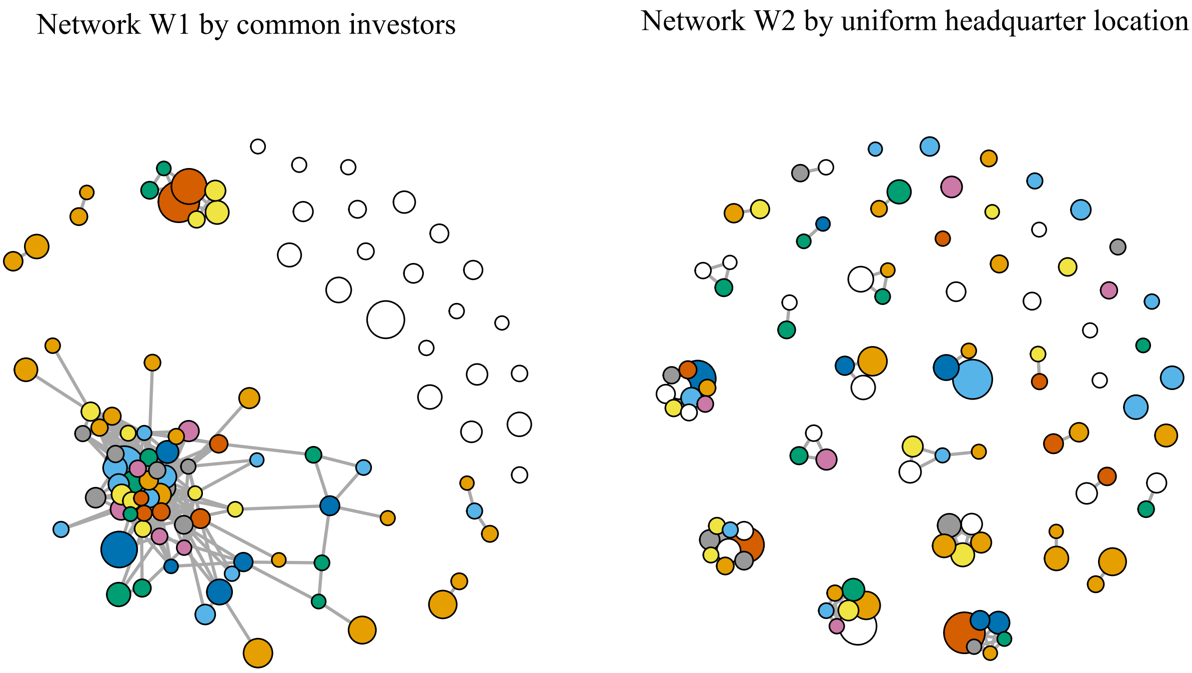


Figure 1: We depict the top 100 market value stocks in the selected 928 firms. The left panel: the common shareholder network W1 of the top 100 market value stocks, constructed by checking whether the stocks are invested in by at least five common mutual fund owners. The right panel: the uniform headquarter location network W2 for the same stocks, constructed by checking whether the headquarters of companies are located in the same state and city. The larger nodes imply higher market capitalization. The darker nodes present higher connectedness in network W1.

For comparison we present the estimation results for the proposed NQADL model together with the NQAD model without the common covariates, and the NQAR model without contemporaneous effects in $\overline{Y}_{it}$. To compare the performance of our proposed model relative to the alternative models, we evaluate the goodness of fit across the different quantiles, following for example Koenker and Machado (1999). Consider a linear model for the conditional quantile function,

$$Q_{Y_{it}}(\tau|X_{it}) = X_{1it}^{\top}\theta_1(\tau) + X_{2it}^{\top}\theta_2(\tau) \tag{24}$$

and let $\widehat{\theta}(\tau)$ denote the estimator of the unrestricted model, which is the minimizer of

$$\widehat{V}(\tau) = \min \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau \left\{Y_{it} - X_{it}^{\top} b\right\}. \tag{25}$$

$\widetilde{\theta}(\tau) = (\widetilde{\theta}_1^{\top}(\tau), 0)^{\top}$ denote the minimizer for the constrained model

$$\widetilde{V}(\tau) = \min \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau \left\{Y_{it} - X_{1it}^{\top} b_1\right\}. \tag{26}$$

$\widehat{\theta}(\tau)$ and $\widetilde{\theta}(\tau)$ denote the unrestricted and restricted quantile regression estimates. Define the goodness-of-fit criterion as

$$R(\tau) = 1 - \frac{\widehat{V}(\tau)}{\widetilde{V}(\tau)}, \tag{27}$$

which represents the overall decreased percentage of loss function in quantile regression of the unrestricted model compared with the restricted model.

First, we report the estimation results for the network with W1 in Table 5. For convenience we present the coefficients and the standard errors multiplied by 10^2 at the different quantiles ($\tau = 0.1, 0.5, 0.9$). The goodness of fit value $R(\tau)$ is reported in the last row of Table 5 (Goodn.fit.). Table 5 reveals that the contemporaneous effects in $\overline{Y}_{it}$ are significant and non-negligible across different quantile levels. The overall loss function of the proposed NQADL model falls off around 4% in comparison with the NQAR model without contemporaneous effects. The NQAD model without common factors performs similarly to the NQADL model. Further, the effects of common factors are significant for extreme values or economic environments, i.e. $\tau = 0.1, 0.9$ in comparison with the central circumstance, i.e. $\tau = 0.5$.

Meanwhile, we display the estimated quantile regression effects across the different quantiles in Figure 2. The dashed line is the quantile regression coefficients on these variables, and the grey areas indicate a rank test-based confidence band across different

	NQADL			NQAR			NQAD		
	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
$\widehat{\gamma}_0$	-2.09^{***}	0.02^{***}	2.10^{***}	-2.16^{***}	0.07^{***}	2.31^{***}	-2.09^{***}	0.02^{***}	2.10^{***}
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
$\widehat{\gamma}_1$	95.54^{***}	79.32^{***}	86.36^{***}	-	-	-	94.92^{***}	79.47^{***}	86.93^{***}
	(0.66)	(0.44)	(0.67)				(0.65)	(0.43)	(0.66)
$\widehat{\gamma}_2$	1.72^{***}	0.81^{**}	1.39^{**}	-0.47	-1.40^{***}	2.52^{***}	1.48^{***}	0.84^{**}	1.91^{***}
	(0.56)	(0.35)	(0.57)	(0.71)	(0.18)	(0.51)	(0.56)	(0.35)	(0.51)
$\widehat{\gamma}_3$	-1.18^{***}	-2.19^{***}	-1.10^{***}	-0.27	-1.88^{***}	-0.75^{**}	-1.18^{***}	-2.23^{***}	-1.08^{***}
	(0.31)	(0.24)	(0.33)	(0.38)	(0.13)	(0.35)	(0.32)	(0.24)	(0.33)
SIZE	0.06^{***}	0.00	-0.05^{***}	0.08^{***}	0.00	-0.08^{***}	0.06^{***}	0.00	-0.05^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BM	0.10^{***}	0.00	-0.10^{***}	0.10^{***}	0.01^{***}	-0.10^{***}	0.10^{***}	0.00	-0.10^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Cash	0.01	0.00	-0.02^{***}	0.02^{***}	0.00	-0.03^{***}	0.01	0.00	-0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PE	0.05^{***}	0.00	-0.02^{***}	0.02^{***}	0.01^{***}	-0.02^{***}	0.05^{***}	0.00	-0.02^{***}
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
VIX	0.01	-0.01	-0.03^{**}	0.10^{***}	-0.08^{***}	-0.10^{***}	-	-	-
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)			
Rm - Rf	0.05^{***}	-0.01^{*}	-0.04^{***}	-0.12^{***}	0.03^{***}	-0.03^{***}	-	-	-
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)			
SMB	-0.01	0.00	-0.02^{**}	-0.09^{***}	-0.05^{***}	-0.02	-	-	-
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)			
HML	0.03^{***}	-0.01^{**}	-0.05^{***}	-0.12^{***}	-0.09^{***}	-0.11^{***}	-	-	-
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)			
Goodn.fit.	-	-	-	4.40	5.49	2.76	0.04	0.02	0.12

Table 5: Estimation with network W1 for a US stock dataset of 928 stocks. The network W1 is constructed by checking whether the stocks are invested in by at least five common mutual fund owners. The parameter estimates ($\times 10^2$) are reported for quantile levels $\tau = 0.1, 0.5, 0.9$, and the value below in parentheses is the corresponding standard error ($\times 10^2$). Structure NQAR (SNQAR) denotes our model, NQAR denotes the model without simultaneous connectedness effects, and SNQAR without common factors model denotes the model excluding the common economic factors. Goodn.fit. ($\times 10^2$) represents the goodness of fit of our model with the others. The significance levels of 1%, 5% and 10% are noted by ***, **, * respectively.

quantile levels, i.e. $\tau = 0.1, 0.2, .., 0.9$. The band clearly excludes the null effect indicated by the solid horizontal line. Figure 2 shows that all the dynamics of parameters $\gamma_1(.), \gamma_2(.), \gamma_3(.)$ at various quantile levels are nearly featured with a "U" shape and the confidence bands consistently exclude the zero line, which implies that both the contemporaneous and lagged network effects are distinctly important at various quantile levels, and the effects are relatively stronger at the tail levels. Additionally, the contemporaneous network effects (γ_1) are comparatively more significant and larger than the others. Besides, most of the common factors affect the stock returns in a decreasing trend as quantile levels increase.

Finally, in order to do the robustness check, when using network W2 to implement the application, we obtain similar results, see Table 6 and Figure 3. The formats follow Table 5 and Figure 2. In Table 6, although the contemporaneous effects of the NQADL model are slightly smaller than the counterparts in Table 5, the patterns in comparison with the NQAR model and NQAD model almost mimic the ones in Table 5. The overall loss function of the proposed NQADL model decreases around 6% for the three selected quantile levels $\tau = 0.1, 0.5, 0.9$ compared with NQAR model without the synchronous network effects. Meanwhile, the scheme of Figure 3 is analogous to that in Figure 2.

6 Conclusion

We propose a dynamic spatial autoregressive quantile network model that allows for temporal and cross-sectional dependence. Using the predetermined network information, we study the dynamic tail event driven risk under various quantile levels within a network topology. The model's distinguishing characteristic is that a given nodal response's behavior is not only influenced by its previous information, but also connected with a weighted aggregation of simultaneous and lagged responses from others. Moreover, common covariates representing macroeconomic environments are also incorporated.

The main challenge of our proposed NQADL model is the potential endogeneity problem due to the simultaneous network effects. We extend the IVQR method in the model estimation, and provide the associated asymptotic theory for the IVQR estimator. Sim-

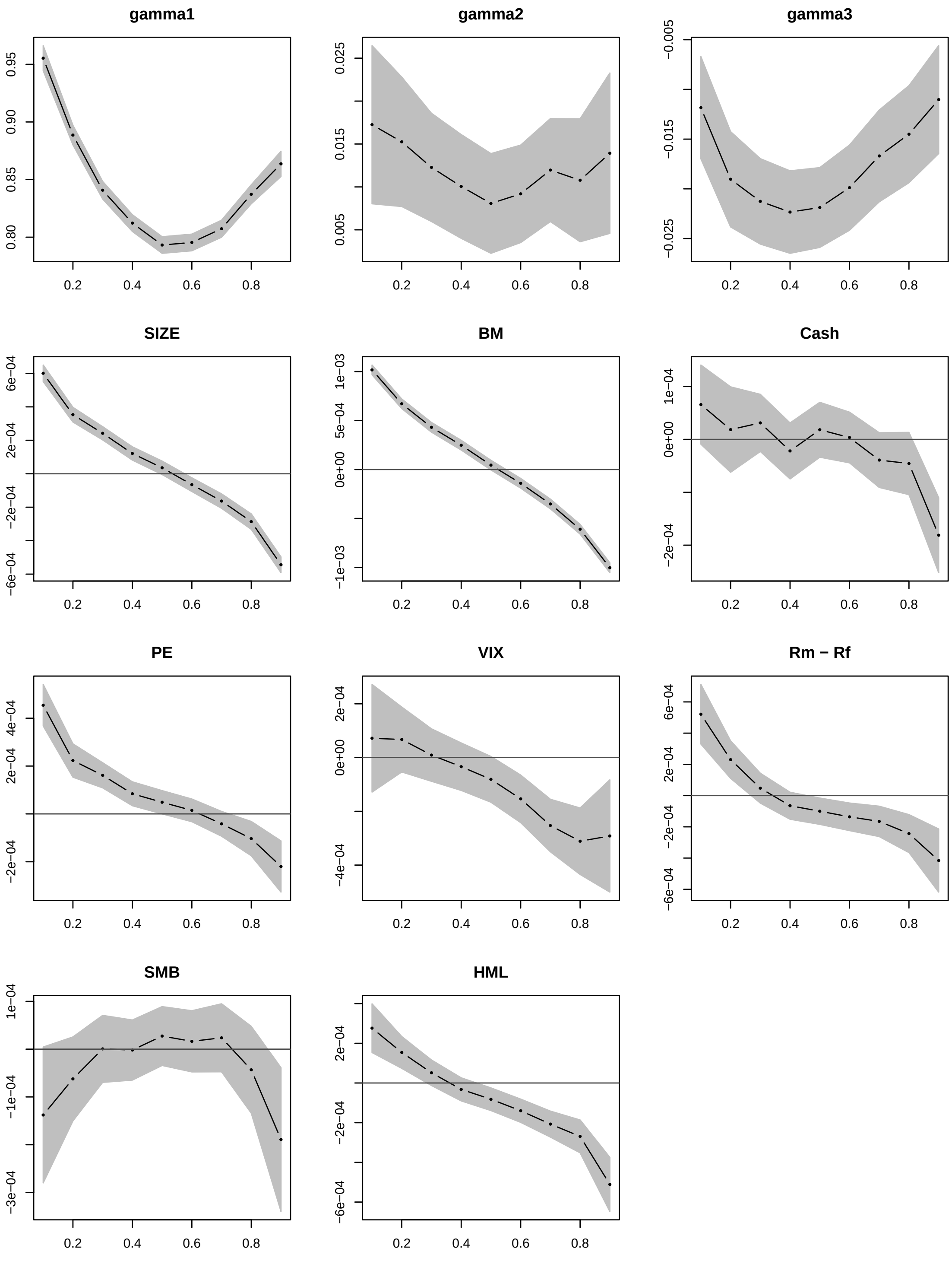


Figure 2: Estimated coefficients associated with some selected variables with the network W1. The dashed line is the estimated coefficients for these variables, and the grey region indicates a rank test-based confidence band for the estimators. The quantile levels $\tau = 0.1, 0.2, 0.3..., 0.9$ are used for calculation. The solid horizontal line is zero. The network W1 is constructed by checking whether the stocks are invested in by at least five common mutual fund owners.

	NQADL			NQAR			NQAD		
	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
$\widehat{\gamma}_0$	-2.04^{***}	0.02^{***}	2.06^{***}	-2.16^{***}	0.07^{***}	2.31^{***}	-2.04^{***}	0.02^{***}	2.07^{***}
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
$\widehat{\gamma}_1$	86.73^{***}	73.06^{***}	79.47^{***}	-	-	-	86.34^{***}	73.24^{***}	79.71^{***}
	(0.59)	(0.42)	(0.58)				(0.58)	(0.42)	(0.58)
$\widehat{\gamma}_2$	1.09^{*}	0.45	1.47^{**}	0.26	-1.57^{***}	2.14^{***}	1.06^{*}	0.54	1.95^{***}
	(0.58)	(0.34)	(0.57)	(0.74)	(0.25)	(0.68)	(0.59)	(0.35)	(0.57)
$\widehat{\gamma}_3$	-1.05^{***}	-1.80^{***}	-0.91^{***}	-0.34	-1.86^{***}	-0.73^{***}	-1.16^{***}	-1.79^{***}	-0.81^{***}
	(0.32)	(0.25)	(0.26)	(0.38)	(0.13)	(0.19)	(0.32)	(0.25)	(0.28)
SIZE	0.06^{***}	0.00	-0.06^{***}	0.08^{***}	0.00	-0.08^{***}	0.06^{***}	0.00	-0.06^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BM	0.10^{***}	0.00	-0.09^{***}	0.10^{***}	0.01^{***}	-0.10^{***}	0.10^{***}	0.00	-0.09^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Cash	0.02^{***}	0.00	-0.02^{***}	0.02^{***}	0.00	-0.03^{***}	0.02^{***}	0.00	-0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PE	0.03^{***}	0.01^{***}	0.00	0.02^{***}	0.01^{***}	-0.02^{***}	0.02^{***}	0.01^{***}	0.00
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
VIX	0.02^{*}	-0.01^{***}	-0.03^{**}	0.10^{***}	-0.08^{***}	-0.11^{***}	-	-	-
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)			
Rm - Rf	0.03^{***}	-0.01	-0.05^{***}	-0.11^{***}	0.03^{***}	-0.03^{**}	-	-	-
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)			
SMB	-0.01	-0.01^{**}	-0.02^{**}	-0.09^{***}	-0.05^{***}	-0.01	-	-	-
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)			
HML	0.01	-0.01^{***}	-0.06^{***}	-0.12^{***}	-0.09^{***}	-0.11^{***}	-	-	-
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)			
Goodn.fit.	-	-	-	6.17	6.66	4.54	0.00	0.02	0.09

Table 6: Estimation with the network W2 for a US stock dataset consisting of 928 stocks. The network W2 is constructed by checking whether the headquarters of companies are located in the same state and city. The parameter estimates ($\times 10^2$) are reported for quantile levels $\tau = 0.1, 0.5, 0.9$, and the value in parentheses below is the corresponding standard error ($\times 10^2$). Structure NQAR (SNQAR) denotes our model, NQAR denotes the model without simultaneous connectedness effects, and SNQAR without common factors model denotes the model excluding the common economic factors. Goodn.fit. ($\times 10^2$) represents the goodness of fit of our model with the others. The significance levels of 1%, 5% and 10% are noted by ***, **, * respectively.

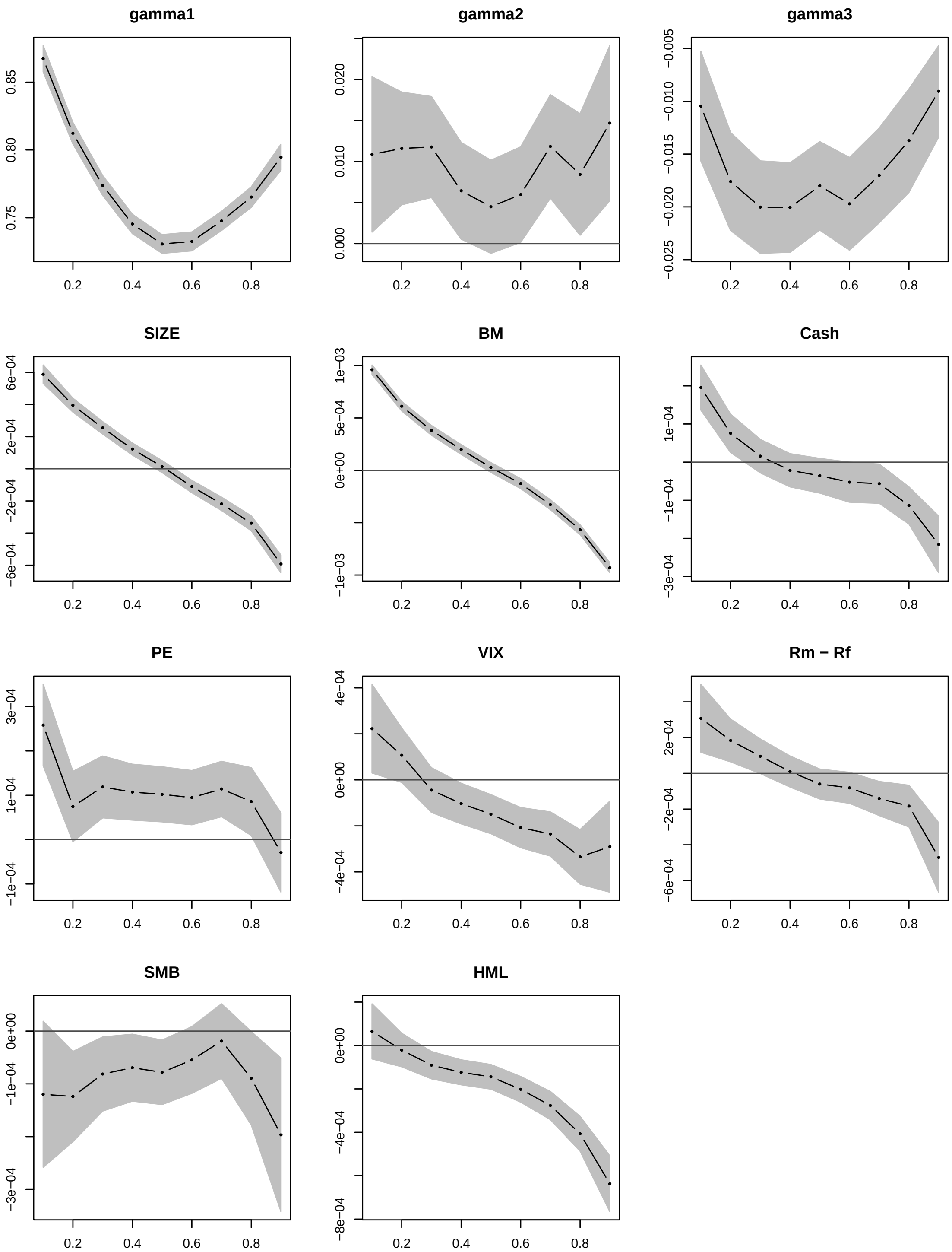


Figure 3: Estimated coefficients of some selected variables with the network W2. The dashed line is the estimated value for these coefficients, and the grey region indicates a rank test-based confidence interval for the estimators. The quantile levels $\tau = 0.1, 0.2, 0.3..., 0.9$ are used for the calculation. The solid horizontal line is zero. The network W2 is constructed by checking whether the headquarters of companies are located in the same state and city.

ulation results under various scenarios show that our model performs well with different network structures and at various quantile levels, especially when the node sizes increase. In addition, an empirical study for US stock markets finds that the contemporary network effects are distinct and significant across different quantile levels while the common covariates affect relatively more significantly when considering extreme tail dependence. Overall, our proposed NQADL model favorably outperforms the NQAR model without synchronous network effects.

7 Appendix

7.1 Proof of Lemma 2.1

(i) Strict Stationarity

We first discuss the strict stationarity of $\{\mathbb{Y}_t\}_t$. Under Assumption 2.1(1), we have $(I - \mathbf{A}_{1t}W)^{-1} = \sum_{k=0}^{\infty}(\mathbf{A}_{1t}W)^k$. Then, we obtain the reduced form of the model (3) by

$$\begin{aligned}\mathbb{Y}_t &= (I - \mathbf{A}_{1t}W)^{-1}H_t\mathbb{Y}_{t-1} + (I - \mathbf{A}_{1t}W)^{-1}(\Gamma + B_t\mathbb{F}_t + V_t) \\ &= M_t\mathbb{Y}_{t-1} + \mathbb{C}_t,\end{aligned}$$

where $M_t = (I - \mathbf{A}_{1t}W)^{-1}H_t$, and $\mathbb{C}_t = (I - \mathbf{A}_{1t}W)^{-1}(\Gamma + B_t\mathbb{F}_t + V_t)$. Hence, this process falls into the class of a general autoregressive process with $\{M_t, \mathbb{C}_t, t \in \mathbb{Z}\}$. According to Theorem 1.3 and Lemma 2.1 of Bougerol and Picard (1992), $\mathbb{Y}_t$ has a strictly stationary solution, if the sequence of random matrices $\{M_t, t \in \mathbb{Z}\}$ satisfy the following two conditions:

a) $\mathsf{E}\log|M_0|_2^+ < \infty$, with $\log|M_0|_2^+ = \max(\log|M_0|_2, 0)$,

b) $\lim_{t\to\infty}|M_0M_{-1}\cdots M_{-t}|_2 = 0$ almost surely.

Recall that $|A|_2 = \sup_{\{v\in\mathbb{R}^d, v\neq 0\}}|Av|_2/|v|_2$, where we recall that $|.|_2$ is the two norm of a vector or matrix. We now prove the two conditions. First, consider:

$$
\begin{aligned}
|M_t|_2 &= |(I - \mathbf{A}_{1t}W)^{-1}H_t|_2 \\
&\leqslant |\sum_{k\geqslant 0}(\mathbf{A}_{1t}W)^k H_t|_2 \quad \text{under (Assumption 2.1 A.1)} \\
&\leqslant \sum_{k\geqslant 0}|(\mathbf{A}_{1t}W)^k H_t|_2 \quad \text{under (Minkowski inequality)} \\
&\leqslant \sum_{k\geqslant 0}|(\mathbf{A}_{1t})|_2^k |W|_2^k |H_t|_2 \quad \text{under } (\mathbf{A.1}) \\
&\leqslant \sum_{k\geqslant 0}|(\mathbf{A}_{1t})|_2^k (|(\mathbf{A}_{2t})|_2 + |(\mathbf{A}_{3t})|_2) \quad \text{under } (\mathbf{A.1}) \\
&= \sum_{k\geqslant 0}\{\max_i |\gamma_1(U_{it})|\}^k (\max_i |\gamma_2(U_{it})| + \max_i |\gamma_3(U_{it})|) \\
&\leqslant \sum_{k\geqslant 0} c_1^k (\max_i |\gamma_2(U_{it})| + \max_i |\gamma_3(U_{it})|) \\
&\leqslant \sum_{k\geqslant 0} c_1^k c_{23} \quad \text{under } (\mathbf{A.2}) \\
&\leqslant c_{23}/(1-c_1).
\end{aligned}
$$

Then, $\mathsf{E}\log|M_0|_2^+ \leqslant \log \mathsf{E}\,|M_0|_2^+ \leqslant \max\{\log(c_{23}/(1-c_1)), 0\} < \infty$, such that the conditions a) holds.

Next, consider the second condition b), which can be written as,

$$
\begin{aligned}
\mathsf{E}\,|M_0 M_{-1}\cdot M_{-t}|_2 &\leqslant \mathsf{E}\,\Pi_{l=-t}^0\{(\sum_{k\geqslant 0} c_1^k)(\max_i |\gamma_2(U_{il})| + \max_i |\gamma_3(U_{il})|)\} \\
&\leqslant (1-c_1)^{-t-1}[\mathsf{E}\{\max_i |\gamma_2(U_{il})| + \max_i |\gamma_3(U_{il})|\}]^{t+1} \\
&\leqslant (1-c_1)^{-t-1} c_{23}^{t+1}.
\end{aligned}
$$

For a small constant $\varepsilon > 0$, we now have:

$$
\begin{aligned}
&\sum_{t=1}^{\infty} \mathrm{P}(|M_0 M_{-1}\cdot M_{-t}|_2 > \varepsilon) \\
\leqslant\ & \sum_{t=1}^{\infty} \frac{\mathsf{E}\,|M_0 M_{-1}\cdot M_{-t}|_2}{\varepsilon} \quad \text{under (Markov's inequality)} \\
\leqslant\ & \sum_{t=1}^{\infty}(1-c_1)^{-t-1} c_{23}^{t+1}/\varepsilon \\
=\ & \frac{(c_{23})^2}{(1-c_1)(1-c_1-c_{23})\varepsilon} < \infty.
\end{aligned}
$$

Then, by Borel-Cantelli lemma, the condition b) holds. Therefore any projection of the process in (3) has a strictly stationary solution. If in addition $\mathsf{Var}(\mathbb{Y}_t)$ and $\Gamma_l = \text{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-l})$ exits, then we can conclude the covariance stationarity of the process.

(ii) Covariance Stationarity

Recall that the model (3) admits the following covariance stationary solution

$$\mathbb{Y}_t = \sum_{l=0}^{\infty} \Pi_l \mathbb{D}_{t-l} = \sum_{l=0}^{\infty} \Pi_l S_{t-l}^{-1} B_{t-l} \mathbb{F}_{t-l} + \sum_{l=0}^{\infty} \Pi_l S_{t-l}^{-1} \mathbf{A}_{0t},$$

where $\mathbb{D}_t = S_t^{-1}(B_t \mathbb{F}_t + \mathbf{A}_{0t})$, $M_t = S_t^{-1} H_t$ and $\Pi_l = M_t \times \cdots \times M_{t-l+1}$ for $l > 1$ with $\Pi_0 = I$ and $\Pi_1 = M_t$. Let $\mathsf{E}\, M_t = M$ and $\mathsf{E}\, \mathbb{D}_t = \mathbb{D}$. Moreover, $|\mathbb{D}_t|_\infty \leqslant (1-c_1)^{-1}(d_f + d_z)$ by Assumption 2.1(3). Thus, the expected value of $\mathbb{Y}_t$ is given by $\mu_{\mathbb{Y}} = (I - M)^{-1}\mathbb{D}$. Further, we have: $|M_t I_N| \leqslant_a c_{23}/(1 - c_1) I_N$ for every t, where $\leqslant_a$ denotes 'element-wise smaller'. The variance and covariance of $\mathbb{Y}_t$ are then given by

$$\Gamma_0 = \mathsf{Var}(\mathbb{Y}_t) = \mathsf{E}\left\{ \left(\sum_{l=0}^{\infty} \Pi_l \mathbb{D}_{t-l} \right) \left(\sum_{l=0}^{\infty} \Pi_l \mathbb{D}_{t-l} \right)^\top \right\} - \mu_{\mathbb{Y}} \mu_{\mathbb{Y}}^\top,$$
$$\Gamma_l = \text{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-l}) = \mathsf{E}\left\{ \left(\sum_{l=0}^{\infty} \Pi_l \mathbb{D}_{t-l} \right) \left(\sum_{l=0}^{\infty} \Pi_l \mathbb{D}_{t-2l} \right)^\top \right\} - \mu_{\mathbb{Y}} \mu_{\mathbb{Y}}^\top.$$

Consider $\mathsf{Var}(\mathbb{Y}_t)$. Let $c' = (1 - c_1)^{-1} c_{23}$, then we have:
$e_i^\top \mathsf{E}\{(\sum_{l \geqslant 0} \Pi_l \mathbb{D}_{t-l})((\sum_{l \geqslant 0} \Pi_l \mathbb{D}_{t-l}))^\top\} e_j = I_l + I_2 + I_3$.

First, we show that $I_1 = \sum_{l_1 > l_2} e_i^\top \otimes e_j^\top M^{l_2} \otimes M^{l_2} \mathsf{E}\{I \otimes M_{t-l_2} \cdots I \otimes M_{t-l_1-1} \text{vec}(\mathbb{D}_t \mathbb{D}_{t-l_2}^\top)\} \leqslant \sum_{l_1 > l_2} |\mathbb{D}|_{\max} \mathsf{E}(d_z + d_f)\{c'\}^{l_1 - l_2} c'^{2l_2} = \sum_{l_1 > l_2} |\mathbb{D}|_{\max} \mathsf{E}(d_z + d_f) c'^{l_1 + l_2} < \infty$. Similarly, $I_2 = \sum_{l_1} e_i^\top \otimes e_j^\top M^{l_1} \otimes M^{l_1} \mathsf{E}\{\text{vec}(\mathbb{D}_{t-l_1} \mathbb{D}_{t-l_1}^\top)\} \leqslant \sigma_{d\max} \sum_{l_1} c'^{2l_1}$ and $I_3 = \sum_{l_2 > l_1} e_i^\top \otimes e_j^\top M^{l_1} \otimes M^{l_1} \mathsf{E}\{M_{t-l_1} \otimes I \cdots M_{t-l_1-1} \otimes I \text{vec}(\mathbb{D}_{t-l_1} \mathbb{D}_t^\top) \leqslant \sum_{l_1 < l_2} |\mathbb{D}|_{\max} \mathsf{E}(d_z + d_f) c'^{l_1 + l_2}$. Thus, we have $I_1, I_2, I_3 < \infty$ under Assumption 2.1. Similarly, we can show that $\Gamma_l = \text{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-l})$ exists. Thus, $a^\top \mathbb{Y}_t$ is covariance stationary.

7.2 Proof of Theorem 1

7.2.1 Invariance Principle

In order to to prove Theorem 1, We shall apply Theorem 3 in Wu (2011). We introduce the functional dependence measure.

Definition 7.1. *Define* $X_t = g(\mathcal{F}_t)$ *with the shift process* $\mathcal{F}_t = (..., \xi_{t-1}, \xi_t)$*. Let* ξ_0 *be replaced by an i.i.d. copy of* ξ_0^**, and* $X_t^* = g(\mathcal{F}_t^*)$ *with* $\mathcal{F}_t^* = (\ldots, \xi_{-1}, \xi_0^*, \xi_1, \ldots, \xi_{t-1}, \xi_t)$*. For* $q \geqslant 1$*, define the functional dependence measure* $\delta_{q,t}(X_t) \stackrel{\text{def}}{=} \|X_t - X_t^*\|_q$*, which measures the dependency of* ξ_0 *on* X_t*. Also, define* $\Delta_{m,q}(X_t) \stackrel{\text{def}}{=} \sum_{t=m}^{\infty} \delta_{q,t}$*, which measures the cumulative effect of* ξ_0 *on* $X_{t\geqslant m}$*. Finally, define the predictive dependence measure by* $\mathcal{P}_j X_t = \mathsf{E}(X_t|\mathcal{F}_j) - \mathsf{E}(X_t|\mathcal{F}_{j-1})$*.*

According to Theorem 3 in Wu (2011), if the condition $\sum_{t=0}^{\infty} \|\mathcal{P}_0(a^\top \tilde{\mathbb{Y}}_t)\|_q < \infty$ holds, then we can obtain the main result (7) in Theorem 1. We now prove that $\Delta_{0,q}(a^\top \mathbb{Y}_t) < \infty$. (We note that $\Delta_{0,q}(a^\top \mathbb{Y}_t) < \infty$ and $\|\mathcal{P}_0(a^\top \tilde{\mathbb{Y}}_t)\|_q$ are equivalent measures.) Notice that

$$\begin{aligned}
\mathcal{P}_0(a^\top \tilde{\mathbb{Y}}_t) &= a^\top(\mathsf{E}(\tilde{\mathbb{Y}}_t|\mathcal{F}_0) - \mathsf{E}(\tilde{\mathbb{Y}}_t|\mathcal{F}_{-1})) \\
&= a^\top(M^t(\mathbb{D}_0 - \mathbb{D}) + \sum_{l\geqslant t} M^{t-1}(M_0 - M)M_{-1}\cdots M_{t-l+1}\mathbb{D}_{t-l}) \\
&\leqslant a^\top(M^t \mathbf{I}_N 2(d_z + d_f) + \sum_{l\geqslant t} M^{t-1}(M_0 - M)M_{-1}\cdots M_{t-l+1}\mathbf{I}_N(d_z + d_f)) \\
&\leqslant a^\top(c'^t + \sum_{l\geqslant t} 2c'^l)\mathbf{I}_N 2(d_z + d_f) \\
&\leqslant 4(d_z + d_f)c'^t/(1 - c')
\end{aligned}$$

where $c' = (1-c_1)^{-1}c_{23}$. Thus, $\Delta_{0,q}(a^\top \mathbb{Y}_t) \lesssim \sum_{t\geqslant 0}(\|d_z\|_q + \|d_f\|_q)c'^t/(1-c') < \infty$. Hence, the conditions are satisfied.

7.3 Proof of Theorem 2 and Theorem 3

As we handle the statistic object involved with weak spatial temporal dependence, we shall condition on $\mathcal{C} \stackrel{\text{def}}{=} \{\mathbb{F}_t\}_t$ thoughout the proof. The E in the proof are conditioning on $\mathcal{C}$ without special notice.

7.3.1 Lemmas for near-epoch dependent (NED) Processes

Let (X_{it}, U_{it}) be the basis of NED processes. Then, we provide a number of Lemmas on the basic properties of NED in random fields. The derivation follows largely Xu and Lee (2015) with substantial modifications to fit to our model setup.

Lemma A.1. *If $\{Y_{it}\}$ and $\{Z_{it}\}$ are both uniformly $L_{2+\delta}$ bounded, and uniformly and geometrically L_2-NED, then $\{Y_{it}Z_{it}\}$ is uniformly and geometrically L_2-NED.*

Lemma A.2. *For $h \geqslant 1$, there exists some $\pi_1 < \infty$, such that the number of all elements in D_{NT} located within a distance $[h, h+1)$ satisfying $\sum_{j \in D_{NT}: h \leqslant \rho(i,j) < h+1} \leqslant \pi_1 h^{d-1}$ for any $i \in D_{NT}$. This is from Lemma A.1 in Jenish and Prucha (2009).*

Lemma A.3. *Suppose W is an $N \times N$ square matrix which can be decomposed into the sum of two $N \times N$ matrices, i.e., $W = A + B$. Denote $|A|_{max} = \max_{ij} |a_{ij}|, i, j = 1, \cdots, N$. Then for any positive integer l, we have $(W^l - B^l)_{ij} \leqslant |A|_{max} \sum_{k=0}^{l-1} \|B\|_\infty^k \|W\|_1^{l-1-k}$.*

Proof. Let $e_k = (0, \cdots, 0, 1, 0, \cdots, 0)^\top$ is the unit column vector with one in its kth entry and zeros in its other entries. By expansion, $W^l - B^l = \sum_{k=0}^{l-1} B^k A W^{l-1-k}$. Then $(W^l - B^l)_{ij} = \sum_{k=0}^{l-1} e_i^\top B^k A W^{l-1-k} e_j$. For any matrix M and a vector e of dimension n, $\|Me\|_\infty \leqslant |M|_{max} \|e\|_1$. Hence, $e_i^\top B^k A W^{l-1-k} e_j \leqslant \|e_i^\top B^k\|_\infty \|A W^{l-1-k} e_j\|_\infty \leqslant \|B^k\|_\infty |A|_{max} \|W^{l-1-k} e_j\|_1 \leqslant |A|_{max} \|B^k\|_\infty \|W\|_1^{l-1-k}$, for any integer $k = 0, \cdots, l-1$. □

Lemma A.4. *For any $\alpha > 0$ and $s \geqslant 2$, $\sum_{h=[s]}^\infty h^{-\alpha-1} < \frac{2^{\alpha+1}}{\alpha} s^{-\alpha}$. $[s]$ denotes the largest integer less than or equal to s.*

Proof. For $h \geqslant 2$, $h \geqslant \frac{h+1}{2}$. When $\alpha > 0$, $\sum_{h=[s]}^\infty h^{-\alpha-1} \leqslant \sum_{h=[s]}^\infty (\frac{h+1}{2})^{-\alpha-1} \leqslant 2^{\alpha+1} \int_s^\infty x^{-\alpha-1} dx = \frac{2^{\alpha+1}}{\alpha} s^{-\alpha}$. Therefore, $\sum_{h=[s]}^\infty h^{-\alpha-1} < \frac{2^{\alpha+1}}{\alpha} s^{-\alpha}$. □

Lemma A.5. *If $\{Y_{it}\}$ and $\{Z_{it}\}$ are both uniformly $L_{2+\delta}$ bounded, and uniformly and geometrically L_p-NED, then $\{Y_{it} - Z_{it}\}$ and $\{Y_{it} + Z_{it}\}$ are both uniformly and geometrically L_p-NED.*

Proof. Define $\|Z_{it} - \mathsf{E}(Z_{it}|\mathscr{F}_{it}(s))\|_p < d_{it}^Z \varphi(s)^Z$ and $\|Y_{it} - \mathsf{E}(Y_{it}|\mathscr{F}_{it}(s))\|_p < d_{it}^Y \varphi(s)^Y$. By Minkowski's inequality, $\|(Y_{it} - Z_{it}) - \mathsf{E}(Y_{it} - Z_{it}|\mathscr{F}_{it}(s))\|_p \leqslant \|Y_{it} - \mathsf{E}(Y_{it}|\mathscr{F}_{it}(s))\|_p +$

$\|Z_{it} - \mathsf{E}(Z_{it}|\mathscr{F}_{it}(s))\|_p < d_{it}^Y \varphi(s)^Y + d_{it}^Z \varphi(s)^Z < d_{it}\varphi(s)$, with $d_{it} = \max(d_{it}^Y, d_{it}^Z)$ and $\varphi(s) = \varphi(s)^Y + \varphi(s)^Z$. Similar results for $\{Y_{it} + Z_{it}\}$. □

Lemma A.6. *(Ibragimov and Linnik (1971)) Let $L_p(\mathscr{F}_1)$ and $L_p(\mathscr{F}_2)$ denote the class of $\mathscr{F}_1-$measurable and $\mathscr{F}_2-$measurable random variables x with $\|x\|_p < \infty$. Let $X \in L_p(\mathscr{F}_1)$ and $Y \in L_p(\mathscr{F}_2)$. Then, for any $1 \leqslant p, q, r < \infty$ such that $p^{-1} + q^{-1} + r^{-1} = 1$,*

$$|Cov(X, Y)| < 4\alpha^{1/r}(\mathscr{F}_1, \mathscr{F}_2)\|X\|_p\|Y\|_q,$$

where $\alpha(\mathscr{F}_1, \mathscr{F}_2) = \sup_{A\in\mathscr{F}_1, B\in\mathscr{F}_2}(|P(AB) - P(A)P(B)|)$.

Lemma A.7. *Under assumptions 2.1 (A.1), 3.1 - 3.2(1),*

(1) $\Gamma_u = \|W\|_1 < \infty$, and $\Gamma_w = \|W\|_\infty = 1$.

(2) For any N and positive integer h, $\|W^h\|_1 < hK\Gamma_u$.

Proof. (1) Using Lemma A.2, $\|W\|_1 = \sup_j \sum_{i=1}^N |\omega_{ij}| = \sup_j \sum_{h=1}^\infty \sum_{i:h\leqslant\rho(i,j)<h+1} |\omega_{ij}| \leqslant \sup_j \sum_{h=1}^\infty \sum_{i:h\leqslant\rho(i,j)<h+1} \pi_0 h^{-c_w} \leqslant \sum_{h=1}^\infty \pi_1 h^{d-1}\pi_0 h^{-c_w} < \infty$, due to $c_w > d$ (Assumption 3.2 (1)). As W is row-normalized in our model, thus $\Gamma_w = \|W\|_\infty = 1$.

(2) Denote an index set V_N with $\sum_{i=1}^N |\omega_{ij}| > \Gamma_w$ if $j \in V_N$ and $\sum_{i=1}^N |\omega_{ij}| \leqslant \Gamma_w$ if $j \notin V_N$. By assumption 3.2(1), $|V_N| \leqslant K$ for any N. Denote $e_k = (0, \cdots, 0, 1, 0, \cdots, 0)^\top$ as a unit column vector with kth entry one and other entries zeros. $e = (1, \cdots, 1)^\top = \sum_{k=1}^N e_k$. Note that $I_N = \sum_{j=1}^N e_j e_j^\top$. The kth column sum of W^h, i.e., $e^\top W^h e_k$ can be transformed as,

$$\begin{aligned}
e^\top W^h e_k &= \sum_{j=1}^N e^\top W e_j e_j^\top W^{h-1} e_k \\
&= \sum_{j\in V_N} e^\top W e_j e_j^\top W^{h-1} e_k + \sum_{j\notin V_N} e^\top W e_j e_j^\top W^{h-1} e_k \\
&\leqslant K\left(\max_{j\in V_N} e^\top W e_j\right)\left(\max_{j\in V_N} e_j^\top W^{h-1} e_k\right) + \left(\max_{j\notin V_N} e^\top W e_j\right)\sum_{j\notin V_N} e_j^\top W^{h-1} e_k \\
&\leqslant K\Gamma_u\|W^{h-1}\|_\infty + \Gamma_w\|W^{h-1}\|_1 \\
&\leqslant K\Gamma_u + \|W^{h-1}\|_1, \quad \forall k = 1, \cdots, N.
\end{aligned}$$

Hence, we have $\|W^h\|_1 \leqslant K\Gamma_u + \|W^{h-1}\|_1$. By deduction, it can be achieved that $\|W^h\|_1 \leqslant (h-1)K\Gamma_u + \Gamma_u \leqslant hK\Gamma_u$. □

7.3.2 Proof of Proposition 3.1

Proof. (1) We first discuss the NED properties of $\{Y_{it}\}_{i,t}$. Following Jenish and Prucha (2012), the NED property is satisfied if random fields are generated from nonlinear Lipschitz type functions on random field $\{X_{it}, U_{it}\}$. Notice that $\int_0^1 \gamma_1(u)du \leqslant \max_u \gamma_1(u)$. Define $\{\mathcal{F}_{it}(s) = \sigma(X_{i',t'}, U_{i',t'}, \mathcal{C}) : |i-i'| \leqslant s, |t-t'| \leqslant s\}$, $\Pi^s_{l,i,t} \overset{\text{def}}{=} \mathsf{E}(\Pi_l|\mathcal{F}_{it}(s))$, $\mathbb{D}^s_{t-l,i,t} = \mathsf{E}(\mathbb{D}_{t-l}|\mathcal{F}_{it}(s))$, and $\tilde{\Pi}_{l-s+1} = M_{t-s-1} \cdots M_{t-l+1}$. $\Pi_{s+1,i,t}$ Conditioning on $\mathcal{C}$,

$$
\begin{aligned}
& ||Y_{it} - \mathsf{E}\{Y_{it}|\mathcal{F}_{it}(s)\}||_2 \\
\leqslant\ & ||e_i^\top(\sum_{l\geqslant 0}\Pi_l\mathbb{D}_{t-l} - \sum_{l\geqslant 0}\mathsf{E}(\Pi_l\mathbb{D}_{t-l}|\mathcal{F}_{it}(s)))||_2 \\
\leqslant\ & ||e_i^\top\sum_{l\leqslant s}(|\Pi_l|_a + |\Pi^s_{l,i,t}|_a)|\mathbb{D}_{t-l} - \mathbb{D}^s_{t-l,i,t}|_a||_2 \\
+\ & ||e_i^\top\sum_{l>s}(|\Pi_{s+1}|_a + |\Pi^s_{s+1,i,t}|_a)|\tilde{\Pi}_{l-s+1}\mathbb{D}_{t-l} - M^{l-s+1}\mu_{\mathbb{D}}|_a||_2 \\
\leqslant\ & T_1 + T_2,
\end{aligned}
$$

Then, we hand the the first via spatial dependency and the second term via temporal dependency.

Let $\mathbb{A} = (\sum_{k\geqslant 0} c_1^k|W|_a^k)(c_2|W|_a + c_3 I)$, and notice that $|\Pi_l|_a + |\Pi^s_{l,i,t}|_a \leqslant 2\mathbb{A}^l$, where $|.|_a$ is element wise absolute value. Because of the row normalization, we have $(\mathsf{E}\,|T_2|^q)^{1/q} \lesssim |\mu_{q,\mathbb{D}}|_\infty (c_{23}/(1-c_1))^{s+1}$, where $|\mu_{q,\mathbb{D}}|_\infty$ is the maximum element of $\max_i(\mathsf{E}\,||\mathbb{D}_{t-l} - \mu_{\mathbb{D}}|_{a,i}|^q)^{1/q}$. Define $\mathcal{B}_{it}(s) = \{(i',t') : |i'-i| \leqslant s, |t'-t| \leqslant s\}$, then we have:

$$
T_1 \leqslant ||\sum_{j\notin\mathcal{B}_{it}(s)} e_i^\top \sum_{l\leqslant s}\mathbb{A}^l_{.j}|(\mathbb{D}_{t-l} - \mathbb{D}^s_{t-l,i,t})|_j||_2, \tag{28}
$$

as the term inside $\mathcal{B}_{it}(s)$ cancels for $(\mathbb{D}_{t-l} - \mathbb{D}^s_{t-l,i,t})$. Note that $||e_i^\top\mathbb{A}^{l-1}||_2 \leqslant (c_{23})^{l-1}/(1-$

$c_1)^{l-1}$ by Assumption 2.1. For $l > 1$, we have

$$\begin{aligned}
&\sum_{j\notin\mathcal{B}_{it}(s)}\sum_{l\leqslant s}||e_i^\top\mathbb{A}^l(\mathbb{D}_{t-l}-\mathbb{D}^s_{t-l,i,t})||_2 \leqslant \sum_{j\notin\mathcal{B}_{it}(s)}\sum_{l\leqslant s}||(|e_i^\top\mathbb{A}^{l-1}|_1|\mathbb{A}(\mathbb{D}_{t-l}-\mathbb{D}^s_{t-l,i,t})|_\infty)||_2\\
&\leqslant \sum_{l\leqslant s}(c_2+c_3)^{(l-1)}/(1-c_1)^{(l-1)}(\{|\mathbb{D}_{t-l}|_\infty \vee |\mu_{\mathbb{D}}|_\infty\}\sum_{j\notin\mathcal{B}_{it}(s)} c_{23}g_{ij})\\
&\leqslant \sum_{l\leqslant s}(c_{23})^l/(1-c_1)^{(l-1)}(1-c)^{-1}\{(d_f+d_z)\vee|\mu_{\mathbb{D}}|_\infty\}(\sum_{j\notin\mathcal{B}_{it}(s)} g_{ij})\\
&\leqslant C(\sum_{j\notin\mathcal{B}_{it}(s)} g_{ij}),
\end{aligned}$$

where C is a constant value and $g_{ij} = |(I-|c_1W|)^{-1}_{ij}|$. We obtain the last step using Assumption 2.1 A.1, A.2 and A.3.

Hence, under the condition $\sup_i \sum_{j\notin\mathcal{B}_{it}(s)} g_{ij} \to 0$ as $s\to\infty$, then $T_1 \to 0$. It is straightforward to prove it from the following results: either $T_1 \lesssim c^{-s}|\mu_{q,\mathbb{D}}|_\infty$ or $T_1 \lesssim s^{-c_w-2}|\mu_{q,\mathbb{D}}|_\infty$ depending on the assumption. Therefore,

$$\|Y_{it}-\mathsf{E}(Y_{it}|\mathscr{F}_{it}(s))\|_2 \lesssim \psi(s), \tag{29}$$

where $\psi(s)\to 0$ as $s\to 0$. Next, we discuss condition $\sup_i \sum_{j\notin\mathcal{B}_{it}(s)} g_{ij} \to 0, s\to\infty$, under Assumption 3.2 (i) and (ii), respectively.

(i) Under Assumption 3.2(1), we decompose matrix W using the properties of nilpotent matrix. For any positive integer h, we construct two $N\times N$ matrices A and B as follows: $a_{ij} = w_{ij}\mathbf{I}\{\rho(i,j) < N-h+1\}$, $b_{ij} = w_{ij}\mathbf{I}\{\rho(i,j)\geqslant N-h+1\}$. Then $W = A+B$ and $a_{ij}b_{ij} = 0$. We then check whether B is a nilpotent matrix, i.e., $B^h = 0$. Under assumption 3.2(1), $|w_{ij}| \leqslant \pi_0\rho(i,j)^{-c_w}$, and by Lemma A.3, we have:

$$\begin{aligned}
(|W|^h)_{ij} =&(W^h-B^h)_{ij} \leqslant |A|_{max}\sum_{k=0}^{h-1}\|B\|_\infty^k\|W^{h-1-k}\|_1 \quad \text{by Lemma A.3}\\
&\leqslant \pi_0\rho(i,j)^{-c_w}\sum_{k=0}^{h-1}\|W\|_\infty^k(h-k-1)K\Gamma_u \quad \text{by Assumption 3.2(1); Lemma A.7(2)}\\
&\leqslant \pi_0\rho(i,j)^{-c_w}\sum_{k=0}^{h-1}(h-k-1)K\Gamma_u\\
&\leqslant \pi_0\rho(i,j)^{-c_w}K\Gamma_u h^2.
\end{aligned}$$

Hence, for any $i \neq j$, using $\Upsilon = \sup_{\tau} |\gamma_1(\tau)|$ (Assumption 2.1 (A.1)), then

$$\begin{aligned} g_{ij} =& |(I - |\gamma_1 W|)^{-1}_{ij}| = \sum_{h=0}^{\infty} |\gamma_1 W|^h_{ij} = \sum_{h=0}^{\infty} |\gamma_1|^h |W|^h_{ij} \\ \leqslant& \sum_{h=0}^{\infty} \Upsilon^h \pi_0 \rho(i,j)^{-c_w} K \Gamma_u h^2 \quad \text{by Assumption 2.1 (A.1)} \\ =& \pi_0 K \Upsilon^2 \Gamma_u \rho(i,j)^{-c_w} \sum_{h=0}^{\infty} \Upsilon^{h-2} h^2 \\ \leqslant& \pi_2 \rho(i,j)^{-c_w}, \quad \text{for some constant } \pi_2 > 0. \end{aligned}$$

For sufficiently large s, we have:

$$\begin{aligned} \sup_i \sum_{j:\rho(i,j)>s} g_{ij} \leqslant& \sup_i \sum_{h=[s]}^{\infty} \sum_{j:h \leqslant \rho(i,j) < h+1} \pi_2 \rho(i,j)^{-c_w} \\ \leqslant& \sum_{h=[s]}^{\infty} \pi_1 h^{d-1} \pi_2 h^{-c_w} = \sum_{h=[s]}^{\infty} \pi_1 \pi_2 h^{-(c_w-d)-1} \quad \text{by Lemma A.2} \\ \leqslant& \pi_1 \pi_2 \frac{2^{c_w-d+1}}{c_w - d} s^{-(c_w-d)} \quad \text{by Lemma A.4} \\ \leqslant& \pi s^{-(c_w-d)}, \quad \text{for some constant } \pi > 0. \end{aligned}$$

Under Assumption 3.2(1), $c_w > d$, as $s \to \infty$, $\sup_i \sum_{j:\rho(i,j)>s} g_{ij} \leqslant \pi s^{-(c_w-d)} \to 0$.

(ii) Next, under Assumption 3.2(2), $\sup_i \sum_{j:\rho(i,j)>s} g_{ij} = \sup_i \sum_{j:\rho(i,j)>s} \sum_{h=0}^{\infty} |\gamma_1|^h |W|^h_{ij}$
$\leqslant \sup_i \sum_{j:\rho(i,j)>s} \sum_{h=[s/\bar{\rho}_0]+1} |\gamma_1|^h |W|^h_{ij} = \sup_i \sum_{h=[s/\bar{\rho}_0]+1} \sum_{j:\rho(i,j)>s} |\gamma_1|^h |W|^h_{ij}$
$\leqslant \sup_i \sum_{h=[s/\bar{\rho}_0]+1} \Upsilon^h \leqslant (1-\Upsilon)^{-1} \Upsilon^{s/\bar{\rho}_0}$.

Under Assumption 2.1 (A.1), $\Upsilon < 1$. Hence, as $s \to \infty$, we have: $\sup_i \sum_{j:\rho(i,j)>s} g_{ij} \leqslant (1-\Upsilon)^{-1} \Upsilon^{s/\bar{\rho}_0} \to 0$.

(2) Next, we discuss the NED properties of $\{u_{it}\}_{i,t}$. Denote $\bar{y}_{it} \overset{\text{def}}{=} e_i^\top W \mathbb{Y}_t$. We first prove that $\{\bar{y}_{it}\}_{it}$ is NED. Note that $\|\bar{y}_{it} - \mathsf{E}(\bar{y}_{it}|\mathscr{F}_{it}(s))\|_2 \leqslant \sum_{j=0}^{N} |W_{ij}| \|Y_{j,t} - \mathsf{E}(Y_{j,t}|\mathscr{F}_{it}(s))\|_2$. Under Assumption 3.2(1), using the result $\|Y_{it} - \mathsf{E}(Y_{it}|\mathscr{F}_{it}(s))\|_2 < Cs^{-(c_w-2)}$ in Proposition 3.1(1), we have: $\|\bar{Y}_{it} - \mathsf{E}(\bar{Y}_{it}|\mathscr{F}_{it}(s))\|_2 \leqslant \sum_{j=0}^{N} |W_{ij}| C s^{-(c_w-d)} \leqslant \|W\|_\infty C s^{d-c_w} \leqslant \pi_3 s^{d-c_w}$ for some positive constant π_3. Hence, $\{\bar{Y}_{it}\}$ is NED process. Similar results can

be obtained under Assumption 3.2(2) and using Proposition 3.1(2). By Lemma A.5, it is easily seen that $\{u_{it}(\gamma_1,\phi,\lambda,\tau)\}_{i,t}$ follow the same NED process.

We now prove that this NED property can be transformed. Let $\tilde{u}_{it}$ be a middle point between u_{it} and 0, then for sufficient small ε, we have:

$$\begin{aligned}
&\|\psi_\tau(u_{it}) - \mathsf{E}[\psi_\tau(u_{it})|\mathscr{F}_{it}(s)]\|_2 = \|\mathbf{I}(u_{it}\geqslant 0) - \mathsf{E}[\mathbf{I}(u_{it}\geqslant 0)|\mathscr{F}_{it}(s)]\|_2 \\
\leqslant\ & \|\mathbf{I}(u_{it}\geqslant 0) - \mathbf{I}\{\mathsf{E}[u_{it}|\mathscr{F}_{it}(s)]\geqslant 0\}\|_2 = \left\{\mathsf{E}\left|\mathbf{I}(u_{it}\geqslant 0) - \mathbf{I}\{\mathsf{E}[u_{it}|\mathscr{F}_{it}(s)]\geqslant 0\}\right|^2\right\}^{\frac{1}{2}} \\
\leqslant\ & \mathrm{P}(u_{it}\geqslant 0, \mathsf{E}[u_{it}|\mathscr{F}_{it}(s)]<0) \\
\leqslant\ & \mathrm{P}(u_{it}\geqslant \varepsilon, \mathsf{E}[u_{it}|\mathscr{F}_{it}(s)]<0) + \mathrm{P}(0<u_{it}<\varepsilon) \\
\leqslant\ & \mathrm{P}(|u_{it} - \mathsf{E}[u_{it}|\mathscr{F}_{it}(s)]|>\varepsilon) + \varepsilon f(\tilde{u}_{it}) \\
\leqslant\ & \mathsf{E}(|u_{it} - \mathsf{E}[u_{it}|\mathscr{F}_{it}(s)]|^2)/\varepsilon^2 + \varepsilon f(\tilde{u}_{it}), \quad \tilde{u}_{it} \quad \text{is a point between} \quad 0 \quad \text{and} \quad \tilde{u}_{it}. \\
\leqslant\ & \psi(s)/\varepsilon^2 + \varepsilon c_u.
\end{aligned}$$

Taking $\varepsilon = \psi(s)^{1/3}$ to be sufficiently small, then we achieve the desired result. Hence, conditioning on $\mathcal{C}$, the transformations $\{\psi_\tau(u_{it})\}_{i,t}$ and $\{\rho_\tau(u_{it})\}_{i,t}$ are also L_2-NED on $\{X_{it}, U_{it}\}_{i,t}$. □

7.3.3 Proof of Theorem 2 and Theorem 3

For convenience we collect some important notations. Define $\eta(\tau)\equiv(\phi(\tau)',\lambda(\tau)')'$, $\pi(\tau)\equiv(\gamma_1(\tau),\phi(\tau),\lambda(\tau)) = (\gamma_1(\tau),\eta(\tau))$. For simplicity, we denote $\eta\equiv(\phi',\lambda')'$, $\pi\equiv(\gamma_1,\phi,\lambda) = (\gamma_1,\eta)$, $\theta\equiv(\gamma_1',\phi')'$. Recall that the true parameter $\eta^0(\tau) = (\phi^0(\tau)',0')'$, and $\rho_\tau(u) = (\tau - 1(u\leqslant 0))u$. Define

$$\begin{aligned}
Q_{NT}(\gamma_1,\eta,\tau) &\equiv \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left[\rho_\tau\left\{Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - X_{it}^\top\phi(\tau) - \Phi_{it}^\top\lambda(\tau)\right\}\right], && (30)\\
Q_\infty(\gamma_1,\eta,\tau) &\equiv \lim_{N\to\infty,T\to\infty}\mathsf{E}[Q_{NT}(\gamma_1,\eta,\tau)], && (31)\\
\widehat{\eta}(\gamma_1,\tau) &\equiv (\widehat{\phi}(\gamma_1,\tau)',\widehat{\lambda}(\gamma_1,\tau)')' \equiv \arg\min_{(\phi,\lambda)\in\mathcal{B}\times\mathcal{G}} Q_{NT}(\gamma_1,\eta,\tau), && (32)\\
\eta^0(\gamma_1,\tau) &\equiv (\phi(\gamma_1,\tau)',\lambda(\gamma_1,\tau)')' = \arg\min_{(\phi,\lambda)\in\mathcal{B}\times\mathcal{G}} Q_\infty(\gamma_1,\eta,\tau). && (33)
\end{aligned}$$

$$
s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) = \psi_\tau \left\{ Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - \Psi_{it}^\top \eta(\gamma_1, \tau) \right\} \Psi_{it}, \quad \Psi_{it} = (X_{it}^\top, \Phi_{it}^\top)^\top, \tag{34}
$$

$$
\check{s}_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) = s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) - \mathsf{E}\, s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau), \tag{35}
$$

$$
\begin{aligned}
G_{NT} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \check{s}_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau), \qquad (36)\\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} [s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) - \mathsf{E}\, s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau)],\\
G_{NT}^0 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \check{s}_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau), \qquad (37)\\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} [s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau)].
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{\gamma}_1(\tau) &\equiv \arg\min_{\gamma_1 \in \mathcal{A}} ||\widehat{\lambda}(\gamma_1, \tau)||, \quad \gamma_1^*(\tau) \equiv \arg\min_{\gamma_1 \in \mathcal{A}} ||\lambda(\gamma_1, \tau)||, \qquad (38)\\
\widehat{\eta}(\tau) &\equiv (\widehat{\phi}(\tau)', \widehat{\lambda}(\tau)')' \equiv \widehat{\eta}(\widehat{\gamma}_1(\tau), \tau), \qquad (39)\\
\eta^0(\tau) &\equiv (\phi^0(\tau)', 0')' \equiv \eta(\gamma_1^0(\tau), \tau). \qquad (40)
\end{aligned}
$$

Proof of Theorem 2

Let τ be fixed. We mainly follow Chernozhukov and Hansen (2006) and prove the theorem in three steps:

Step (i) [Identification]: By Assumption 3.6 (R3), $\theta^0(\tau) = (\gamma_1^0(\tau), \phi^0(\tau))$ is the unique solution to $S_\infty(\theta, \tau) = 0$, which implies that it uniquely solves the equation

$$
\lim_{N\to\infty, T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathsf{E}\left[\psi_\tau \left\{ Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - X_{it}^\top \phi(\tau) - \Phi_{it}^\top 0 \right\} \Psi_{it} \right]. \tag{41}
$$

In view of the global convexity of $Q_\infty(\gamma_1, \eta, \tau)$ in η for each γ_1 and τ, there is a fact that if $\eta^0(\gamma_1, \tau) = (\phi^0(\gamma_1, \tau), \lambda^0(\gamma_1, \tau))$ is in the interior of $\mathcal{B} \times \mathbb{G}$, then $\eta^0(\gamma_1, \tau)$ uniquely solves the first order condition of minimizing $Q_\infty(\gamma_1, \eta, \tau)$ over η:

$$
\lim_{N\to\infty, T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathsf{E}\left[\psi_\tau \left\{ Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - X_{it}^\top \phi(\gamma_1, \tau) - \Phi_{it}^\top \lambda(\gamma_1, \tau) \right\} \Psi_{it} \right] = 0. \tag{42}
$$

We need to find $\gamma_1^*(\tau)$ by minimizing $||\lambda(\gamma_1, \tau)||$ over γ_1 subject to the constraint in

(42). By (41), it is clear that $\gamma_1^*(\tau) = \gamma_1^0(\tau)$ makes $||\lambda(\gamma_1(\tau), \tau)|| = 0$ and $\gamma_1^0(\tau)$ satisfies (42). That is, $\gamma_1^*(\tau) = \gamma_1^0(\tau) \in \arg\min_{\gamma_1 \in \mathcal{A}} ||\lambda(\gamma_1, \tau)||$ subject to the constraint in (42). It is also the unique solution by (41). Hence, $\phi(\gamma_1^*(\tau), \tau) = \phi(\gamma_1^0(\tau), \tau) = \phi^0(\tau)$ by (42).

Step (ii) [Consistency]:

In Proposition 3.1, we established that the process $\{\rho_\tau(u_{it})\}_{i,t}$ is L_2-NED on $\{X_{it}, U_{it}\}_{i,t}$. According to Theorem 1 in Jenish and Prucha (2012), under Assumption 3.4, we have the uniform consistency $\sup_{\gamma_1,\phi,\lambda,\tau} |Q_{NT}(\gamma_1, \phi, \lambda, \tau) - \mathsf{E}\, Q_{NT}(\gamma_1, \phi, \lambda, \tau)| = \mathcal{o}_p(1)$.

By the bounded density condition in 3.6 (R2), $Q_\infty(\gamma_1, \eta, \tau)$ is continuous over $\mathcal{A} \times (\mathcal{B} \times \mathcal{G}) \times \mathcal{T}$. And by Lemma B2, $\sup_{(\gamma_1,\eta,\tau)\in\mathcal{A}\times(\mathcal{B}\times\mathcal{G})\times\mathcal{T}} ||Q_{NT}(\gamma_1, \eta, \tau) - Q_\infty(\gamma_1, \eta, \tau)|| = \mathcal{o}_p(1)$. This implies the uniform convergence by Lemma B.1 in Chernozhukov and Hansen (2006) $\sup_{(\gamma_1,\tau)\in\mathcal{A}\times\mathcal{T}} ||\widehat{\eta}(\gamma_1, \tau) - \eta^0(\gamma_1, \tau)|| = o_p(1)$ (*). It follows that $\sup_{(\gamma_1,\tau)\in\mathcal{A}\times\mathcal{T}} ||\widehat{\lambda}(\gamma_1, \tau) - \lambda^0(\gamma_1, \tau)|| = o_p(1)$, which by Lemma B.1 in Chernozhukov and Hansen (2006) again implies $\sup_{\tau\in\mathcal{T}} ||\widehat{\gamma}_1(\tau) - \gamma_1^0(\tau)|| = o_p(1)$, which by (*) implies $\sup_{\tau\in\mathcal{T}} ||\widehat{\phi}(\tau) - \phi^0(\tau)|| = o_p(1)$ and $\sup_{\tau\in\mathcal{T}} ||\widehat{\lambda}(\widehat{\gamma}_1(\tau), \tau) - 0|| = o_p(1)$.

Step (iii)[Asymptotics]: Consider a small ball $B_{\epsilon_{NT}}(\gamma_1^0(\tau))$ of radius ϵ_{NT} centered at $\gamma_1^0 \equiv \gamma_1^0(\tau)$ for each τ, while balls' radius ϵ_{NT} is independent of τ and $\epsilon_{NT} \to 0$ slowly enough. Let any value $\tilde{\gamma}_1 \equiv \tilde{\gamma}_1(\tau) \in B_{\epsilon_{NT}}(\gamma_1^0(\tau))$. By the computational properties of the ordinary quantile regression estimator $\widehat{\eta}(\tilde{\gamma}_1, \tau)$, Theorem 3.3 in Koenker and Bassett (1978),

$$O(n^{-1/2}) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \psi_\tau \left\{ Y_{it} - \tilde{\gamma}_1(\tau)\overline{Y}_{it} - \Psi_{it}^\top \widehat{\eta}(\tilde{\gamma}_1, \tau) \right\} \Psi_{it}. \tag{43}$$

Let $s_{it}(\tilde{\gamma}_1, \widehat{\eta}(\tilde{\gamma}_1, \tau), \tau) = \psi_\tau \left\{ Y_{it} - \tilde{\gamma}_1(\tau)\overline{Y}_{it} - \Psi_{it}^\top \widehat{\eta}(\tilde{\gamma}_1, \tau) \right\} \Psi_{it}$,
$G_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} [s_{it}(\tilde{\gamma}_1, \widehat{\eta}(\tilde{\gamma}_1, \tau), \tau) - \mathsf{E}\, s_{it}(\tilde{\gamma}_1, \widehat{\eta}(\tilde{\gamma}_1, \tau), \tau)]$, and
$G_{NT}^0 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau)$. Using Lemma B.1, which implies that for any $\sup_{\tau\in\mathcal{T}} ||\tilde{\gamma}_1(\tau) - \gamma_1^0(\tau)|| = o_p(1)$, it is the case that

$\sup_{\tau\in\mathcal{T}}||G_{NT}-G^0_{NT}||=o_p(1)$. Then the above equation (43) can be transformed as,

$$\begin{aligned}O(n^{-1/2})=&\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T s_{it}(\tilde\gamma_1,\widehat\eta(\tilde\gamma_1,\tau),\tau),\\=&\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T[s_{it}(\tilde\gamma_1,\widehat\eta(\tilde\gamma_1,\tau),\tau)-\mathsf{E}\,s_{it}(\tilde\gamma_1,\widehat\eta(\tilde\gamma_1,\tau),\tau)] \qquad (44)\\&+\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\mathsf{E}\,s_{it}(\tilde\gamma_1,\widehat\eta(\tilde\gamma_1,\tau),\tau),\\=&G^0_{NT}+o_p(1)+\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\mathsf{E}\,s_{it}(\tilde\gamma_1,\widehat\eta(\tilde\gamma_1,\tau)). \qquad (45)\end{aligned}$$

By mean value theorem and dominated convergence arguments, we have

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\mathsf{E}\,s_{it}(\tilde\gamma_1,\widehat\eta(\tilde\gamma_1,\tau))\\=&\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\mathsf{E}\,\psi_\tau\left\{Y_{it}-\tilde\gamma_1(\tau)\overline{Y}_{it}-\Psi_{it}^\top\widehat\eta(\tilde\gamma_1,\tau)\right\}\Psi_{it}\\=&(J_{\gamma_1}(\tau)+o_p(1))\sqrt{NT}(\tilde\gamma_1(\tau)-\gamma_1^0(\tau))+(J_\eta(\tau)+o_p(1))\sqrt{NT}(\widehat\eta(\tilde\gamma_1,\tau)-\eta^0(\tau)), \qquad (46)\end{aligned}$$

where

$$\begin{aligned}J_{\gamma_1}(\tau)&=\frac{\partial S_\infty(\theta,\tau)}{\partial\gamma_1}\Big|_{\gamma_1=\gamma_1^0}, &(47)\\J_\eta(\tau)&=\frac{\partial S_\infty(\pi,\tau)}{\partial(\phi,\lambda)}\Big|_{\phi=\phi^0,\lambda=0}, &(48)\\J(\tau)&=\frac{\partial S_\infty(\pi,\tau)}{\partial(\gamma_1,\phi)}\Big|_{\gamma_1=\gamma_1^0,\phi=\phi^0,\lambda=0}, &(49)\end{aligned}$$

with dimensions $(q+4+(p+1)m)\times1$, $(q+4+(p+1)m)\times(q+4+(p+1)m)$ and $(q+4+(p+1)m)\times(q+4+(p+1)m)$. Putting (45) and (46) together, we have

$$\begin{aligned}O(n^{-1/2})=&G^0_{NT}+o_p(1)+(J_{\gamma_1}(\tau)+o_p(1))\sqrt{NT}(\tilde\gamma_1(\tau)-\gamma_1^0(\tau)) \qquad (50)\\&+(J_\eta(\tau)+o_p(1))\sqrt{NT}(\widehat\eta(\tilde\gamma_1,\tau)-\eta^0(\tau)),\end{aligned}$$

which implies that for any $\sup_{\tau\in\mathcal{T}}||\tilde\gamma_1(\tau)-\gamma_1^0(\tau)||=o_p(1)$,

$$\sqrt{NT}(\widehat\eta(\tilde\gamma_1,\tau)-\eta^0(\tau))=-J_\eta^{-1}(\tau)G^0_{NT}-J_\eta^{-1}(\tau)J_{\gamma_1}(\tau)[1+o_p(1)]\sqrt{NT}(\tilde\gamma_1(\tau)-\gamma_1^0(\tau))+o_p(1). \qquad (51)$$

Then

$$\sqrt{NT}(\widehat{\phi}(\tilde{\gamma}_1,\tau)-\phi^0(\tau)) = \bar{J}_\phi(\tau)G^0_{NT} - \bar{J}_\phi(\tau)J_{\gamma_1}(\tau)[1+o_p(1)]\sqrt{NT}(\tilde{\gamma}_1(\tau)-\gamma^0_1(\tau)) + o_p(1), \tag{52}$$

$$\sqrt{NT}(\widehat{\lambda}(\tilde{\gamma}_1,\tau)-0) = \bar{J}_\lambda(\tau)G^0_{NT} - \bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)[1+o_p(1)]\sqrt{NT}(\tilde{\gamma}_1(\tau)-\gamma^0_1(\tau)) + o_p(1), \tag{53}$$

where partition conformably $J_\eta^{-1}(\tau) = [\bar{J}'_\phi(\tau), \bar{J}'_\lambda(\tau)]'$.

By Step (ii), with probability approaching one,

$$\widehat{\gamma}_1(\tau) = \arg\min_{\tilde{\gamma}_1 \in B_{\epsilon_{NT}}(\gamma^0_1)} ||\widehat{\lambda}(\tilde{\gamma}_1,\tau)||, \text{ for all } \tau \in \mathcal{T}. \tag{54}$$

As we discussed in part 3.2.1, the process $\{\check{s}_{it}\}_{i,t}$ is NED process, where $\check{s}_{it} \equiv \check{s}_{it}(\gamma^0_1,\eta^0(\gamma^0_1,\tau),\tau) = s_{it}(\gamma^0_1,\eta^0(\gamma^0_1,\tau),\tau) - \mathsf{E}\, s_{it}(\gamma^0_1,\eta^0(\gamma^0_1,\tau),\tau)$. By Theorem 2 in Jenish and Prucha (2012), under Assumption 3.3(ii) and 3.5, we have $G^0_{NT} \xrightarrow{d} N(0,\Omega^0)$, where $\Omega^0 = \tau(1-\tau)\lim_{N\to\infty,T\to\infty} \mathsf{E}(\Psi_{it}\Psi'_{it})$. Hence G^0_{NT} is $O_p(1)$, then we have

$$\sqrt{NT}||\widehat{\lambda}(\tilde{\gamma}_1,\tau)|| = ||O_p(1) - \bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)[1+o_p(1)]\sqrt{NT}(\tilde{\gamma}_1(\tau)-\gamma^0_1(\tau))||. \tag{55}$$

It follows from (54) and (55) that $\sqrt{NT}(\widehat{\gamma}_1(\tau)-\gamma^0_1(\tau)) = O_p(1)$ by the full rank properties of $\bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)$. Consequently, following Lemma B.1 in Chernozhukov and Hansen (2006), combing (53) and (55),

$$\begin{aligned}\sqrt{NT}(\widehat{\gamma}_1(\tau)-\gamma^0_1(\tau)) &= \arg\min_{s\in\mathbb{R}} ||-\bar{J}_\lambda(\tau)G^0_{NT} - \bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)s|| + o_p(1)\\ &= [J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)]^{-1}[J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)]G^0_{NT} + o_p(1). \end{aligned} \tag{56}$$

Plugging into (51), simple algebra shows that

$$\begin{aligned}&\sqrt{NT}(\widehat{\eta}(\widehat{\gamma}_1(\tau),\tau)-\eta^0(\tau)) \\ &= J_\eta^{-1}(\tau)\left[I - J_{\gamma_1}(\tau)[J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)]^{-1}J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)\right]G^0_{NT} + o_p(1).\end{aligned} \tag{57}$$

Due to invertibility of $J_{\gamma_1}(\tau)\bar{J}_\lambda(\tau)$, we can have,

$$\begin{aligned}\sqrt{NT}(\widehat{\lambda}(\widehat{\gamma}_1(\tau),\tau)-0)=&-\bar{J}_\lambda(\tau)\left[I-J_{\gamma_1}(\tau)[J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)]^{-1}\bar{J}_\lambda(\tau)\right]G^0_{NT}+o_p(1)\\=&0\times O_p(1)+o_p(1).\end{aligned}$$

Using the fact that $(\tilde{\gamma}_1(\tau),\widehat{\eta}(\tilde{\gamma}_1(\tau),\tau))=(\widehat{\gamma}_1(\tau),\widehat{\eta}(\tau))=(\widehat{\gamma}_1(\tau),\widehat{\phi}(\tau),0+o_p(\frac{1}{\sqrt{(NT)}}))$, as $min\{N,T\}\to\infty$, we have

$$\sqrt{NT}\left\{\widehat{\theta}(\tau)-\theta^0(\tau)\right\}=-J^{-1}(\tau)G^0_{NT}(\theta^0,\tau)+o_p(1). \tag{58}$$

Recall $\Omega_0=\tau(1-\tau)\,\mathsf{E}(\Psi_{it}\Psi_{it}^\top)$. As mentioned before, using the properties of NED process $\{\check{s}_{it}\}_{i,t}$, where
$\check{s}_{it}\equiv\check{s}_{it}(\gamma_1^0,\eta^0(\gamma_1^0,\tau),\tau)=s_{it}(\gamma_1^0,\eta^0(\gamma_1^0,\tau),\tau)-\mathsf{E}\,s_{it}(\gamma_1^0,\eta^0(\gamma_1^0,\tau),\tau)$, by Theorem 2 in Jenish and Prucha (2012), under Assumption 3.3(ii) and 3.5, conditioning on $\mathcal{C}$, we have $\Omega_{0F}^{-1/2}G^0_{NT}\xrightarrow{d}N(0,I)$.

$$\begin{aligned}&\begin{pmatrix}\sqrt{NT}(\widehat{\gamma}_1(\tau)-\gamma_1^0(\tau))\\\sqrt{NT}(\widehat{\phi}(\tau)-\phi^0(\tau))\end{pmatrix}\\&=\begin{pmatrix}[J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)]^{-1}[J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)]\\\bar{J}_\phi(\tau)\left[I-J_{\gamma_1}(\tau)[J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)J_{\gamma_1}(\tau)]^{-1}J'_{\gamma_1}(\tau)\bar{J}'_\lambda(\tau)\bar{J}_\lambda(\tau)\right]\end{pmatrix}G^0_{NT}+o_p(1).\end{aligned} \tag{59}$$

Then conditioning on $\mathcal{C}$, we have $G^0_{NT}\xrightarrow{d}N(0,\Omega_0)$. Recall the conditional version of $J(\tau)$ as $J(\tau)^*$. Recall that the unconditional version of Ω_0 is defined by $\Omega_0^*=\tau(1-\tau)\,\mathsf{E}(\Psi_{it}\Psi_{it}^\top)$(not conditioning on $\mathcal{C}$). As we assume $\Omega_0^{-1}\Omega_0^*\to_p I$ and $J(\tau)^{-1}J(\tau)^*\to_p I$ where I is an identity matrix. Then the conclusion follows.

■

7.3.4 Lemma B.1

For convenience we collect some important notations. Note that the parameter set $\pi(\tau) \equiv (\gamma_1(\tau), \phi(\tau), \lambda(\tau)) = (\theta(\tau), \lambda(\tau)) = (\gamma_1(\tau), \eta(\tau))$, with $\theta(\tau) \equiv (\gamma_1(\tau), \phi(\tau))$ and $\eta(\tau) \equiv (\phi(\tau), \lambda(\tau))$. For simplicity, we denote $\pi \equiv (\gamma_1, \phi, \lambda) = (\theta, \lambda) = (\gamma_1, \eta)$, with $\theta \equiv (\gamma_1', \phi')'$ and $\eta \equiv (\phi', \lambda')'$. The true parameter set $\pi^0 \equiv (\gamma_1^0, \phi^0, \lambda^0) = (\theta^0, 0)$, with $\theta^0 \equiv \theta^0(\tau) \equiv (\gamma_1^0(\tau), \phi^0(\tau))$ and $\lambda^0 = 0$. Recall that

$$
\begin{aligned}
u_{it} =& Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - \Psi_{it}^\top \eta(\gamma_1, \tau), \quad \Psi_{it} = (X_{it}^\top, \Phi_{it}^\top)^\top, \\
u_{it}^* =& Y_{it} - \gamma_1^0(\tau)\overline{Y}_{it} - \Psi_{it}^\top \eta^0(\gamma_1^0, \tau), \\
s_{it}(\theta, \lambda, \tau) =& s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) = \psi_\tau \left\{ Y_{it} - \gamma_1(\tau)\overline{Y}_{it} - \Psi_{it}^\top \eta(\gamma_1, \tau) \right\} \Psi_{it}, \\
\check{s}_{it}(\theta, \lambda, \tau) =& \check{s}_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) = s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) - \mathsf{E}\, s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau), \\
G_{NT} =& \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \check{s}_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau), \\
=& \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau) - \mathsf{E}\, s_{it}(\gamma_1, \eta(\gamma_1, \tau), \tau)], \\
G_{NT}^0 =& \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \check{s}_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau), \\
=& \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau) - \mathsf{E}\, s_{it}(\gamma_1^0, \eta^0(\gamma_1^0, \tau), \tau)].
\end{aligned}
$$

We need to prove that for any $\sup_{\tau \in \mathcal{T}} ||\widehat{\pi}(\tau) - \pi^0(\tau)|| = o_p(1)$, it is the case that $\sup_{\tau \in \mathcal{T}} ||G_{NT} - G_{NT}^0|| = o_p(1)$. First for any estimator $\widehat{\pi}(\tau) = (\widehat{\theta}(\tau), \widehat{\lambda}(\tau)) = (\widehat{\gamma}_1(\tau), \widehat{\phi}(\tau), \widehat{\lambda}(\tau))$ which satisfying $|\widehat{\theta}(\tau) - \theta^0(\tau)|_a \leqslant \delta_1$ and $|\widehat{\lambda}(\tau) - 0|_a \leqslant \delta_2$ with a constant vector $\delta = (\delta_1^\top, \delta_2^\top)^\top$, we define

$$
\tilde{\psi}_{it}(\tau, \delta) = \check{s}_{it}(\widehat{\theta}, \widehat{\lambda}, \tau) - \check{s}_{it}(\theta^0, 0, \tau). \tag{60}
$$

Lemma B.1. *We denote c_1, c_2 as two constants, $B_\tau \in (0, 1)$ is a compact set*

$$
\sup_{\tau \in B_\tau} \sup_{|\delta_1|_1 \leqslant c_1/\sqrt{NT}} \sup_{|\delta_2|_1 \leqslant c_2/\sqrt{NT}} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \tilde{\psi}_{it}(\tau, \delta) \right\| = \mathcal{O}_p(1). \tag{61}
$$

Proof.

We denote $|.|_a$ as the element wise absolute value. Let $d = q+4+(p+1)m$, we define the function class, with m_1, m_2 as two constants,

$$\begin{aligned}
&\mathcal{V}(m_1, m_2, B_\tau) \\
\stackrel{\text{def}}{=} \; & \{(\theta, \tau, \lambda) \mapsto \{[\tau - \mathbf{I}(y - \gamma_1 \overline{y} - x^\top \phi - \Phi^\top \lambda \leqslant 0)]\Psi \\
& -\{[\tau - \mathbf{I}(y - \gamma_1^0 \overline{y} - x^\top \phi^0 - \Phi^\top \lambda^0 \leqslant 0)]\Psi\}, |\theta - \theta^0|_1 \leqslant m_1, |\lambda - \lambda^0|_1 \leqslant m_2, \tau \in B_\tau\}.
\end{aligned}$$

$\mathcal{V}$ is VC subgraph with index $v \geqslant d+2$ in view of Lemma 9.12 i) in Kosorok (2007) and A.7 in Andrews et al. (1993). $\mathcal{V}$ has the envelop function $V(.)$. With probability measure Q and L_2 norm $\|V\|_{Q,2} = (\int |V|_2^2 dQ)^{1/2}$. Then we assume that covering numbers of VC-classes of functions $\mathcal{N}(\varepsilon\|V\|_{Q,2}, \mathcal{V}, L_2(Q)) \lesssim (\varepsilon)^{-(v-1)}$ by Theorem 9.3 of Kosorok (2007).

Denote $T_{NT}(f) = G_{NT}(f) - G^0_{NT}(f) = (\sqrt{NT})^{-1} \sum_i \sum_t \tilde{\psi}_{it}(\tau, \delta)$, and $J_{NT}(f) = (NT)^{-1} \sum_i \sum_t (s_{it}(\theta^0 + \delta_1, \lambda + \delta_2, \tau) - s_{it}(\theta^0, 0, \tau))$. We define $\widetilde{\mathcal{V}} = \mathcal{V}(\delta_1 \sqrt{NT}^{-1}, \delta_2 \sqrt{NT}^{-1}, B_\tau)$, and the rate of the cover of the envelope for $\widetilde{\mathcal{V}}$ is $\|\tilde{V}\|_2 \lesssim \{NT\}^{-1/4}$.

Then we define the A_ε as the $\varepsilon\|\tilde{V}\|_{Q,2}$ cover of the functional class $\mathcal{V}$, where for each f in v we shall define the the closest element to it as $\pi(f)$ and $|\pi(f) - f|_{Q,2} \leqslant \varepsilon$. It is not hard to see that $|\mathcal{A}_\varepsilon| \lesssim (\varepsilon)^{-(v-1)}$. Also we denote $P_{NT}(f, g) = (NT)^{-1} \sum_{i,t} |f_{i,t} - g_{i,t}|$. We shall assume that for our choice of ε, $\mathcal{N}(\varepsilon\|V\|_{Q,2}, \mathcal{V}, L_{1,n}(Q)) \lesssim_p \mathcal{N}(\varepsilon\|V\|_{Q,2}, \mathcal{V}, L_2(Q))$. The one step chaining gives us,

$$\begin{aligned}
& \sup_{\tau \in B_\tau} \sup_{|\delta_1|_1 \leqslant c_1 1/\sqrt{NT}} \sup_{|\delta_2|_1 \leqslant c_2 1/\sqrt{NT}} |(\sqrt{NT})^{-1} \sum_i \sum_t \tilde{\psi}_{it}(\tau, \delta)| && (62) \\
& \leqslant \sup_{f \in \widetilde{\mathcal{V}}} |T_{NT}(f)| \\
& = \sqrt{NT} \sup_{f \in \widetilde{\mathcal{V}}} \Big| [J_{NT}(f) - J_{NT}\{\pi(f)\} - \mathsf{E}\, J_{NT}(f) + \mathsf{E}\, J_{NT}\{\pi(f)\}] && (63) \\
& \qquad + [J_{NT}\{\pi(f)\} - \mathsf{E}\, J_{NT}\{\pi(f)\}] \Big| \\
& \lesssim_p 2\sqrt{NT}\varepsilon(NT)^{-1/4} + \sqrt{NT} \max_{f \in A_\varepsilon} |J_{NT}(f) - \mathsf{E}\, J_{NT}(f)| \\
& = 2(NT)^{1/4}\varepsilon + K_{NT}, && (64)
\end{aligned}$$

where $(NT)^{-1/4}$ corresponds to the rate of the envelope.

We let Θ_ε corresponds to the discretized function set A_ε. Here, K_{NT} involves partial sums, which is handled via the NED property and the continuity of the function with respect to the parameter, see more details in Lemma B.2.

By Lemma B.2, we have the following rate,

$$\begin{aligned}
&\mathrm{P}(\sup_{\tau,\theta\in\Theta_\varepsilon} \frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{\psi}_{it}(\tau,\delta) \geqslant c) \\
&\leqslant \mathsf{E}\left\{ \sup_{\tau,\theta\in\Theta_\varepsilon} \frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{\psi}_{it}(\tau,\delta) \right\}^2 / c^2 \\
&\leqslant (NT)^{-1/4}/c^2.
\end{aligned}$$

We can pick for example $(NT)^{-1/8}/c = o(1)$, and then $K_{NT} = o(1)$. Also $\varepsilon/(NT)^{-1/4} = o(1)$. □

Lemma B.2. *Denote* $\eta_{it}(\tau,\delta) \stackrel{\text{def}}{=} \sup_{\tau,\theta\in\Theta_\varepsilon}\{\check{s}_{it}(\theta,\lambda,\tau) - \check{s}_{it}(\theta^0,0,\tau)\}$*, where* Θ_ε *corresponds to the discretized function set* A_ε*. For each* τ *and* $\|\delta\| \leqslant M < \infty$*, if* $c_w > (1+\frac{1}{1-q})d$*, we have*

$$\mathsf{Var}\left[\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \eta_{it}(\tau,\delta)\right] = \mathcal{O}(\{NT\}^{-\frac{1}{4}}).$$

Proof.

For simplicity, we denote $\sup_{\theta\in\Theta_\varepsilon}(\psi_\tau(u_{it}(\gamma_1,\phi,\lambda,\tau)) - \tilde{\psi}_\tau(u^*_{it}))\Psi_{it}$ as η_{it}. It is not hard to see that η_{it} is NED. For any $i \in 1,\cdots,N$, $t \in 1,\cdots,T$ and any $s > 0$, let

$$R^s_{it} = \mathsf{E}(\eta_{it}|\mathscr{F}_{it}(s)), \quad T^s_{it} = \eta_{it} - R^s_{it}.$$

By the Jenson and Lyapunov inequalities, we have for all $i \in 1,\cdots,N$, $t \in 1,\cdots,T$ and any $1 \leqslant q \leqslant 2+\delta$,

$$\mathsf{E}\,|R^s_{it}|^q = \mathsf{E}\{|\,\mathsf{E}(\eta_{it}|\mathscr{F}_{it}(s))|^q\} \leqslant \mathsf{E}\{\mathsf{E}(|\eta_{it}|^q|\mathscr{F}_{it}(s))\} = \mathsf{E}\,|\eta_{it}|^q,$$

and thus

$$\|R_{it}^s\|_q \leqslant \|\eta_{it}\|_q \leqslant \|\eta_{it}\|_{2+\delta}$$
$$\|T_{it}^s\|_q \leqslant \|\eta_{it}\|_q + \|R_{it}^s\|_q \leqslant 2\|\eta_{it}\|_q \leqslant 2\|\eta_{it}\|_{2+\delta}.$$

Therefore, both R_{it}^s and T_{it}^s are uniformly $L_{2+\delta}$ bounded. Since the process $\{\tilde{\psi}_\tau\}_{i,t}$ is uniformly L_2-NED on $\{X_{it}, U_{it}\}_{i,t}$ and w.l.g. the NED-scaling factors can be chosen as one, then

$$\sup_{i,t} \|T_{it}^s\|_2 \leqslant \varphi(s),$$

Furthermore, let $\sigma(R_{it}^s)$ denote the $\sigma-$field generated by R_{it}^s. Since $\sigma(R_{it}^s) \subseteq \mathscr{F}_{it}(s)$, the mixing coefficients of R_{it}^s satisfy

$$\overline{\alpha}_R(1,1,h) \leqslant \begin{cases} 1, & h \leqslant 2s, \\ \overline{\alpha}(Ms^d, Ms^d, h-2s), & h > 2s, \end{cases}$$

where $\overline{\alpha}(u,v,h)$ are the mixing coefficients of the input process $\{X_{it}, U_{it}\}$, since the $s-$neighborhood of any point on D contains at most Ms^2 points of D for some M that does not depend on s, see Lemma A.1 of Jenish and Prucha (2009).

We decompose η_{it} and $\eta_{i',t'}$ as

$$\eta_{it} = R_{it}^{[s/3]} + T_{it}^{[s/3]}, \quad \eta_{i',t'} = R_{i',t'}^{[s/3]} + T_{i',t'}^{[s/3]}.$$

where $s = \rho((i,t),(i',t'))$. Then,

$$\begin{aligned} |\operatorname{Cov}(\eta_{it}, \eta_{i',t'})| =& |\operatorname{Cov}(R_{it}^{[s/3]} + T_{it}^{[s/3]}, R_{i',t'}^{[s/3]} + T_{i',t'}^{[s/3]})| \\ \leqslant& |\operatorname{Cov}(R_{it}^{[s/3]}, R_{i',t'}^{[s/3]})| + |\operatorname{Cov}(R_{it}^{[s/3]}, T_{i',t'}^{[s/3]})| \\ &+ |\operatorname{Cov}(T_{it}^{[s/3]}, R_{i',t'}^{[s/3]})| + |\operatorname{Cov}(T_{it}^{[s/3]}, T_{i',t'}^{[s/3]})|. \end{aligned}$$

We then bound each item on the right side of the last inequality.

First, using Lemma A.6 with $p = 2+\delta, q = 2$, and $h = 2(2+\delta)/\delta$ yields the following bound on the first term:

$$
\begin{aligned}
|\operatorname{Cov}(R_{it}^{[s/3]}, R_{i',t'}^{[s/3]})| &\leqslant 4\|R_{it}^{[s/3]}\|_{2+\delta}\|R_{i',t'}^{[s/3]}\|_2 \overline{\alpha}_R^{\delta/(4+2\delta)}(1,1,[s/3]) \\
&\leqslant 4\|\eta_{it}\|_{2+\delta}\|\eta_{it}\|_2 \overline{\alpha}^{\delta/(4+2\delta)}(M[s/3]^d, M[s/3]^d, s-2[s/3]) \\
&\leqslant C_1\|\eta_{it}\|_{2+\delta}\|\eta_{it}\|_2[s/3]^{d\varsigma_0}\widehat{\alpha}^{\delta/(4+2\delta)}([s/3]),
\end{aligned}
$$

where $\varsigma_0 = \delta\varsigma/(4+2\delta)$.

Second, by the Cauchy-Schwartz inequality, the second and third terms are bounded by:

$$
|\operatorname{Cov}(R_{it}^{[s/3]}, T_{i',t'}^{[s/3]})| \leqslant 4\|R_{it}^{[s/3]}\|_2\|T_{i',t'}^{[s/3]}\|_2 \leqslant 4\|\eta_{it}\|_2\varphi([s/3]).
$$

Similarly, the fourth term can be bounded as:

$$
|\operatorname{Cov}(T_{it}^{[s/3]}, T_{i',t'}^{[s/3]})| \leqslant 4\|T_{it}^{[s/3]}\|_2\|T_{i',t'}^{[s/3]}\|_2 \leqslant 8\|\eta_{it}\|_2\varphi([s/3]).
$$

Collecting the above inequalities, we have

$$
|\operatorname{Cov}(R_{it}^{[s/3]}, R_{i',t'}^{[s/3]})| \leqslant \|\eta_{it}\|_2\left\{C_1\|\eta_{it}\|_{2+\delta}[s/3]^{d\tau_0}\widehat{\alpha}^{\delta/(4+2\delta)}([s/3]) + C_2\varphi([s/3])\right\}. \qquad (65)
$$

Using the above inequality as well as the bound and definition of random fields, we have

$$
\begin{aligned}
&\mathsf{Var}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{it}\right]\\
\leqslant&\frac{1}{NT}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\mathsf{Var}(\eta_{it})+\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{(i',t')\neq(i,t)}|\operatorname{Cov}(R_{it}^{[s/3]},R_{i',t'}^{[s/3]})|\right\}\\
\leqslant&4\|\eta_{it}\|_2^2+C_1\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{(i',t')\neq(i,t)}\|\eta_{it}\|_2\|\eta_{it}\|_{2+\delta}[s/3]^{d\varsigma_0}\widehat{\alpha}^{\delta/(4+2\delta)}([s/3])\\
&+C_2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{(i',t')\neq(i,t)}\|\eta_{it}\|_2\varphi([s/3])\\
\leqslant&4\|\eta_{it}\|_2^2+C_1\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{\infty}\sum_{(i',t'):h\leqslant\rho((i,t),(i',t'))/3<h+1}\\
&\|\eta_{it}\|_2\|\eta_{it}\|_{2+\delta}[\rho((i,t),(i',t'))/3]^{d\varsigma_0}\widehat{\alpha}^{\delta/(4+2\delta)}([\rho((i,t),(i',t'))/3])\\
&+C_2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{\infty}\sum_{(i',t'):h\leqslant\rho((i,t),(i',t'))/3<h+1}\|\eta_{it}\|_2\varphi([\rho((i,t),(i',t'))/3])\\
\leqslant&4\|\eta_{it}\|_2^2+C_3\|\eta_{it}\|_2\left\{\sum_{h=1}^{\infty}h^{d(\varsigma_0+1)-1}\widehat{\alpha}^{\delta/(4+2\delta)}(h)+\sum_{h=1}^{\infty}h^{d-1}\varphi(h)\right\},
\end{aligned}
$$

where the second inequality is by substituting equation (65), the third inequality is by using the properties of random field, and the last inequality is by Lemma A.2 and the $L_{2+\delta}$-bound property of $\{\eta_{it}\}$.

We discuss $\sum_{h=1}^{\infty}h^{d-1}\varphi(h)$ under the aforementioned two cases of the NED coefficients of $\{\eta_{it}\}_{i=1}^{n}$, $\varphi(s)$: (1) Under assumptions 3.1-3.2(1) and 3.3(ii), the NED coefficients $\varphi(s)=s^{-(1-q)(c_w-d)}$, then $\sum_{h=1}^{\infty}h^{d-1}\varphi(h)=\sum_{h=1}^{\infty}h^{d-(1-q)(c_w-d)-1}<\infty$, when $c_w>(1+\frac{1}{1-q})d$. (2) Under assumptions 3.1-3.2(2) and 3.3(ii), the NED coefficients $\varphi(s)=\Upsilon^{(1-q)s/\bar{\rho}_0}$, then $\sum_{h=1}^{\infty}h^{d-1}\varphi(h)=\sum_{h=1}^{\infty}h^{d-1}\Upsilon^{(1-q)h/\bar{\rho}_0}<\infty$, due to $\Upsilon<1$. Therefore, $\sum_{h=1}^{\infty}h^{d-1}\varphi(h)<\infty$.

Further, under assumption 3.3(i), $\sum_{h=1}^{\infty}h^{d(\varsigma_0+1)-1}\widehat{\alpha}^{\delta/(4+2\delta)}(h)<\infty$, combined with $L_{2+\delta}$-bound of $\{\eta_{it}\}$, we obtain that

$$
\mathsf{Var}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{it}\right]\leqslant C\max_{i,t}\|\eta_{it}\|_2
$$

for some $C<\infty$.

Next, we analyze $\|\eta_{it}\|_2$. Note that

$$
\begin{aligned}
\left|\mathsf{E}\left(\eta_{it}^\top \eta_{it}\right)\right| \leqslant & 2\left|\mathsf{E}\left\{\left[\mathbf{I}(-(NT)^{-\frac{1}{2}}\delta^\top|\xi_{it}|_a \leqslant u_{it}^* < (NT)^{-\frac{1}{2}}\delta^\top|\xi_{it}|_a)\right]\{\Psi_{it}^\top\Psi_{it} \vee \mathsf{E}|\Psi_{it}|_a^\top \mathsf{E}|\Psi_{it}|_a\}\right\}\right| \\
\leqslant & 2\left|\mathsf{E}\int_{-NT^{-\frac{1}{2}}\delta^\top|\xi_{it}|_a}^{NT^{-\frac{1}{2}}\delta^\top|\xi_{it}|_a}\{\Psi_{it}^\top\Psi_{it} \vee \mathsf{E}|\Psi_{it}|_a^\top \mathsf{E}|\Psi_{it}|_a\}f(u)du\right| \\
= & 4\left|\mathsf{E}\{NT\}^{-\frac{1}{2}}\delta^\top|\xi_{it}|_a f(u)\{\Psi_{it}^\top\Psi_{it} \vee \mathsf{E}|\Psi_{it}|_a^\top \mathsf{E}|\Psi_{it}|_a\}\right|
\end{aligned}
$$

with $u \in (0, \{NT\}^{-\frac{1}{2}}\delta^\top|\xi_{it}|_a)$ and $f(u) \leqslant \pi_5$ (by Assumption 3.6 (R2)) is the density function of u_{it}^* conditioning on $\mathcal{C}$ and ξ_{it}. Therefore,

$$
\begin{aligned}
\|\eta_{it}\|_2 = \left[\mathsf{E}\left(\eta_{it}^\top\eta_{it}\right)\right]^{1/2} \leqslant & \left|\{NT\}^{-\frac{1}{2}}\mathsf{E}\left\{\delta^\top\xi_{it}f(u)\{\Psi_{it}^\top\Psi_{it} \vee \mathsf{E}|\Psi_{it}|_a^\top \mathsf{E}|\Psi_{it}|_a\}\right\}\right|^{1/2} \\
\leqslant & \pi_5\{NT\}^{-\frac{1}{4}}\mathsf{E}\left\{\left|\delta^\top\xi_{it}\{\Psi_{it}^\top\Psi_{it} \vee \mathsf{E}|\Psi_{it}|_a^\top \mathsf{E}|\Psi_{it}|_a\}\right|\right\}^{1/2}.
\end{aligned}
$$

By Assumption 3.3 (ii), the last term $\mathsf{E}\left\{\left|\delta^\top\xi_{it}\{\Psi_{it}^\top\Psi_{it} \vee \mathsf{E}|\Psi_{it}|_a^\top \mathsf{E}|\Psi_{it}|_a\}\right|\right\}^{1/2}$ is bounded.

Hence, we obtain that

$$
\mathsf{Var}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\eta_{it}\right] \leqslant C\max_i\|\eta_{it}\|_2 = \mathcal{O}_p(\{NT\}^{-\frac{1}{4}})
$$

for some $C < \infty$. □

References

Acemoglu, D., Ozdaglar, A. and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks, *American Economic Review* **105**(2): 564–608.

Andrews, D. W. et al. (1993). *Empirical process methods in econometrics*, Citeseer.

Anton, M. and Polk, C. (2014). Connected stocks, *The Journal of Finance* **69**(3): 1099–1127.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks, *Science* **286**(5439): 509–512.

Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression, *Journal of the American Statistical Association* **73**(363): 618–622.

Betz, F., Hautsch, N., Peltonen, T. A. and Schienle, M. (2016). Systemic risk spillovers in the european banking and sovereign network, *Journal of Financial Stability* **25**: 206–224.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics* **52**(1-2): 115–127.

Chen, C. Y.-H., Härdle, W. K. and Okhrin, Y. (2019). Tail event driven networks of SIFIs, *Journal of Econometrics* **208**(1): 282–298.

Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects, *Econometrica* **73**(1): 245–261.

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models, *Journal of Econometrics* **132**(2): 491–525.

Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach, *Journal of Econometrics* **142**(1): 379–398.

Clauset, A., Shalizi, C. R. and Newman, M. E. (2009). Power-law distributions in empirical data, *SIAM review* **51**(4): 661–703.

Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: models & applications*, Taylor & Francis.

Diebold, F. X. and Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms, *Journal of Econometrics* **182**(1): 119–134.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles, *Journal of Business & Economic Statistics* **22**(4): 367–381.

Fafchamps, M. and Gubert, F. (2007). Risk sharing and network formation, *American Economic Review* **97**(2): 75–79.

Frölich, M. and Melly, B. (2013). Unconditional quantile treatment effects under endogeneity, *Journal of Business & Economic Statistics* **31**(3): 346–357.

Galvao Jr, A. F., Montes-Rojas, G. and Park, S. Y. (2013). Quantile autoregressive distributed lag model with an application to house price returns, *Oxford Bulletin of Economics and Statistics* **75**(2): 307–321.

Garcia, D. and Norli, Ø. (2012). Geographic dispersion and stock returns, *Journal of Financial Economics* **106**(3): 547–565.

Härdle, W. K., Wang, W. and Yu, L. (2016). Tenet: Tail-event driven network risk, *Journal of Econometrics* **192**(2): 499–513.

Hautsch, N., Schaumburg, J. and Schienle, M. (2014). Forecasting systemic impact in financial networks, *International Journal of Forecasting* **30**(3): 781–794.

Hautsch, N., Schaumburg, J. and Schienle, M. (2015). Financial network systemic risk contributions, *Review of Finance* **19**(2): 685–738.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs, *Journal of the American Statistical Association* **76**(373): 33–50.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields, *Journal of Econometrics* **150**(1): 86–98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence, *Journal of Econometrics* **170**(1): 178–190.

Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica: journal of the Econometric Society* pp. 33–50.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression, *Journal of the American Statistical Association* **94**(448): 1296–1310.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression, *Journal of the American Statistical Association* **101**(475): 980–990.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference*, Springer Science & Business Media.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica* **72**(6): 1899–1925.

Machado, J. A. and Silva, J. S. (2019). Quantiles via moments, *Journal of Econometrics* **213**(1): 145–173.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures, *Journal of the American Statistical Association* **96**(455): 1077–1087.

Pirinsky, C. and Wang, Q. (2006). Does corporate headquarters location matter for stock returns?, *The Journal of Finance* **61**(4): 1991–2015.

Su, L. and Hoshino, T. (2016). Sieve instrumental variable quantile regression estimation of functional coefficient models, *Journal of Econometrics* **191**(1): 231–254.

Su, L. and Yang, Z. (2011). Instrumental variable quantile estimation of spatial autoregression models, *Manuscript* .

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs, *Journal of the American Statistical Association* **82**(397): 8–19.

White, H., Kim, T.-H. and Manganelli, S. (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles, *Journal of Econometrics* **187**(1): 169–188.

Wu, W. B. (2011). Asymptotic theory for stationary processes, *Statistics and its Interface* **4**(2): 207–226.

Wüthrich, K. (2019). A closed-form estimator for quantile treatment effects with endogeneity, *Journal of Econometrics* **210**(2): 219–235.

Wüthrich, K. (2020). A comparison of two quantile models with endogeneity, *Journal of Business & Economic Statistics* **38**(2): 443–456.

Xu, X. and Lee, L.-f. (2015). Maximum likelihood estimation of a spatial autoregressive tobit model, *Journal of Econometrics* **188**(1): 264–280.

Zhao, Y., Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models, *The Annals of Statistics* pp. 2266–2292.

Zhu, X., Chang, X., Li, R. and Wang, H. (2019). Portal nodes screening for large scale social networks, *Journal of Econometrics* **209**(2): 145–157.

Zhu, X., Huang, D., Pan, R. and Wang, H. (2018). Multivariate spatial autoregression for large scale social networks, *Journal of Econometrics* **To appear**.

Zhu, X. and Pan, R. (2018). Grouped network vector autoregression, *Statistica Sinca* .

Zhu, X., Pan, R., Li, G., Liu, Y., Wang, H. et al. (2017). Network vector autoregression, *The Annals of Statistics* **45**(3): 1096–1123.

Zhu, X., Wang, W., Wang, H. and Härdle, W. K. (2019). Network quantile autoregression, *Journal of Econometrics* **212**(1): 345–358.

Zhu, X. et al. (2020). Nonconcave penalized estimation in sparse vector autoregression model, *Electronic Journal of Statistics* **14**(1): 1413–1448.

IRTG 1792 Discussion Paper Series 2020

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Estimation and Determinants of Chinese Banks' Total Factor Efficiency: A New Vision Based on Unbalanced Development of Chinese Banks and Their Overall Risk" by Shiyi Chen, Wolfgang K. Härdle, Li Wang, January 2020.

002 "Service Data Analytics and Business Intelligence" by Desheng Dang Wu, Wolfgang Karl Härdle, January 2020.

003 "Structured climate financing: valuation of CDOs on inhomogeneous asset pools" by Natalie Packham, February 2020.

004 "Factorisable Multitask Quantile Regression" by Shih-Kang Chao, Wolfgang K. Härdle, Ming Yuan, February 2020.

005 "Targeting Cutsomers Under Response-Dependent Costs" by Johannes Haupt, Stefan Lessmann, March 2020.

006 "Forex exchange rate forecasting using deep recurrent neural networks" by Alexander Jakob Dautel, Wolfgang Karl Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2020.

007 "Deep Learning application for fraud detection in financial statements" by Patricia Craja, Alisa Kim, Stefan Lessmann, May 2020.

008 "Simultaneous Inference of the Partially Linear Model with a Multivariate Unknown Function" by Kun Ho Kim, Shih-Kang Chao, Wolfgang K. Härdle, May 2020.

009 "CRIX an Index for cryptocurrencies" by Simon Trimborn, Wolfgang Karl Härdle, May 2020.

010 "Kernel Estimation: the Equivalent Spline Smoothing Method" by Wolfgang K. Härdle, Michael Nussbaum, May 2020.

011 "The Effect of Control Measures on COVID-19 Transmission and Work Resumption: International Evidence" by Lina Meng, Yinggang Zhou, Ruige Zhang, Zhen Ye, Senmao Xia, Giovanni Cerulli, Carter Casady, Wolfgang K. Härdle, May 2020.

012 "On Cointegration and Cryptocurrency Dynamics" by Georg Keilbar, Yanfen Zhang, May 2020.

013 "A Machine Learning Based Regulatory Risk Index for Cryptocurrencies" by Xinwen Ni, Wolfgang Karl Härdle, Taojun Xie, August 2020.

014 "Cross-Fitting and Averaging for Machine Learning Estimation of Heterogeneous Treatment Effects" by Daniel Jacob, August 2020.

015 "Tail-risk protection: Machine Learning meets modern Econometrics" by Bruno Spilak, Wolfgang Karl Härdle, October 2020.

016 "A data-driven P-spline smoother and the P-Spline-GARCH models" by Yuanhua Feng, Wolfgang Karl Härdle, October 2020.

017 "Using generalized estimating equations to estimate nonlinear models with spatial data" by Cuicui Lu, Weining Wang, Jeffrey M. Wooldridge, October 2020.

018 "A supreme test for periodic explosive GARCH" by Stefan Richter, Weining Wang, Wei Biao Wu, October 2020.

019 "Inference of breakpoints in high-dimensional time series" by Likai Chen, Weining Wang, Wei Biao Wu, October 2020.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2020

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

020 "Long- and Short-Run Components of Factor Betas: Implications for Stock Pricing" by Hossein Asgharian, Charlotte Christiansen, Ai Jun Hou, Weining Wang, October 2020.

021 "Improved Estimation of Dynamic Models of Conditional Means and Variances" by Weining Wang, Jeffrey M. Wooldridge, Mengshan Xu, October 2020.

022 "Tail Event Driven Factor Augmented Dynamic Model" by Weining Wang, Lining Yu, Bingling Wang, October 2020.

023 "The common and speci
fic components of infl ation expectation across European countries" by Shi Chen, Wolfgang Karl Härdle, Weining Wang, October 2020.

024 "Dynamic Spatial Network Quantile Autoregression" by Xiu Xu, Weining Wang, Yongcheol Shin, October 2020.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.