Latent Quantile Network: Estimation and Inference

Stan Koobs* Ryo Okui† Yutao Sun‡ Wendun Wang§

February 22, 2024

Abstract

We propose methods for the estimation of an unknown network, specifically the corresponding adjacency matrix, from a panel data set in which the individuals are connected through the network. In particular, we allow this network to be quantile-dependent which involves links that mutate across data quantiles. To address this, we utilize a linear quantile regression model, treating the adjacency matrix entries as model parameters. We impose a sparsity assumption on the network and employ standard regularization techniques to improve estimation efficiency. Furthermore, we enable valid post-selection inference on other policy/treatment variables in the model. Simulation studies are conducted to investigate the performance of our methods. In addition, we apply our methods to study sovereign credit risks and find interactions of risks that are not explained by geographical factors.

JEL Classification: C21, C23

Keywords: Spillover effects; Quantile regression; Panel data; High-dimensional parameters; Double machine learning

*Econometric Institute, Erasmus University Rotterdam; Tinbergen Institute. Email: koobs@ese.eur.nl.

†Faculty of Economics, Graduate School of Economics, University of Tokyo. Email: okuiryo@e.u-tokyo.ac.jp.

‡Institute for Advanced Economic Research (IAER) at Dongbei University of Finance and Economics. Email: yutao.sun@dufe.edu.cn.

§Econometric Institute, Erasmus University Rotterdam; Tinbergen Institute. Email: wang@ese.eur.nl.

1 Introduction

There has been an increasing amount of evidence that individuals are more connected than ever, interacting with each other and shaping their decisions through these connections. Information about such connections, referred to as networks, has been widely utilized in empirical studies across various domains including education (De Giorgi and Pellizzari, 2014), portfolio allocation (Bursztyn et al., 2014), crime (Malm and Bichler, 2011), financial market contagion (Forbes and Rigobon, 2002), R&D collaborations (König et al., 2019), and more. For comprehensive reviews, we refer to (Jackson et al., 2017) and (de Paula, 2017).

One of the main challenges in empirical applications is that the true underlying network is rarely observed by researchers. A common solution is to rely on a postulated network constructed from measures of geographic or economic distance, among other factors. However, this approach may have flaws as the postulated network may not be consistent with its true underlying counterpart for at least two reasons. First, the formation of such a network is often driven simultaneously by a combination of multiple economic, geographic, or social factors (Carrell et al., 2013) and by unobservables (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Han et al., 2019). Second, the exact mechanism of formation, including the roles of the driving factors and their interactions, is often unclear to researchers. Additionally, the network may vary across different locations of the data distribution, exhibiting quantile dependency. A motivating example is provided by Zhu et al. (2019), who find that dependencies of volatilities differ across quantile levels in the Chinese financial market, with much stronger dependencies at tail levels. This raises another challenge in capturing the quantile-varying feature of the network. Most existing methods, however, rely on the assumption of a constant network, thus unable to capture this feature.

The first contribution of this paper is that we propose a new method for estimating an unobserved, and most importantly, a fully unspecified and quantile-dependent network. We consider a panel dataset where individual outcomes are influenced by the characteristics of their connected peers through a network. These peer characteristics are considered to have distributional effects on an individual's outcome, and we allow the network to exhibit quantile dependency, leading to a quantile regression approach. Notably, the features of the network need not be specified a priori. Our primary interest lies in estimating the entries of the adjacency matrix associated with the unknown network, along with doing valid inference on other model parameters as well. All parameters are permitted to vary based on the quantiles of interest. This approach diverges from relevant works such as de Paula et al. (2018) and Manresa

(2016), where quantile-varying networks are generally not allowed. Our problem is high-dimensional, as the number of parameters to be estimated is roughly the square of the number of the cross-sectional units. Similar to de Paula et al. (2018) and Manresa (2016), we consider a sparse network where the number of actual links is small relative to the number of possible ones. It is common to employ sparsity in the network literature and this is well-justified by empirical evidence (see, e.g., Karlan et al., 2009; Leider et al., 2009; De Weerdt and Dercon, 2006; Cai et al., 2015). To accommodate the sparsity, we invoke regularization approaches, in particular, ℓ_1-penalized quantile regression (Belloni and Chernozhukov, 2011) using adaptive weights (Zou, 2006) in our estimation.

Utilizing regularization techniques for network estimation can pose challenges when conducting inference on another policy or treatment variable in the model. Only under highly unrealistic assumptions, the network will be selected perfectly (where 'selection' refers to setting the corresponding element in the network matrix to nonzero). However, in practice, moderate selection errors are common and can have serious consequences for inference, as demonstrated by Leeb and Pötscher (2005). To address such challenges, we draw upon insights from the emerging debiased machine learning (DML) literature (Chernozhukov et al., 2018; Belloni et al., 2013, 2014; Zhang and Zhang, 2014; van de Geer et al., 2014). In particular, we will adopt the methodology outlined by Belloni et al. (2019), which has been developed for cross-sectional quantile regression analysis. The second main contribution of this paper is that we adapt their estimation techniques to the panel data framework, where the confounding factors in our model comprise of a fixed effect and potential network effects. The estimators rely on constructing an orthogonal score function, either explicitly or implicitly, resulting in a moment condition robust against first-order mistakes in estimating network parameters. The approach involves an initial estimate of the network matrix using ℓ_1-penalized quantile regression methods and partialling out its confounding influence on the treatment variable. For partialling out this effect, we employ a post-Lasso technique as proposed by Belloni et al. (2012). Subsequently, we introduce two treatment effect estimators: one based on an explicit moment condition derived from the orthogonal score function and another utilizing quantile regression incorporating all variables selected in prior steps. The latter approach bears resemblance to the "post-double selection" method introduced by Belloni et al. (2013, 2014). To conduct inference on these variables, we rely on the asymptotic normality results established by Belloni et al. (2019), which provide confidence regions with valid asymptotic coverage.

In the simulations presented in this paper, we first assess the performance of the estimation of the quantile-varying network effects. When examining a large number

of time periods relative to the number of cross-sectional units ($T > N$), we observe the capability to capture underlying network links, with performance improving as the number of time periods increases. Moreover, in terms of inference on the treatment variable, we compare the performance of the two proposed estimators with ℓ_1-penalized, post-ℓ_1-penalized, and ordinary quantile regression. Our estimators substantially reduce bias compared to the ℓ_1-penalized and post-ℓ_1-penalized estimators. Furthermore, we demonstrate valid coverage while demonstrating greater efficiency than ordinary quantile regression.

We apply our method to investigate the determinants of sovereign credit risk, explicitly taking into account the international connection of the risk. Cross-country spillovers of risks are expected to play an important role in the sovereign credit risks due to deep integration and globalization, and the interactions between countries may also evolve over time. Several important insights emerge from our analysis. First, while the geographic pattern remains salient in the network structure, some strong connections are not captured by the geographic locations but economic relations, such as strategic economic partnership or heavy trading relationship. Second, our empirical estimates generally confirm the sparsity of the network. We also find that the network structure is not always symmetric, suggesting that the risk spillovers are directional.

The rest of the paper is organized in the following way. Section 2 reviews the literature and links with related studies. Section 3 sets up the model and explains our estimation strategy and subsequent inference. To evaluate the finite-sample performance of our approach, we provide simulation studies in Section 4. An empirical study on investigating international interactions of sovereign credit risks is provided in Section 5. Finally, Section 6 concludes.

2 Literature review

The literature on recovering networks is extensive, albeit with limitations. One significant body of research examines postulated networks based on characteristics such as geographic locations (Ciccarelli and Elhorst, 2018; Ho et al., 2018), social homophily (Topa, 2001; Sacerdote, 2001), and economic distance (Qu and Lee, 2015). Identification results and various estimation methods for models dependent on this type of network have been established, including works by Ord (1975); Lin and Lee (2010); Lee (2007); Bramoullé et al. (2009); Goldsmith-Pinkham and Imbens (2013); Lee et al. (2010). However, accurately specifying a network is challenging since the underlying network is often poorly understood. Moreover, deviations of the postulated network from the underlying one could result in highly misleading estimates

(Manresa, 2016). Endogeneity problems may also arise because the variables used to construct the adjacency matrix, such as "economic distance," may be correlated with the final outcome or with some unobserved characteristics affecting both the network and outcome (Qu and Lee, 2015; Hsieh and Lee, 2016).

To avoid a fully postulated network and the potential endogeneity problem, another line of methods has been proposed. These methods aim to model, rather than postulate, the network formation process based on individual decisions or characteristics. For example, Goldsmith-Pinkham and Imbens (2013) propose a strategic network formation model where units form a link if they find positive net utility in doing so. Similarly, Hsieh and Lee (2016) parameterize the network formation process using a logit function, where the probability of each link is determined by a pre-specified set of covariates. Hsieh et al. (2020) model the network formation via an exponential probability distribution, while Johnsson and Moon (2021) adopt a control function approach. While these modeling strategies relax the assumption of a fully postulated network, they still necessitate specification of the network formation process on different aspects, such as driving variables, link probabilities, or utility functions. Any misspecification of the formation process could lead to inconsistency.

The formation of networks is also examined in other contexts, and this stream of literature focuses on explaining how the connections form and evolve. For example, Jackson et al. (2012) propose a game theoretic foundation for network formation in the framework of informal favor exchange. Haag and Lagunoff (2006) study the formation of cliques or clusters in a network. Currarini et al. (2009) examine, in the context of sociology, the effect of individual characteristics on the formation of links, and McPherson et al. (2001) formulate, mathematically, the concept of homophily and propose an economic model for friendship formation. Graham (2017) further pursues the concept of homophily and proposed an econometric model and estimation procedures to detect homophily with heterogeneous agents. Dzemski (2019) studies a dyadic link formation model where direct links are formed between agents with homophily and reciprocity. Most of these techniques are in contexts of individual or household behaviors, and thus ad hoc to different extents.

Two very related papers are by Manresa (2016) and de Paula et al. (2018). They allow for a general and fully unspecified network which they estimate from the data. Manresa (2016) considers panel data and a conditional mean model in which the outcome of a unit depends on the characteristics of his/her own, and that of others. A double pooled LASSO procedure has been proposed to recover the latent network together with the network externalities, or "spillover" in the terminology of Manresa (2016). de Paula et al. (2018) consider a similar model in which and in addition to the covariates, the outcome also generates externalities; and thus, allowing the

presence of both endogenous and exogenous social effects, in the terminology of Manski (1993). They discuss the identification of the network and propose to estimate the adjacency matrix using the GMM regularized by the adaptive elastic net. Our work is similar in that we also allow for a general and fully unspecified network, and estimate the adjacency matrix using regularization methods. We consider, however, only the exogenous social effect, since dealing with exogeneity in a quantile regression setting is difficult. In addition, instead of the conditional mean model, we consider a conditional quantile model in which the network may vary across different quantile levels and we propose several estimators to enable post-selection valid inference for a treatment variable.

Another important related paper is the work by Hautsch et al. (2015), who also use ℓ_1-penalized quantile regression to estimate quantile-varying network matrices in a financial network. Their estimation procedure is slightly different as they only have unit-specific parameters which allows them to estimate all parameters per unit (dropping the panel aspect). Our contribution is that we also allow for common parameters in the model and the estimation of the network is only a preliminary step to doing further inference on these. To better understand the network estimation, we also perform a more extensive simulation study. Other related papers include, for instance, Zhu et al. (2019) who follow a vector autoregression approach where the conditional quantile of the vector of outcomes is affected by its temporal and spatial lags. Like most existing studies but unlike ours, Zhu et al. (2019) require prior knowledge of the adjacency matrix. In their analysis of financial risk contagion mechanisms that arise from common shared ownership, the adjacency matrix is constructed from the top ten shareholders' characteristics. Another example is Han et al. (2019) who model the dynamic network formation and interactions jointly. They assume that the network is unweighted so that the entries of the adjacency matrix are binary. Each of the entries follows a probability distribution which is further modelled using a logit function of time-varying observables and unobservables, including observed homophily measures, persistence and transitivity, and unobserved homophily and heterogeneity. We resemble this study in that we allow latent and possibly time-varying (in the sense of quantile-varying) networks as well, while not imposing any specific functional forms or variable specifications for the network formation process. Instead, our network recovery is achieved by exploiting repeated observations of individuals in a panel data set.

In addition, many existing methods consider the presence of fixed effects. For instance, Manresa (2016) allows for individual effects whereas de Paula et al. (2018) allow, as an extension, individual and time effects. Our approaches are alike, as we consider the presence of individual effects. The inclusion of fixed effects may

induce the so-called incidental parameter problems of Neyman and Scott (1948). Our ambition is not to provide solutions to the incidental parameter problem in this context. Therefore, we mostly focus on settings where the number of time periods is relatively large compared to the number of cross-sectional units which allows us to utilize the results derived by Kato et al. (2012) and Galvao et al. (2020). We refer to Fernández-Val and Weidner (2018) for a recent survey of the development in this field.

This paper also contributes to the literature on applying DML techniques to panel data models. While DML techniques are quite general, there are no such examples provided in Chernozhukov et al. (2018). One important paper in this literature is Belloni et al. (2016), where valid inference techniques for high-dimensional linear panel data models are developed. More recent work has focused on dynamic panel data models (Semenova et al., 2023). To the best of our knowledge, DML techniques have not been applied to panel data models that account for cross-sectional dependence, as we do by incorporating exogenous spillover effects.

3 Model setup and estimation

In this section, we begin by presenting a linear panel data model incorporating quantile-dependent spillover effects. This model posits that a unit's dependent variable is impacted not just by its own characteristics but also by those of other units. As the number of units increases, the number of parameters to estimate grows roughly quadratically. Therefore, we require a penalized estimation approach capable of handling high-dimensional parameters. Next to that, to conduct inferince on other treatment variables in the model, we propose several estimators that integrate the network estimation as the initial step.

3.1 Model

We consider a panel data set where the individuals are connected through an unknown but exogenous network, and the purpose of our work is to recover the network (i.e., its adjacency matrix) using quantile level information. Let y_{it} be a continuous scalar outcome variable for individual $i = 1, \ldots, N$ at time period $t = 1, \ldots, T$. Next, let x_{it} and z_{it} represent scalar covariates. The distinction between x_{it} and z_{it} lies in the fact that the former denotes an individual characteristic that may generate spillovers, while z_{it} comprises another observable that does not lead to spillovers. For simplicity, we maintain x_{it} and z_{it} as scalars, although this can readily be extended

to multivariate covariates. We consider the following data generating process (DGP)

$$y_{it} = \sum_{j=1}^{N} W_{ij}(U_{it}) x_{jt} + \theta(U_{it}) z_{it} + \alpha_i(U_{it}), \tag{3.1}$$

where U_{it} is an i.i.d. uniform random variable on $(0,1)$, $W_{ij}(U_{it})$ denotes the (i,j)th element of the unknown adjacency matrix $W(U_{it})$ which captures the exogenous social effect of x_{jt} on y_{it}, $\theta(U_{it})$ is an unknown scalar parameter that captures the effect of the individual's characteristics on its own outcome, and $\alpha_i(U_{it})$ is the individual effect.

Equation (3.1) corresponds to a random coefficient representation of a quantile regression model where the DGP itself does not have an explicit error term. All parameters, including the adjacency matrix, are functions of the common random variable U_{it} and, thus, the randomness comes solely from U_{it}. This is similar to other works of, e.g., Su and Yang (2007) and Graham et al. (2015). We use this representation as it is convenient to choose a particular network matrix as a function of the data quantile. Another common set-up is the location-scale representation in, e.g., Koenker and Xiao (2002) for which we will also demonstrate simulation results in Appendix C. However, for now we focus on the random coefficient representation.

A further step is necessary to transform the random coefficient representation in (3.1) into a linear quantile regression model. Specifically, we assume that for each $x_t := (x_{1t}, \ldots, x_{NT})'$ and z_{it}, the right-hand side of (3.1) is monotonically increasing in U_{it}. This ensures that the quantile function of y_{it} also monotonically increases in U_{it}, a standard assumption in the quantile regression literature and for the rest of this paper we will refer to it as the monotonocity assumption. Doksum (1974) interprets the disturbance U_{it} as individual "ability" or "proneness". It, however, shall be noted that this monotonicity assumption does not always hold in empirical studies. When this assumption is violated, one may follow Bondell et al. (2010) by adding a monotonicity constraint in the estimation; or follow Chernozhukov et al. (2010) by considering a reordering.

Next, we show an example where this assumption is satisfied.

Example 3.1 *Let us consider a simplified version of (3.1) where* $\theta(U_{it}) = 0$, $\alpha_i(U_{it}) = 0$ *for all* i. *Consider* x_{it} *being non-negative such that* $x_{it} \geq 0$ *for all* i *and* t. *It can then be verified that the monotonicity assumption is satisfied for every* $W_{ij}(U_{it})$ *satisfying* $W_{ij}(U_{it}) \leq W_{ij}(U'_{it})$ *for* $U_{it} \leq U'_{it}$.

A special case of such a $W(U_{it})$ *can be, e.g. for* $N = 3$,

$$W(U_{it}) := (1 + U_{it}) \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{for } U_{it} < 0.5,$$

$$W(U_{it}) := (1 + U_{it}) \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{for } U_{it} \geq 0.5.$$

This provides an example of how particular links in the network can become nonzero for higher quantiles. In practice, one could think about financial applications where a network of firms becomes more "dense" in times of high volatility.

Under the monotonicity condition, the conditional quantile function of y_{it} can be written as

$$Q_{it}(\tau) := Q_{y_{it}}(\tau | x_t, z_{it}) = \sum_{j=1}^{N} W_{ij}(\tau) x_{jt} + \theta(\tau) z_{it} + \alpha_i(\tau), \tag{3.2}$$

or more compactly in matrix notation,

$$Q_t(\tau) = W(\tau) x_t + \theta(\tau) z_t + \alpha(\tau),$$

where $\tau \in (0,1)$ is a quantile of interest, $Q_t(\tau) := (Q_{1t}(\tau), \ldots, Q_{NT}(\tau))'$, $z_t := (z_{1t}, \ldots, z_{NT})'$, and $\alpha(\tau) := (\alpha_1(\tau), \ldots, \alpha_n(\tau))'$.

3.2 Estimation of the network

The purpose of our work is to estimate the unknown adjacency matrix $W(\tau)$ together with other model parameters in (3.1). In addition, as many empirical applications find that $W(\tau)$ is high-dimensional but sparse, we employ ℓ_1-penalized quantile regression (Belloni and Chernozhukov, 2011) in combination with the adaptive LASSO from Zou (2006) in the estimation to improve the estimation efficiency. This means we focus on the following optimization problem

$$(\widehat{\theta}(\tau), \widehat{W}(\tau), \widehat{\alpha}(\tau)) \in \underset{\theta, W, \alpha}{\arg\min} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau \left(y_{it} - z_{it}\theta - \sum_{j=1}^{N} x_{jt} W_{ij} - \alpha_i \right) + D(W), \tag{3.3}$$

where $\rho_\tau(u) := (\tau - \mathbb{1}\{u \leq 0\})u$ denotes the check function (Koenker and Bassett, 1978) and $D(W)$ the penalization on the spillover effects. This term is given by

$$D(W) = \lambda \sum_{i=1}^{N} \sum_{j=1}^{N} \mu_{ij} |W_{ij}|. \tag{3.4}$$

where λ and μ_{ij} are, respectively, the penalty parameter and weights involved in the adaptive LASSO procedure. In particular, we set $\mu_{ij} = 1/\sqrt{|\breve{W}_{ij}(\tau)|}$ where $\breve{W}_{ij}(\tau)$ denotes an estimate of $W_{ij}(\tau)$ using unregularized quantile regression.

In our model, we have both unit-specific parameters like $W_{i1}(\tau), \ldots, W_{jN}(\tau)$ and $\alpha_i(\tau)$, as well as a common parameter $\theta(\tau)$. This makes our estimation procedure slightly different from Hautsch et al. (2015) as they consider a similar quantile-varying network model but without a common parameter. To estimate our model, we first write the DGP (3.1) in full matrix notation. We can write $y = S\beta(U)$ where $y = (y_{11}, y_{12}, \ldots, y_{NT})$ and

$$S = \begin{pmatrix}
1 & \mathbf{0}'_{N-1} & x'_1 & \mathbf{0}'_{N(N-1)} & z_{11} \\
1 & \mathbf{0}'_{N-1} & x'_2 & \mathbf{0}'_{N(N-1)} & z_{12} \\
\vdots & & \vdots & & \vdots \\
1 & \mathbf{0}'_{N-1} & x'_T & \mathbf{0}'_{N(N-1)} & z_{1T} \\
0 & 1 \; \mathbf{0}'_{N-2} & \mathbf{0}'_N & x'_1 \; \mathbf{0}'_{N(N-1)} & z_{21} \\
0 & 1 \; \mathbf{0}'_{N-2} & \mathbf{0}'_N & x'_2 \; \mathbf{0}'_{N(N-1)} & z_{22} \\
\vdots & & \vdots & & \vdots \\
\mathbf{0}'_{N-1} & 1 & \mathbf{0}'_{N(N-1)} & x'_T & z_{NT}
\end{pmatrix}$$

which is a $NT \times (N + N^2 + 1)$ matrix. Lastly, $\beta(U) = (\alpha(U)', \text{rvec}(W(U))', \theta(U))'$, where rvec denotes the vec operation by row-wise stacking. This system can be estimated by ℓ_1-penalized quantile regression estimation methods. For moderate values of N, this can already become computationally expensive. We directly utilize the existing code from `R` package 'quantreg' by Koenker (2023) which proves to be quite fast. For the rest of this paper, we refer to the above estimation method as the ℓ_1-penalized network estimation.

Lastly, a model selection technique needs to be invoked for the selection of λ. There are two commonly used methods: the cross validation and information criteria (IC) minimization. While theoretically both may apply, our preference is to consider cross validation (see, e.g., Arlot and Celisse, 2010), because of its simplicity, where IC-based approaches, e.g., Lee et al. (2014), would depend on extra tuning parameters

which, in turn, adds extra complexity to our model. For our purpose, we partition the data set along the time dimension. In typical panel settings, it would be more natural to partition the individuals. However, in our setting, the individuals are network-connected, which renders partitioning the individuals infeasible; whereas by partitioning the time periods, we implicitly rule out the possibility of temporal dependency in the model as well as the data generating process.

3.3 Estimation of the treatment effect

Let us now direct our attention to the estimation of the treatment effect $\theta(\tau)$. In addition to the model in (3.1), we introduce the following equation to account for confounding in the model:

$$z_{it} = \sum_{j=1}^{N} W^c_{0,ij} x_{jt} + v_{it}. \tag{3.5}$$

Here, W^c_0 denotes the "confounding matrix" which is the matrix of parameters that determines how the x regressors affect the treatment z and where the subscript "0" denotes its true value. Specifically, $W^c_{0,ij}$ denotes the (i,j)th element of W^c_0 which indicates the influence of x_{jt} on z_{it}. Additionaly, v_{it} is a disturbance term. This equation tracks the dependence of the treatment variable on the other regressors, similar to the methodology discussed by Chernozhukov et al. (2018). For now, we assume that $W^c_{0,ij}$ is quantile-invariant. While this assumption could be extended to cases where $W^c_{0,ij}$ also varies with the quantile, for simplicity, we assume it remains constant across all quantiles. Equation (4.1) is crucial for understanding how different estimation techniques for $W(\tau)$ affect the estimates of $\theta(\tau)$.

One possible way to estimate the treatment effect is to use the network estimation method described above, which also delivers an estimate of the treatment effect given by $\widehat{\theta}(\tau)$. However, a drawback of this estimate is that the estimates of $W(\tau)$ suffer from bias due to the regularization bias introduced by the ℓ_1-penalized quantile regression. Although no penalization is applied to $\theta(\tau)$, this regularization bias extends to $\widehat{\theta}(\tau)$ when the regressors are correlated with the treatment. In practical scenarios, there is often strong correlation between the treatment and other regressors, leading to significant implications for treatment inference when using such penalized estimation techniques. We refer to $\widehat{\theta}(\tau)$ as the "naive" estimator.

A more sensible way to estimate $\theta(\tau)$ without regularization bias involves using techniques that mitigate the regularization bias in the estimates of $W(\tau)$. One popular technique to do this is the post-Lasso (Belloni et al., 2012), which runs

ordinary quantile-regression on the regressors selected by the first-stage ℓ_1-penalized quantile regression. This method performs well under perfect selection, i.e., when all relevant variables are included in the model and irrelevant ones are excluded. Nonetheless, this method commonly encounters moderate selection mistakes in the first-stage. When variables are left out of the model which are correlated with the treatment, this leads to omitted variable bias. As has been shown in the work by Leeb and Pötscher (2005), disregaring the model-selection step can therefore lead to invalid inference. We refer to this estimation technique as the post-ℓ_1-penalized quantile regression.

The goal of this paper is to utilize the emerging DML literature to obtain estimators which are insensitive to such model selection mistakes coming from the network estimation. Here, we will mainly follow the methodology proposed by Belloni et al. (2019) who offer multiple estimators for high-dimensional sparse quantile regression models in the cross-sectional case. The contribution we make is to extend these methods to panel data quantile regression where the high-dimensionality comes from the spillover effects.

The key to do this is to use a moment condition for $\theta(\tau)$ that satisfies Neyman orthogonality. Intuitively, one could think of this as a moment condition that is insensitive to first-order changes in the nuisance parameters, which are W, α and W^c for our model. For this purpose, we define the following orthogonal score function

$$\psi_{it}(\theta) = \left(\tau - \mathbb{1}\left\{y_{it} \leq \sum_{j=1}^{N} W_{ij}(\tau)\, x_{jt} + \theta z_{it} + \alpha_i(\tau)\right\}\right)\left(z_{it} - \sum_{j=1}^{N} W^c_{0,ij} x_{jt}\right).$$

Using (3.2), one then obtains a moment condition for $\theta(\tau)$ given by

$$\mathbb{E}\left[\left(\tau - \mathbb{1}\left\{y_{it} \leq \sum_{j=1}^{N} W_{ij}(\tau)\, x_{jt} + \theta(\tau) z_{it} + \alpha_i(\tau)\right\}\right)\left(z_{it} - \sum_{j=1}^{N} W^c_{0,ij} x_{jt}\right)\right] = 0. \tag{3.6}$$

Let $f_{it} = f_{y_{it}}(Q_{it}(\tau)|x_t, z_{it})$ denote the conditional density of y_{it} conditional on x_t, z_{it} evaluated at the conditional function $Q_{it}(\tau)$. Then, under the assumptions that $\mathbb{E}[f_{it} x_t v_{it}] = \mathbf{0}$ and $\mathbb{E}[f_{it} v_{it}] = 0$ for all i and t, the moment condition in (3.6) satisfies Neyman orthogonality with respect to first-order changes in the values of the nuisance parameters W, α, and W^c. Mathematically, this means that the Gateaux derivative of the moment condition w.r.t. any of these nuisance parameters evaluated at the true parameter value equals zero. For further details on this condition, we refer the reader to Appendix A.

Using the above moment condition, we turn to the first estimator we propose where we construct this orthogonal score function explicitly. The estimation proce-

dure is described in the algorithm below. Penalty parameter choices are deferred to Remark 1.

Orthogonal score algorithm

Step 1: *Estimate* $(\widehat{\theta}(\tau), \widehat{W}(\tau), \widehat{\alpha}(\tau))$ *from post-*ℓ_1*-penalized quantile regression of* y_{it} *on* z_{it} *and* x_t *described in Section 3.2.*

Step 2: *Compute* $(\tilde{\theta}(\tau), \tilde{W}(\tau), \tilde{\alpha}(\tau))$ *from quantile regression of* y_{it} *on the treatment and the regressors selected in Step 1.*

Step 3: *Compute* $\tilde{W}^c$ *from the post-Lasso estimator of* z_{it} *on* x_t*.*

Step 4: *Construct the score function*

$$\widehat{\psi}_{it}(\theta) = \left(\tau - \mathbb{1}\left\{y_{it} \leq z_{it}\theta + \sum_{j=1}^{N} \tilde{W}_{ij}(\tau)x_{jt} + \tilde{\alpha}_i(\tau)\right\}\right)\left(z_{it} - \sum_{j=1}^{N} \tilde{W}^c_{ij}x_{jt}\right).$$

Step 5: *Using the Neyman-type score statistic*

$$L_{NT}(\theta) = \frac{|\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\psi}_{it}(\theta)|^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\psi}^2_{it}(\theta)},$$

set $\check{W}(\tau) = \tilde{W}(\tau)$*,* $\check{\alpha}(\tau) = \tilde{\alpha}(\tau)$ *and* $\check{\theta}^{OS}(\tau) \in \arg\min_{\theta\in\Theta_\tau} L_{NT}(\theta)$*.*

In Step 5 of the Orthogonal Score (OS) algorithm, we minimize the Neyman-type score over the following set

$$\Theta_\tau = \{\theta \in \mathbb{R} : |\theta - \tilde{\theta}| \leq 10(\mathbb{E}_{NT}[z^2_{it}])^{-1/2}/\log(NT)\},$$

where we use $\mathbb{E}_{NT}$ to abbreviate the panel sample average $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}$.

In addition to the above estimator, we also offer a second estimator for $\theta(\tau)$, in the same fashion as Belloni et al. (2019). This estimator can be viewed as a panel quantile regression version of the "post-double selection" method by Belloni et al. (2013, 2014). The estimation procedure is summarized in the algoritm below.

Double Selection algorithm

Step 1: *Estimate* $(\widehat{\theta}(\tau), \widehat{W}(\tau), \widehat{\alpha}(\tau))$ *from* ℓ_1*-penalized quantile regression of* y_{it} *on* z_{it} *and* x_t*.*

Step 2: *Compute $\tilde{W}^c$ from the Lasso estimator of z_{it} on x_t.*

Step 3: *Compute $(\check{\theta}^{DS}(\tau), \check{W}(\tau), \check{\alpha}(\tau))$ from quantile regression of y_{it} on z_{it} and $\{x_{jt} : x_{jt}$ is selected in either Step 1 or Step 2$\}$.*

It is worth noting that the Double Selection (DS) algorithm does not explicitly feature an orthogonal score function. However, as noted by Belloni et al. (2019), an implicit construction of the orthogonal score function occurs within the optimality condition of the quantile regression in Step 3. This indicates that while the algorithm may not explicitly define the orthogonal score function, its presence is embedded within the optimization process of the quantile regression.

Remark 1 (Penalty parameters) *For both the ℓ_1-penalized quantile regression (step 1 of both algorithms) and the (post-)Lasso (step 3 in OS, step 2 in DS), we use the adaptive Lasso weights from (3.4). The penalty parameter λ is chosen using cross-validation by partitioning along the time dimension.*

Belloni et al. (2019) also propose "weighted" versions of the OS and DS algorithm where the auxiliary regression equation (4.1) is weighted by f_{it}. Under particular conditions, this leads to more a efficient estimator. The problem is that one does need to estimate f_{it} which contains estimation error. In our simulations, this generally led to worse finite-sample results. Therefore, we have deferred these estimators to Appendix B and only focus on the "unweighted" versions above.

3.4 Inference on the treatment

To conduct inference, we employ the results derived by Belloni et al. (2019). Under mild moment conditions and approximate sparsity assumptions, they establish asymptotic normality for both the OS and DS estimator. Under the appropriate regularity conditions, the panel version of their results becomes

$$\sigma_{NT}^{-1}\sqrt{NT}(\check{\theta}(\tau) - \theta(\tau)) \rightsquigarrow \mathbb{U}_{NT}(\tau) + o_p(1), \text{ and } \mathbb{U}_{NT}(\tau) \rightsquigarrow \mathcal{N}(0,1),$$

where $\check{\theta}(\tau)$ can refer to either the OS or DS estimator and $\sigma_{NT}^2 = \tau(1-\tau)\mathbb{E}\left[\mathbb{E}_{NT}[v_{it}^2]^{-1}\right]$. The result still applies when we replace σ_{NT}^2 by a consistent estimator. For the OS estimator, we use the following estimator that follows from the moment condition

$$\widehat{\sigma}_{OS}^2 = \left(\mathbb{E}_{NT}\left[\widehat{f}_{it} z_{it} \tilde{v}_{it}\right]\right)^{-2} \mathbb{E}_{NT}\left[\left(\mathbb{1}\left\{y_{it} \leq \check{W}_i(\tau) x_t + z_{it}\check{\theta}(\tau) + \check{\alpha}_i(\tau)\right\} - \tau\right)^2 \tilde{v}_{it}^2\right]. \tag{3.7}$$

Moreover, for the DS estimator we estimate the standard error $\widehat{\sigma}^2_{DS}$ by paired bootstrap of the quantile regression in Step 3 of the algorithm. This leads to confidence regions of the form

$$\mathcal{C}^k_{\xi,N,T} = \left\{\theta \in \mathbb{R} : |\theta - \check{\theta}(\tau)| \leq \widehat{\sigma}_k \Phi^{-1}(1-\xi/2)/\sqrt{NT}\right\}, \text{ for } k = \text{OS or } k = \text{DS}. \tag{3.8}$$

4 Simulation

In this section, we demonstrate our procedure by simulations under various designs. We do not use any network-specific information in our procedure. In particular, all networks are considered directed and weighted during the estimation.

4.1 Design and evaluation

We consider the following DGP

$$\begin{aligned} z_{it} &= \sum_{j=1}^{N} W^c_{ij} x_{jt} + v_{it}, \\ y_{it} &= \sum_{j=1}^{N} W_{ij}(U_{it}) x_{jt} + z_{it}\theta(U_{it}) + \alpha_i(U_{it}). \end{aligned} \tag{4.1}$$

Firstly, we set $\alpha_i(U_{it}) = \alpha_i(1+U_{it})$ for $\alpha_i \sim \chi^2(1)$. For simplicity, we have still taken $\alpha_i(U_{it})$ to be independent of the regressors like a random effect. For the treatment parameter, we $\theta(U_{it}) = 0.6(1+U_{it})$. Now for the regressors, we take x_{it} iid from truncated normal with lower bound at zero and mean at 5. This ensures the regressors are non-negative which allows us to utilize spillover matrices of the form in Example 3.1 and have the monotonicity assumption satisfied.

In particular, we set $W(U_{it}) = \tilde{W}(U_{it})\gamma(U_{it})$ where $\gamma(U_{it}) = 0.6(1+U_{it})$. Now we generate $\tilde{W}(U_{it})$ in an Erdos-Renyi (ER) type procedure which we outline below. The ith row of $\tilde{W}_i(U_{it})$ is set as

$$\tilde{W}_i\left(U_{it}\right) = \tilde{W}_i + V_i\left(U_{it}\right), \tag{4.2}$$

where

$$\tilde{W}_i = \left(\tilde{W}_{i1}, \ldots, \tilde{W}_{ii-1}, 1, \tilde{W}_{ii+1}, \ldots, \tilde{W}_{iN}\right),$$

$$V_i(U_{it}) = (V_{i1}(U_{it}), \ldots, V_{ii-1}(U_{it}), 0, V_{ii+1}(U_{it}), \ldots, V_{iN}(U_{it})).$$

Here, $\tilde{W}_i$ is the U_{it}-invariant component and $V_i(U_{it})$ generates the U_{it}-dependency. Note that the above already implies that we have a vector of ones on the diagonal for every value of U_{it}. For the non-diagonal elements in $\tilde{W}_i$, one element is randomly selected independent of τ and set to 1, while other elements are set to 0. This means that $\tilde{W}_i$ contains two ones and only zeros for the rest. For the U_{it}-dependency, we set $V_{ij}(U_{it}) = \tilde{W}_{ij} + 1$ if

$$\tilde{W}_{ij-1} = 1, \qquad U_{it} \geq 0.5, \qquad j \neq 1, \qquad j \neq i,$$

and $V_{ij}(U_{it}) = 0$ for all other j. This process is essentially adding a nonzero element on each row to the right of a nonzero element for $U_{it} \geq 0.5$ on the condition that the target position is not on the diagonal line and not out of range. In other words, another nonzero elements is added as long the nonzero element is not on the $(i-1)$th or the Nth position. Given that the covariates are non-negative, such a $W_i(U_{it})$ guarantees the monotonicity of the quantile function. Note that every row $W_i(U_{it})$ depend on its own U_{it} and is, thus, (i,t)-variant. In other words, every observation (i,t) could potentially face a distinct $W_i(U_{it})$. This is the same for other model parameters and is how a random coefficient models are formulated. A simulation design involving the location-scale shift model can be found in Appendix C.

Lastly, W^c is taken to be quantile-invariant for simplicity, however one could also allow this to be quantile-variant. All diagonals elements are set to 0.8 and the non-diagonal element are generated using ER (so only one off-diagonal element in every row is nonzero) and set to 0.5. Notice that x_{it} is now positively correlated with z_{it} for all (i,t). Moreover, v_{it} is drawn i.i.d. from $\mathcal{N}(0,1)$.

We set $N \in \{5, 10\}$ while $T \in \{50, 100, 200\}$ and we look at $\tau \in \{0.2, 0.4, 0.6, 0.8\}$ and let the number of replications be $R = 500$, unless otherwise stated. To evaluate the performance of the network estimation, we consider the percentages of correctly and incorrectly estimated links. These percentages are calculated in the following way. Denote $\mathcal{W}_N(\tau) := \{(i,j) : W_{ij}(\tau) \neq 0\}$ and $\mathcal{W}_Z(\tau) := \{(i,j) : W_{ij}(\tau) = 0\}$. We count, for each replication,

$$p_{qr}(\tau) := \frac{\left|\widehat{\mathcal{W}}_q(\tau) \cap \mathcal{W}_r(\tau)\right|}{|\mathcal{W}_r(\tau)|}, \qquad q, r \in \{N, Z\},$$

where $|\cdot|$ denotes the cardinality, $\widehat{\mathcal{W}}_N(\tau) := \{(i,j) : \widehat{W}_{ij}(\tau) \neq 0\}$, and $\widehat{\mathcal{W}}_Z(\tau) := \{(i,j) : \widehat{W}_{ij} = 0\}$. All p_{qr} are then averaged over the replications. Note that p_{NN} is the percentage of correctly estimated nonzero links and p_{ZZ} is the percentage

correctly estimated zero links. Moreover, for the performance of different treatment effect estimators, we consider both bias and coverage. For bias, we average $\widehat{\theta}(\tau)^{(r)} - \theta(\tau)$ over the replications, where the superscript (r) denotes the estimate coming from replication r. For coverage, we compute the confidence regions, such as the one in (3.8) for OS and DS and consider the percentage over all replications in which the true value is contained in this interval.

4.2 Network estimation results

We begin by examining the estimation of the network effects, as understanding the quality of these estimates and how they interplay with the corresponding regularization is crucial for understanding subsequent inference on the treatment parameter. In Table 1, we demonstrate p_{NN} (in the nonzero column) and p_{ZZ} (in the zero column) for several values of N, T and τ. Notably, we observe that the majority of percentages are 50% or higher, and with increasing T, our method demonstrates improved performance in both capturing the true non-zero links and the true zero links. Additionally, we observe a decline in performance as N increases, particularly evident in the percentage of correctly captured non-zero links. This outcome is anticipated, considering that the number of parameters in the network matrix increases quadratically with N. This means that for $N = 10$, the estimation entails 100 entries instead of 25 which is four times as large.

Next to that, when we compare the performance around different quantiles, we note that the performance is generally worse for the middle quantiles $\tau = 0.4$ and $\tau = 0.6$. Usually one would expect the best performance around the median as this quantile contains the most information. However, for this network, some zero entries in fact become nonzero at the median. This means that it is expected that the quantiles around the median struggle to correctly classify these links which explains the decline in performance. We also note that the percentage of correctly captures nonzeros is higher for $\tau = 0.8$ compared to $\tau = 0.2$. Our intuition on this is that the role of the network becomes more significant as τ increases because $W(\tau)$ increases with τ due to $\gamma(\tau)$. Lastly, the performance of the method also depends on the strength of confounding, i.e. the magnitude of the coefficient in W^c. In Appendix D.1, we demonstrate that all percentages increase when one uses weaker confounding.

This demonstrates that for the Erdos-Renyi networks, we can estimate the links in the matrix relatively well for large T, but there still remains some moderate selection mistake. This is important to keep in mind when comparing the different treatment effect estimators.

Table 1: Updated average percentages of correct nonzero and zero links of τ-dependent Erdos-Renyi networks. For all settings, we used $R = 500$, adaptive penalization and cross-validation to select the penalty parameter.

$\tau = 0.2$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	48.67%	75.82%	49.20%	73.53%
	100	58.47%	84.61%	54.93%	88.31%
	200	74.07%	91.52%	73.33%	92.21%
$\tau = 0.4$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	58.93%	61.38%	54.87%	57.46%
	100	59.26%	74.31%	54.00%	72.89%
	200	71.27%	83.38%	62.27%	84.09%
$\tau = 0.6$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	61.09%	62.08%	58.43%	56.71%
	100	65.38%	74.31%	57.16%	69.70%
	200	74.29%	83.38%	67.25%	79.04%
$\tau = 0.8$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	65.42%	73.18%	61.50%	70.42%
	100	78.26%	79.12%	73.76%	83.78%
	200	93.17%	88.94%	91.08%	88.30%

4.3 Inference results

For brevity, in this section, we will restrict our attention to four estimation techniques of the treatment effect:

1. Post-ℓ_1 estimator: refers to the estimate from Step 2 of the OS algorithm,
2. Quantile regression (QR) estimator: refers to applying ordinary QR to the model with all possible variables,
3. OS estimator,
4. DS estimator.

We have also ran experiments with the "naive" estimator but this estimator, as expected, performed worse than the post-ℓ_1 estimator. Next to that, we have also used a "weighted" version of the OS estimator but this also led to worse performance than the OS itself. Therefore, we have left these out here but they are included in similar tables in Appendix D.2.

We first consider the bias of these estimators. These can be found in Table 2 which showcases results for various values of N, T, and τ. Primarily, it is evident that in most cases, the bias is strongest for the post-ℓ_1 estimator, which aligns with expectations as this estimator is susceptible to omitted variable bias owing to selection mistakes. OS clearly improves upon the post-ℓ_1 estimator as it has smaller bias for all settings. QR and DS perform generally even better than OS, with no clear standout performer among them. The disparity between the post-ℓ_1 estimator and the other estimators seems to be most notable for larger values of T where in some instances the bias is reduced by over 90% when using QR or DS. This demonstrates the faster convergence rates of these estimators, whereas the post-ℓ_1 estimator is not root-T consistent (Belloni et al., 2019).

Furthermore, we observe a general trend across most estimators wherein bias decreases with increasing T, which is in line with expectations. However, this trend is not consistent in all cases, likely due to minimal differences in true bias and the limited number of replications in the current simulations. Additionally, it is noteworthy that there is no clear pattern between bias and N. According to the findings presented in Table 1, an increase in N correlates with decreased accuracy in network estimation. Conversely, a larger N results in a greater number of units containing information about the common parameter $\theta(\tau)$. The results suggest that neither of these opposing factors exerts clear dominance over the other.

Next, we turn to coverage of the different confidence regions of the true parameter. These are constructed using the approximate normality from (3.8). In Table 3, we report the coverage and average standard deviation of the different estimators, with the nominal coverage probability being 95% for $\tau = 0.2$. Results for the remaining quantile levels are given in Appendix D.3. Ideally, we should see the coverage to also be 95%. We have omitted the confidence regions resulting from the OS estimator in combination with standard error estimate (3.7) as this proved to be more unstable in finite sample than the confidence regions coming from DS in combination with bootstrapped standard errors. Lastly, we compute the standard errors for the post-ℓ_1 estimator by bootstrapping the post-stage quantile regression, and for QR standard errors we also use bootstrap.

We immediately observe that the post-ℓ_1 estimator experiences serious undercoverage. The coverage by QR and DS is consistently above 95%. DS is the closest to

Table 2: Average bias of post-ℓ_1, QR, OS, and DS estimators of treatment effect confounded by τ-dependent Erdos-Renyi networks. For all settings, we used $R = 500$, adaptive penalization and cross-validation to select the penalty parameter.

$\tau = 0.2$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.133	0.039	0.119	0.040	0.132	0.008	0.103	0.025
	100	0.127	0.008	0.107	0.012	0.128	0.008	0.088	0.010
	200	0.096	0.006	0.064	0.008	0.092	0.001	0.052	−0.006
$\tau = 0.4$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.103	0.039	0.101	0.040	0.082	0.010	0.069	0.010
	100	0.063	−0.037	0.050	−0.038	0.079	−0.017	0.067	−0.009
	200	0.077	−0.017	0.046	−0.012	0.076	−0.030	0.039	−0.012
$\tau = 0.6$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.108	−0.010	0.089	0.023	0.087	−0.054	0.062	−0.037
	100	0.106	0.009	0.087	0.012	0.117	0.015	0.079	0.019
	200	0.097	0.012	0.067	0.016	0.103	0.012	0.069	0.018
$\tau = 0.8$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.127	−0.006	0.115	0.020	0.114	−0.025	0.086	0.010
	100	0.087	−0.002	0.059	0.010	0.112	−0.004	0.083	0.012
	200	0.039	0.001	0.023	0.001	0.045	0.002	0.025	0.002

95% in all settings, whereas using QR seems to be rather conservative. From the standard deviation used for the confidence intervals, we observe that the length of the DS intervals is also always smaller. This demonstrates that both QR and DS lead to valid inference for this model where DS is more efficient. Another benefit of the DS estimator is that it can also be used in cases where $N > T$, whereas the QR estimator is infeasible here.

Table 3: Coverage and average standard deviation of confidence regions resulting from the post-ℓ_1, QR, and DS estimators of the treatment effect confounded by τ-dependent Erdos-Renyi networks under nominal coverage probability of 95%. For all settings, we used $\tau = 0.2$, $R = 500$, adaptive penalization and cross-validation to select the penalty parameter.

Design		Coverage			SD		
N	T	Post	QR	DS	Post	QR	DS
5	50	83.7%	98.9%	97.1%	0.232	0.387	0.330
5	100	76.0%	97.4%	95.4%	0.155	0.219	0.202
5	200	74.0%	96.6%	96.2%	0.112	0.142	0.138
10	50	81.4%	99.7%	97.4%	0.172	0.398	0.270
10	100	72.3%	99.8%	96.6%	0.108	0.197	0.149
10	200	69.7%	96.5%	95.7%	0.078	0.108	0.097

5 Empirical example: interactions of sovereign credit risk

In this section, we estimate the international connection of sovereign credit risk. This type of risk has received extensive attention especially since the breakout of a wide range of financial crises when government debts of many countries have increased dramatically. The large and rapidly increasing size of the sovereign debt markets also urges a good understanding of the nature of sovereign credit risk. We focus on the sovereign credit default swap (CDS) spreads as a proxy of sovereign credit risk. A CDS contract is an insurance contract that protects the buyer from the credit event, and its spread is the insurance premium that buyers have to pay, and thus reflects the credit risk. Existing studies that examine the determinants of sovereign credit risk primarily focus on the country-specific type of risk and global macroeconomic forces, but ignore the international interactions of the risk.

Cross-country spillovers of risks are expected to play an important role in the sovereign credit risks due to deep integration and globalization. The economic and financial performance of a country not only affects her own sovereign credit risk, but also the risk of other countries that are economically and politically tied. This suggests that geographic neighbors or a single economic index may not be sufficient to capture the cross-country spillovers of risks. The network structure is likely to be driven by a set of observables and unobservables, and needs to be recovered with some

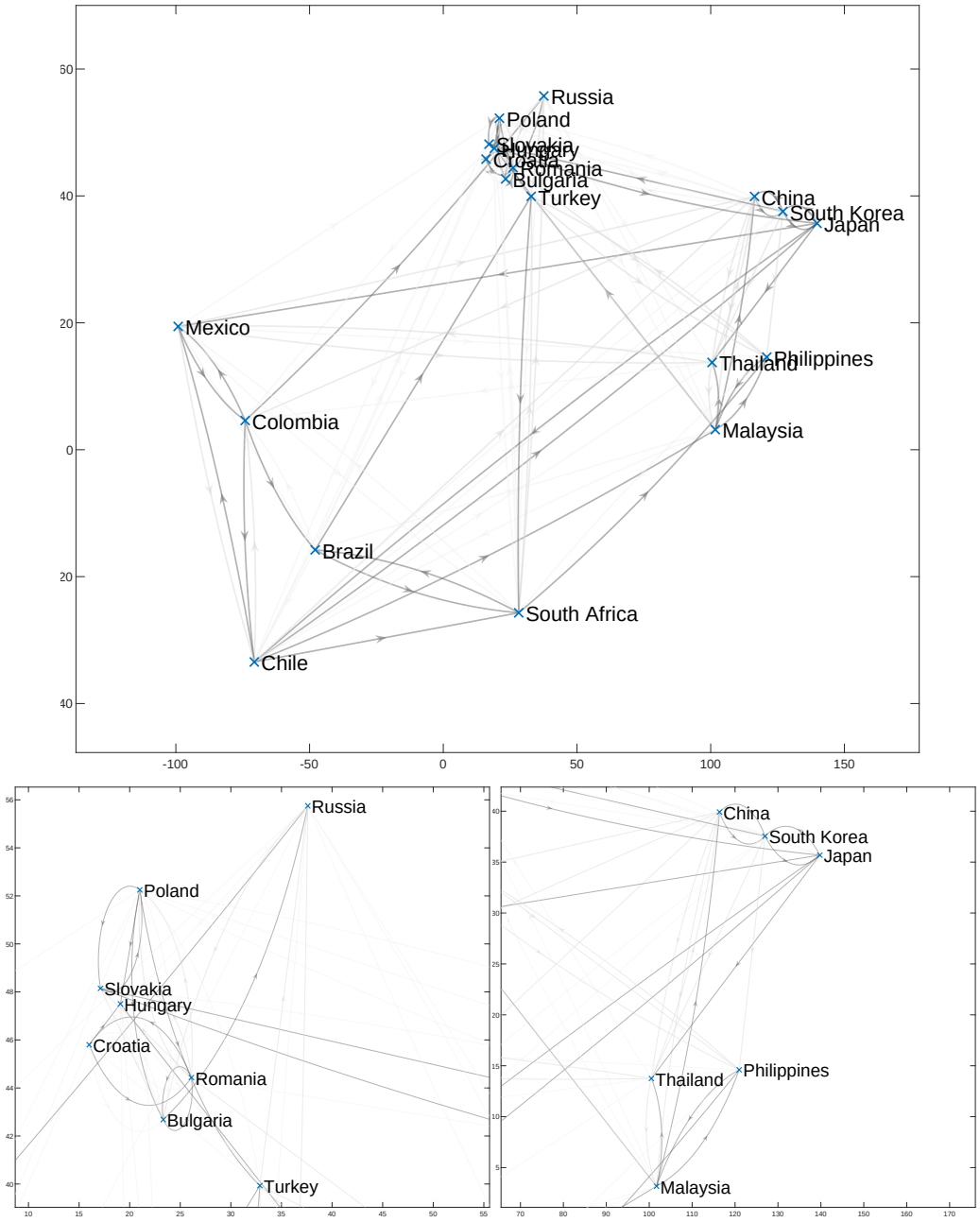
data-driven methods. Ignoring the interaction (if there is any) in the sovereign credit risk regression, as existing studies in this literature, leads to biased estimates of the effects of determinants, even if the global determinants are controlled. Identifying the network structure is also of great interest and importance in practice as it provides policy makers and investors with extra information from other sovereigns to better control the risk in the own sovereign, diversify the investment, and even possibly use this information as an early warning to avoid future default.

The interactions between countries may also evolve over time. The network structure in tranquil periods is not necessarily identical to that in the turmoil periods given that many studies have found evidence of financial contagion or stronger spillovers during financial crisis or recessions. Even if the network structure remains stable, the effects of determinants are also likely to vary across regimes (Qian et al., 2017; Dieckmann and Plank, 2012) and across countries (Longstaff et al., 2011). Thus motivated, we employ the network quantile regression model to identify the possibly quantile-specific network structure and effects of determinants as

$$Q_{it}\left(\tau\right)=\sum_{j=1}^{N}W_{ij}\left(\tau\right)x_{jt}+\alpha_{i}\left(\tau\right), \tag{5.1}$$

where y_{it} is the CDS spread for country i at time t and $Q_{it}(\tau)$ is the associated conditional quantile at τ. We follow Longstaff et al. (2011) to focus on spreads of five-year sovereign credit default swaps, and focus on the effects of domestic macroeconomic variables on the the CDS spreads, namely the local stock market returns. We also include control variables of the U.S. stock market returns, treasury yields, high-yield corporate bond spreads, equity premium, volatility risk premium, equity flows, and bond flows. We use an extended data set of Wang et al. (2019) that contains 19 countries, i.e. Brazil, Bulgaria, Chile, China, Colombia, Croatia, Hungary, Japan, Korea, Malaysia, Mexico, Philippines, Poland, Romania, Slovak, South Africa, Thailand, and Turkey. We use the monthly data starting from January 2003 to January 2016 resulting in 156 time observations. We apply the proposed network estimation method to estimate (5.1). Figure 1 presents the estimated network of the country CDS spreads at $\tau=0.5$. We also include the estimates at $\tau=0.9$ in Appendix E.

Figure 1: Estimated network structure at $\tau = 0.5$



Notes: The darkness of an edge represents the strength of the corresponding link: the darker the stronger. The global overview is in the top panel whereas details on eastern European and Asian countries are in, respectively, the bottom left and right panels.

Several important results emerge from the analysis. First, we find that some geographic pattern remain salient in the network structure. The connections among Eastern European countries are particularly strong, while the Asian and Latin American countries are also interconnected within the continent. For example, in Eastern Europe, the CDS spreads of Bulgaria are highly affected by the spreads and economic performance of Romania and Poland. Croatia is most influenced by Romania as well. In Asia, Malaysia is strongly affected by the Philippines, while Thailand is most influenced by Malaysia and Japan. In Latin America, Brazil, Colombia, and Mexico are all interconnected. These finding are well in line with the strong economic ties between these countries.

Second, we also find that some strong connections are not captured by the geographic locations. For example, the CDS spreads of Chile are affected not only by Latin American countries but also by Japan, and similarly Japan is also most affected by Chile (and other Asian countries). In fact, although Chile is geographically remote from Japan, their economic relations have been strong and persistent as a result of their agreement for a Strategic Economic Partnership. This results in frequent economic activities including a large amount of import and export between the two countries. Hence, it is not surprising that the CDS spreads of Chile got affected by those of Japan when the gross government debt to GDP ratio of Japan kept increasing and remained the highest over years around the world. Another example is Turkey, which is remote from Brazil and South Africa geographically. But the CDS spreads of Turkey is largely influenced by the risk level and economic status of Brazil while it affects South Africa strongly. Further examination reveals that these countries indeed have strong economic ties with Turkey. Brazil is the first strategic partner and biggest trade partner of Turkey in South America, and South Africa is the leading trade partner of Turkey in the Sub-Saharan Africa, and the trade with Turkey generates more than 30 percent of the GDP of all Sub-Saharan African countries.

Third, we find that the spillover effects and network structure are not always symmetric. Although Turkey is highly affected by Brazil, the spillover from Turkey to Brazil is rather limited. Similarly, Russia is largely affected by Bulgaria and China is largely affected by Thailand, but these two connections are both directional. Interestingly, the source countries that generate spillovers in such asymmetric relations are typically characterized by less developed and unstable economies and financial markets, and thus the asymmetry in the volatility of economy may be a possible explanation of directional links.

Finally, the network seems to vary across quantiles. For instance at $\tau = 0.9$ (see Figure 2 in Appendix E) China is strongly connected with many economies such as

Russia, Poland, Thailand, etc. whereas at $\tau = 0.5$ many of these connections are only weak. This implies that at different quantiles of the CDS spreads, the international interactions may exhibit different behaviors.

This demonstrates that sovereign credit risk is highly connected across different nations. These network estimates are crucial to include in further policy / treatment analysis.

6 Concluding remarks

We study the estimation of unobserved networks, representing socioeconomic or spatial interactions, from panel data sets. Such a situation arises naturally in many empirical studies. We consider a quantile model in which outcome distributions of individuals are affected by peer characteristics, in addition to their own, through an unobserved network. Our approach allows the network to change across the quantiles of the data distributions and, therefore, estimates the network for each quantile of interest. The network structures are largely unrestricted, except that sparsity assumptions are adopted to improve the efficiency of the estimated network. In addition, we propose estimators to conduct valid post-selection inference on other policy/treatment variables in the model.

Simulations studies are carried out to demonstrate the performance of our approaches. These demonstrate that our method is well able to capture the underlying network links with the performance increasing with the number of time periods. Nonetheless, moderate selection mistakes persist, carrying over to bias and invalid inference when employing (post-)ℓ_1-penalized estimation techniques. Our estimators, which are based on a Neyman orthogonal moment condition, demonstrate reduced susceptibility resulting from such model selection mistakes. In particular, we achieve a bias reduction of up to 90% compared to conventional estimation techniques. Moreover, we show that our estimator leads to valid inference while being more efficient when comparing with ordinary quantile regression. Moreover, our empirical study illustrates that interactions among sovereign credit risks cannot be solely attributed to geographical factors. It also reveals the varying nature of these interactions across different quantiles.

Several issues are not addressed during our analysis and, therefore, may require further studies. The most important issue is perhaps to allow for the endogenous effect where peer outcomes may also affect each other through the unobserved network. We briefly looked into this issue and find that the inclusion of such an effect would induce the endogeneity. While it could be addressed relatively easily in a conditional mean model, there has not been much much investigation for the quantile

regression. This is partly due to the well-known complication that a quantile function does not possess a linearity and that the monotonicity of the quantile function must be honored. This renders many well-known treatments under a conditional mean model infeasible in this context. In addition, we do not allow for temporal dependency in the data while it could be of particular interest in many applications. The incorporation of temporal dependency should be generally possible by including lagged variables into the model, with the model selection procedure adapted to this. In addition, time-specific effects are also not allowed in our analysis. The reason is that the asymptotic properties of a quantile fixed-effect model with both individual effects, as we include, and time effects is not well-studied. Based on the existing literatures, our conjecture about this is that estimators of such models would be inconsistent so that certain bias correction techniques shall be invoked. During our analysis, we only consider the adaptive LASSO as a vehicle for regularization. However, the literatures on regularization approaches is expanding rapidly; and thus, other tools, such as adaptive elastic net, SCAD, etc., are also worth considering. Similarly, other model selections techniques beyond cross validation may also be of interest.

References

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*, 40–79.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*, 2369–2429.

Belloni, A. and V. Chernozhukov (2011). ℓ_1-penalized quantile regression for high dimensional sparse models. *The Annals of Statistics 39*, 82–130.

Belloni, A., V. Chernozhukov and C. Hansen (2013). Inference for high-dimensional sparse econometric models. *Advances in Economics and Econometrics: The 2010 World Congress of the Econometric Society 3*, 245–295.

Belloni, A., V. Chernozhukov and C. Hansen (2014). Inference on treatment effects after selection amongst high-dimensional controls. *The Review of Economic Studies 81*, 608–650.

Belloni, A., V. Chernozhukov, C. Hansen, and D. Kozbur (2016). Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics 34*, 590–605.

Belloni, A., V. Chernozhukov, and K. Kato. (2019). Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *Journal of the American Statistical Association 114*, 749–758.

Bondell, H. D., B. J. Reich, and H. Wang (2010). Noncrossing quantile regression curve estimation. *Biometrika 97*(4), 825–838.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*, 41–55.

Bursztyn, L., F. Ederer, B. Ferman, and N. Yuchtman (2014). Understanding mechanisms underlying peer effects: evidence from a field experiment on financial decisions. *Econometrica 82*(4), 1273–1301.

Cai, J., A. De Janvry, and E. Sadoulet (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics 7*(2), 81–108.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica 81*, 855–882.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*, C1-C68.

Ciccarelli, C. and J. P. Elhorst (2018). A dynamic spatial econometric diffusion model with common factors: The rise and spread of cigarette consumption in italy. *Regional Science and Urban Economics 72*, 131–142.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

De Giorgi, G. and M. Pellizzari (2014). Understanding social interactions: Evidence from the classroom. *The Economic Journal 124*(579), 917–953.

de Paula, A. (2017). Econometrics of network models. In B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*, pp. 268–323. Cambridge: Cambridge University Press.

de Paula, A., I. Rasul, and P. Souza (2018). Recovering social networks from panel data: identification, simulations and an application. *Working paper*.

De Weerdt, J. and S. Dercon (2006). Risk-sharing networks and insurance against illness. *Journal of Development Economics 81*(2), 337–356.

Dieckmann, S. and T. Plank (2012). Default risk of advanced economies: An empirical analysis of credit default swaps during the financial crisis. *Review of Finance 16*, 903–934.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics 2*, 267–277.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *The Review of Economics and Statistics 101*(5), 763—776.

Fernández-Val, I. and M. Weidner (2018). Fixed effects estimation of large-t panel data models. *Annual Review of Economics 10*, 109–138.

Forbes, K. J. and R. Rigobon (2002). No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance 57*(5), 2223–2261.

Galvao, A., J. Gu, and S. Volgushev (2020). On the unbiased asymptotic normality of quantile regression with fixed effects. *Journal of Econometrics 218*, 178–215.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*, 253–264.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Graham, B. S., J. Hahn, A. Poirier, and J. L. Powell (2015). Quantile regression with panel data. *Working paper*.

Haag, M. and R. Lagunoff (2006). Social norms, local interaction, and neighborhood planning. *International Economic Review 47*(1), 265–296.

Han, X., C.-S. Hsieh, and S. I. Ko (2019). Spatial modeling approach for dynamic network formation and interactions. *Journal of Business & Economic Statistics 39*, 120–135.

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

Ho, C.-Y., W. Wang, and J. Yu (2018). International knowledge spillover through trade: A time-varying spatial panel data approach. *Economics Letters 162*, 30 – 33.

Hsieh, C.-S. and L. F. Lee (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics 31*, 301–319.

Hsieh, C.-S., L.-F. Lee, and V. Boucher (2020). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Quantitative Economics 11*(4), 1349–1390.

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review 102*(5), 1857–97.

Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. *Journal of Economic Literature 55*(1), 49–95.

Johnsson, I. and R. Moon (2021). Estimation of peer effects in endogenous social networks: Control function approach. *Review of Economics and Statistics 103*(2), 329–345.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009). Trust and social collateral. *The Quarterly Journal of Economics 124*(3), 1307–1361.

Kato, K., A.F. Galvao Jr, and G.V. Montes-Rojas (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics 170*, 76-91.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. (2023). quantreg: Quantile Regression. R package version 5.97. Available at http://CRAN.R-project.org/package=quantreg

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46,* 33–50.

Koenker, R. and Z. Xiao (2002). Inference on the quantile regression process. *Econometrica 70*(4), 1583–1612.

König, M. D., X. Liu, and Y. Zenou (2019). R&d networks: Theory, empirics, and policy implications. *The Review of Economics and Statistics 101*(3), 476–491.

Lee, E. R., H. Noh, and B. U. Park (2014). Model selection via bayesian information criterion for quantile regression models. *Journal of the American Statistical Association 109*(505), 216–229.

Lee, L.-f. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 140*(2), 333–374.

Lee, L.-f., X. Liu, and X. Lin (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal 13*(2), 145–176.

Leeb, H. and B. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*, 21-59.

Leider, S., M. M. Möbius, T. Rosenblat, and Q.-A. Do (2009). Directed altruism and enforced reciprocity in social networks. *The Quarterly Journal of Economics 124*(4), 1815–1851.

Lin, X. and L.-f. Lee (2010). Gmm estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics 157*, 34–52.

Longstaff, F. A., J. Pan, L. H. Pedersen, and K. J. Singleton (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics 3*, 75–103.

Malm, A. and G. Bichler (2011). Networks of collaborating criminals: Assessing the structural vulnerability of drug markets. *Journal of Research in Crime and Delinquency 48*(2), 271–297.

Manresa, E. (2016). Estimating the structure of social interactions using panel data. *Working paper*.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*(1), 415–444.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*(1), 1–32.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association 70*, 120–126.

Qian, Z., W. Wang, and K. Ji (2017). Sovereign credit risk, macroeconomic dynamics, and financial contagion: Evidence from Japan. *Macroeconomic Dynamics 21*, 2096–2120.

Qu, X. and L.-f. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics 184*, 209–232.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Semenova, V., M. Goldman, V. Chernozhukov, and M. Taddy (2023). Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence. *Quantitative Economics 14*, 471-510.

Su, L. and Z. Yang (2007). Instrumental variable quantile estimation of spatial autoregressive models. *Working paper*.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies 68*(2), 261–295.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42,* 1166–1202.

Wang, W., X. Zhang, and R. Paap (2019). To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions? *Journal of Applied Econometrics 34*, 724–745.

Zhang, C. and S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society, Series B 76*, 217–242.

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

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*, 1418–1429.

A Neyman orthogonality

Let $w_{it} = \{y_{it}, x_t, z_{it}\}$ denote the data relevant for unit i at time t. Let us define the following moment function

$$m(w_{it}, W, \alpha, W^c, \theta) = \left(\tau - \mathbb{1}\left\{y_{it} \leq \sum_{j=1}^{N} W_{ij} x_{jt} + \theta z_{it} + \alpha_i\right\}\right)\left(z_{it} - \sum_{j=1}^{N} W^c_{ij} x_{jt}\right)$$

such that the moment condition in (3.6) becomes

$$\mathbb{E}\left[m(w_{it}, W(\tau), \alpha(\tau), W_0^c, \theta(\tau))\right] = 0.$$

The Neyman orthogonality condition now boils down to the following conditions on the derivatives w.r.t. the nuisance parameters W, α, and W^c,

$$\begin{aligned}
&\partial_W \mathbb{E}\left[m(w_{it}, W, \alpha(\tau), W_0^c, \theta(\tau))\right]\big|_{W=W(\tau)} = \mathbf{0},\\
&\partial_\alpha \mathbb{E}\left[m(w_{it}, W(\tau), \alpha, W_0^c, \theta(\tau))\right]\big|_{\alpha=\alpha(\tau)} = \mathbf{0},\\
&\partial_{W^c} \mathbb{E}\left[m(w_{it}, W(\tau), \alpha(\tau), W^c, \theta(\tau))\right]\big|_{W^c=W_0^c} = \mathbf{0}.
\end{aligned}$$

Let us verify the first condition. Note that for observation w_{it}, the derivative w.r.t. W_j where $j \neq i$ always vanishes. Therefore, we only need to consider the derivative w.r.t. W_i. Furthermore, let $F_{y_{it}}(\cdot|x_t, z_{it})$ denote the cdf of y_{it} conditional on x_t and z_{it}. Using the law of iterated expectations, we then obtain that

$$\begin{aligned}
\partial_{W_i} \mathbb{E}\left[m(w_{it}, W, \alpha(\tau), W_0^c, \theta(\tau))\right]\big|_{W=W(\tau)} &= \mathbf{0}\\
\partial_{W_i} \mathbb{E}\left[\left(\tau - \mathbb{1}\left\{y_{it} \leq \sum_{j=1}^{N} W_{ij} x_{jt} + \theta(\tau) z_{it} + \alpha_i(\tau)\right\}\right) v_{it}\right]\Bigg|_{W=W(\tau)} &= \mathbf{0}\\
\partial_{W_i} \mathbb{E}\left[\left(\tau - F_{y_{it}}\left(Q_{it}(\tau) + \sum_{j=1}^{N}\left(W_{ij} - W_{ij}(\tau)\right) x_{jt} \middle| x_t, z_{it}\right)\right) v_{it}\right]\Bigg|_{W=W(\tau)} &= \mathbf{0}\\
\mathbb{E}\left[-f_{y_{it}}\left(Q_{it}(\tau) + \sum_{j=1}^{N}\left(W_{ij} - W_{ij}(\tau)\right) x_{jt} \middle| x_t, z_{it}\right) x_t v_{it}\right]\Bigg|_{W=W(\tau)} &= \mathbf{0}
\end{aligned}$$

$$\mathbb{E}\left[f_{it}x_t v_{it}\right] = \mathbf{0},$$

which gives the condition mentioned in Section 3.3. The derivatives and corresponding conditions w.r.t. other nuisance parameters can be verified in a similar way.

B Additonal estimators

As has been motivated by Belloni et al. (2019), using a weighted version of the auxiliary regression (4.1) could possibly reduce the asymptotic variance of the treatment effect estimator. Recall that $f_{it} = f_{y_{it}}(Q_{it}(\tau)|x_t, z_{it})$ denotes the conditional density of y_{it} conditional on x_t, z_{it} evaluated at the conditional function $Q_{it}(\tau)$. We now focus on the following model

$$\begin{aligned} f_{it}z_{it} &= f_{it}\sum_{j=1}^{N} W^c_{0,ij}x_{jt} + v_{it}, \\ y_{it} &= \sum_{j=1}^{N} W_{ij}(U_{it})x_{jt} + z_{it}\theta(U_{it}) + \alpha_i(U_{it}). \end{aligned} \tag{B.1}$$

For estimation of f_{it}, we can use that $F_{y_{it}}(Q_{y_{it}}(\tau|x_t, z_{it})) = \tau$ implies that

$$\frac{1}{f_{it}} = \frac{\partial Q_{y_{it}}(\tau|x_t, z_{it})}{\partial \tau}.$$

Therefore, Belloni et al. (2019) propose the estimator

$$\widehat{f}_{it} = \frac{2h}{\widehat{Q}_{y_{it}}(\tau + h|x_t, z_{it}) - \widehat{Q}_{y_{it}}(\tau - h|x_t, z_{it})}, \tag{B.2}$$

where $h = h_n \to 0$ and denotes a bandwidth parameter and $\widehat{Q}_{y_{it}}(u|x_t, z_{it})$ denotes an estimate of the conditional u-quantile $Q_{y_{it}}(u|x_t, z_{it})$ of observation (i, t) based on ℓ_1-penalized quantile regression. For the bandwidth parameter, it is suggested to use $h := \{n^{-1/6}, \tau(1-\tau)/2\}$. Incorporating these into the orthogonal score algoritm leads to the following weighted algorithm

Weighted orthogonal score algorithm

Step 1: *Estimate* $(\widehat{\theta}(\tau), \widehat{W}(\tau), \widehat{\alpha}(\tau))$ *from post-*ℓ_1*-penalized quantile regression of* y_{it} *on* z_{it} *and* x_t *described in Section 3.2.*

Step 2: *Compute* $(\tilde{\theta}(\tau), \tilde{W}(\tau), \tilde{\alpha}(\tau))$ *from quantile regression of* y_{it} *on the treatment and the regressors selected in Step 1.*

Step 3: *Estimate the conditional density* $\widehat{f}$ *via* (B.2).

Step 4: *Compute* $\tilde{W}^c$ *from the post-Lasso estimator of* $\widehat{f}_{it} z_{it}$ *on* $\widehat{f}_t x_t$.

Step 5: *Construct the weighted score function*

$$\widehat{\psi}_{it}(\theta) = \left(\tau - \mathbb{1}\left\{y_{it} \leq z_{it}\theta + \sum_{j=1}^{N} \tilde{W}_{ij}(\tau) x_{jt} + \tilde{\alpha}_i(\tau)\right\}\right) \widehat{f}_{it} \left(z_{it} - \sum_{j=1}^{N} \tilde{W}^c_{ij}(\tau) x_{jt}\right).$$

Step 6: *Using the Neyman-type score statistic*

$$L_{nT}(\theta) = \frac{|\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{\psi}_{it}(\theta)|^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{\psi}^2_{it}(\theta)},$$

set $\check{W}(\tau) = \tilde{W}(\tau)$, $\check{\alpha}(\tau) = \tilde{\alpha}(\tau)$ *and* $\check{\theta}^{OS}(\tau) \in \arg\min_{\theta \in \Theta_\tau} L_{nT}(\theta)$.

Furthermore, these can also be incorporated in the Double Selection algorithm as summarized below

Weighted Double Selection algorithm

Step 1: Estimate $(\widehat{\theta}(\tau), \widehat{W}(\tau), \widehat{\alpha}(\tau))$ from ℓ_1-penalized quantile regression of y_{it} on z_{it} and x_t.

Step 2: Estimate the conditional density $\widehat{f}$ via (B.2).

Step 3: Compute $\tilde{W}^c$ from the Lasso estimator of $\widehat{f}_{it} z_{it}$ on $\widehat{f}_t x_t$.

Step 4: Compute $(\check{\theta}^{DS}(\tau), \check{W}(\tau), \check{\alpha}(\tau))$ from quantile regression of $\widehat{f}_{it} y_{it}$ on $\widehat{f}_{it} z_{it}$ and $\{\widehat{f}_{jt} x_{jt} : x_{jt}$ is selected in either Step 1 or Step 2$\}$.

C Location-scale shift model

Instead of using the random coefficient representation, we now consider a DGP that falls into the class of location-scale shift models. This is a very common setup setup

in the quantile regression literature (Koenker, 2005; Koenker and Xiao, 2002). In particular, we focus on the following DGP

$$\begin{aligned} z_{it} &= \sum_{j=1}^{N} W_{ij}^{c} x_{jt} + v_{it}, \\ y_{it} &= \eta_i + x_{it} + z_{it} + \left(1 + \tilde{W}_{ii} x_{it} + \tilde{W}_{ij} \mathbb{I}_{\{\epsilon_{it} \geq z_\gamma\}} x_{jt} + z_{it}\tilde{\theta}\right) \epsilon_{it}. \end{aligned} \tag{C.1}$$

First of all, note that the first equation (the confounding equation) in (C.1) is the same as in (4.1). The second equation shows the location-scale shift model. Here, η_i denotes a random effect of unit i, $\tilde{W}$ denotes the $N \times N$ matrix containing spillover parameters, $\tilde{\theta}$ is a treatment effect parameter, and $\epsilon_{it} \sim F$ where z_γ denotes the γ-th quantile of F. Note that the random variable $\mathbb{I}_{\{\epsilon_{it} \geq z_\gamma\}} \epsilon_{it}$ follows a mixture distribution with γ probability mass at zero and a continuous Gaussian tail past z_γ with total mass $1 - \gamma$. This ensures that the effect of x_{jt} on y_{it} is zero for $\epsilon_{it} \leq z_\gamma$ and nonzero for $\epsilon_{it} \geq z_\gamma$. Let $M = \mathbb{I}_{\{\epsilon_{it} \geq z_\gamma\}} \epsilon_{it}$ and let F_M denote its cdf, we observe that

$$\begin{aligned} F_M^{-1}(\tau) &= 0 && \text{for } \tau \leq \gamma, \\ F_M^{-1}(\tau) &= F^{-1}(\tau) && \text{for } \tau \geq \gamma. \end{aligned}$$

One then easily verifies that for the linear QR model

$$Q_{y_{it}}(\tau | x_t, z_{it}) = \alpha_i(\tau) + \sum_{j=1}^{N} W_{ij}(\tau) x_{jt} + \theta(\tau) z_{it},$$

we get that

$$\begin{aligned} \alpha_i(\tau) &= \eta_i + F^{-1}(\tau) \\ W_{ii}(\tau) &= 1 + \tilde{W}_{ii} F^{-1}(\tau) \\ W_{ij}(\tau) &= \tilde{W}_{ij} F_M^{-1}(\tau) \\ W_{ik}(\tau) &= 0 \text{ for } k \notin \{i, j\} \\ \theta(\tau) &= 1 + \tilde{\theta} F^{-1}(\tau). \end{aligned}$$

The most important part of this DGP is how we choose $W(\tau)$. For now, we have chosen $W(\tau)$ so that it is a continuous function of τ. For the diagonal elements this holds if F^{-1} is continuous. We set $F = \mathcal{N}(\mu, 1/2)$ where μ is chosen to guarantee continuity of the non-diagonal elements of W. In particular, we set it so that $F^{-1}(\gamma) = 0$. This ensures the continuity of $F_m^{-1}(\tau)$ and thereby the continuity of

$W_{ij}(\tau)$. For the rest, the regressors and treatment are generated in the same way as in Section 4. Next to that, the non-zero entries of $\tilde{W}$ are chosen using ER (similar to Section 4). All nonzero elements in $\tilde{W}$ are set equal to 1. Lastly, we have that $\eta_i \sim U(0,1)$ and $\tilde{\theta} = 1$.

The results of the network estimation can be found in Table 4. An important difference with Table 1 is that we now see better performance around the middle quantiles. This is anticipated as there is most information here and for this particular network there are no link switches around the median, as was the case with Table 1. Furthermore, note that around the $\tau = 0.8$, the effect of the non-diagonal elements is still smaller than 0.1, so we classify these elements as zeros. We observe that these effects do not strongly deteriorate performance. The performance is slightly worse as in the middle quantiles but this is expected due to the loss of information. On the other side, we do better than in the $\tau = 0.2$ quantile which is due to the fact the effects are stronger in the upper tails and therefore easier to distinguish from true zeros.

Moreover, in Table 5 we show the average bias of the post-ℓ_1, QR, OS and DS estimators. Here, the conclusions are largely the same as in Table 2. The post-ℓ_1 has the most bias compared to the other three. For the other three, there is no single estimator with the smallest bias in general, although QR and DS seem to do slightly better than OS. The coverage corresponding to the post-ℓ_1, QR and DS can be found in Table 6. Here, we see the same patterns as before in Table 3.

Table 4: Average percentages of correct nonzero and zero links of τ-dependent location-scale shift networks. Here, a zero correspond to a link with strength below 0.1 in the true network, which has been estimated to be zero. For all settings, we used $R = 200$, adaptive penalization and cross-validation to select the penalty parameter.

$\tau = 0.2$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	62.40%	57.90%	62.60%	56.38%
	100	63.80%	73.35%	64.00%	74.36%
	200	66.20%	85.20%	70.90%	86.57%
$\tau = 0.4$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	73.60%	53.75%	73.80%	53.93%
	100	80.80%	69.90%	80.40%	72.00%
	200	92.00%	82.25%	93.00%	83.08%
$\tau = 0.6$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	77.40%	53.25%	78.40%	52.28%
	100	84.60%	67.55%	83.70%	68.73%
	200	93.00%	80.15%	95.10%	82.10%
$\tau = 0.8$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	73.60%	49.65%	71.20%	50.92%
	100	77.40%	65.55%	77.70%	67.42%
	200	88.80%	77.78%	87.00%	79.53%

Table 5: Average bias of post-ℓ_1, QR, OS, and DS estimators of treatment effect confounded by τ-dependent location-scale shift networks. For all settings, we used $R = 200$, adaptive penalization and cross-validation to select the penalty parameter.

$\tau = 0.2$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.089	0.039	0.088	0.032	0.086	0.013	0.070	0.039
	100	0.057	−0.026	0.024	−0.025	0.085	0.029	0.067	0.033
	200	0.078	−0.038	0.044	0.041	0.075	0.018	0.043	0.021
$\tau = 0.4$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.063	0.028	0.060	0.026	0.040	0.009	0.050	0.013
	100	0.042	−0.006	0.028	0.013	0.053	0.022	0.043	0.018
	200	0.054	0.026	0.027	0.024	0.013	−0.001	−0.006	−0.001
$\tau = 0.6$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	−0.044	−0.044	−0.019	−0.031	0.035	−0.040	0.018	−0.026
	100	0.023	−0.012	0.001	−0.011	0.045	0.005	0.037	0.023
	200	0.028	0.010	0.017	0.009	0.004	−0.020	−0.007	0.001
$\tau = 0.8$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	−0.034	−0.033	−0.024	−0.032	0.024	−0.036	0.018	−0.005
	100	0.043	−0.038	0.016	−0.030	0.008	−0.072	0.006	−0.023
	200	0.020	−0.025	0.011	−0.019	0.025	−0.036	0.003	−0.009

Table 6: Coverage and average standard deviation of confidence regions resulting from the post-ℓ_1, QR, and DS estimators of the treatment effect confounded by τ-dependent location-scale shift networks under nominal coverage probability of 95%. For all settings, we used $\tau = 0.2$, $R = 200$, adaptive penalization and cross-validation to select the penalty parameter.

Design		Coverage			SD		
N	T	Post	QR	DS	Post	QR	DS
5	50	89.0%	96.0%	96.0%	0.241	0.347	0.316
5	100	91.0%	95.0%	95.0%	0.165	0.225	0.209
5	200	77.0%	95.0%	93.0%	0.115	0.153	0.145
10	50	88.0%	99.0%	93.0%	0.167	0.278	0.228
10	100	83.0%	95.0%	96.0%	0.110	0.170	0.149
10	200	76.0%	98.0%	96.0%	0.080	0.112	0.102

D Tables

D.1 Weaker confounding

For the confounding matrix W_{ij}^c we now set all diagonals elements equal to 0.3 and the non-diagonal element are generated using ER (so only one off-diagonal element in every row is nonzero) and set to 0.2. Notice that this leads to weaker correlation between the treatment and the other regressors. The results from estimating the network matrix can be found in Table 7. Furthermore, the bias of the estimators can be found in 8. Lastly, the coverage of the resulting confidence regions and the corresponding standard deviation are reported in Table 9. Generally, we see that the ℓ_1-penalized estimation techniques perform better compared to Section 3.2. Furthermore, the benefit of using OS and DS estimators that take into account the confounding effect of the other regressors on the treatment reduces, as these confounding effects are smaller.

Table 7: Average percentages of correct nonzero and zero links of τ-dependent Erdos-Renyi networks. For all settings, we used $R = 1000$.

$\tau = 0.2$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	61.14%	80.64%	56.41%	78.65%
	100	75.21%	88.84%	71.02%	90.60%
	200	90.66%	93.92%	90.76%	94.13%
$\tau = 0.4$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	62.51%	64.39%	58.48%	61.33%
	100	68.27%	76.56%	60.19%	74.54%
	200	82.77%	85.37%	72.13%	85.48%
$\tau = 0.6$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	66.95%	62.90%	59.46%	60.23%
	100	71.63%	74.52%	63.20%	72.19%
	200	83.35%	82.77%	74.21%	81.58%
$\tau = 0.8$		$N = 5$		$N = 10$	
	T	Nonzero	Zero	Nonzero	Zero
	50	73.91%	75.17%	65.83%	73.55%
	100	89.22%	84.61%	83.87%	84.42%
	200	97.40%	91.09%	97.11%	90.12%

Table 8: Average bias of post-ℓ_1, QR, OS, and DS estimators of treatment effect confounded by τ-dependent Erdos-Renyi networks. For all settings, we used $R = 500$, adaptive penalization and cross-validation to select the penalty parameter.

$\tau = 0.2$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.076	0.017	0.028	0.025	0.074	0.026	0.021	0.022
	100	0.047	0.008	0.010	0.011	0.046	0.010	0.010	0.001
	200	0.016	−0.002	0.003	0.000	0.019	0.006	0.004	−0.002
$\tau = 0.4$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.030	−0.020	−0.023	−0.022	0.029	−0.015	−0.010	−0.008
	100	0.021	−0.018	−0.020	−0.025	0.012	−0.013	−0.018	−0.022
	200	0.009	−0.009	−0.007	−0.006	0.012	−0.018	−0.017	−0.017
$\tau = 0.6$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.048	0.018	−0.008	−0.005	0.034	−0.020	0.012	0.009
	100	0.037	0.014	0.004	0.007	0.048	0.016	0.014	0.017
	200	0.011	−0.006	−0.003	0.001	0.051	0.015	0.013	0.014
$\tau = 0.8$		$N = 5$				$N = 10$			
	T	Post	QR	OS	DS	Post	QR	OS	DS
	50	0.038	−0.008	0.005	−0.003	0.039	−0.028	0.005	−0.004
	100	0.019	−0.001	0.001	0.005	0.021	−0.009	−0.005	−0.003
	200	0.010	0.001	0.002	0.004	0.007	0.001	0.000	0.001

Table 9: Coverage and average standard deviation of confidence regions resulting from the post-ℓ_1, QR, and DS estimators of the treatment effect confounded by τ-dependent Erdos-Renyi networks under nominal coverage probability of 95%. For all settings, we used $\tau = 0.2$, $R = 500$, adaptive penalization and cross-validation to select the penalty parameter.

Design		Coverage			SD		
N	T	Post	QR	DS	Post	QR	DS
5	50	90.5%	99.0%	96.0%	0.221	0.311	0.256
5	100	92.0%	95.0%	95.0%	0.145	0.169	0.158
5	200	92.0%	95.5%	95.5%	0.101	0.109	0.108
10	50	92.0%	100%	98.0%	0.170	0.341	0.215
10	100	90.5%	99.5%	98.5%	0.102	0.157	0.119
10	200	93.5%	97.0%	95.0%	0.071	0.083	0.076

D.2 Naive and weighted OS estimator results

In Table 10, the average bias of the the naive, post-ℓ_1, OS, weighted OS, and DS estimators can be found. All settings are the same as in Section 4. The naive estimator is described in Section 3.3. Moreover, the weighted OS algorithm is given in Appendix B. We note that the naive estimator possesses more bias than the post-ℓ_1 estimator, which is expected due to the regularization bias in the naive estimator. Next to that, the weighted OS seems to be rather unstable and perform worse than the OS estimator. We expect this to happen due to the estimation noise in the weights $\widehat{f}_{it}$.

Table 10: Average bias of the naive, post-ℓ_1, OS, weighted OS (OSw), and DS estimators of treatment effect confounded by τ-dependent Erdos-Renyi networks. For all settings, we used $R = 500$, adaptive penalization and cross-validation to select the penalty parameter.

$\tau = 0.2$				$N = 5$					$N = 10$		
	T	Naive	Post	OS	OSw	DS	Naive	Post	OS	OSw	DS
	50	0.223	0.133	0.119	0.132	0.040	0.209	0.132	0.103	0.131	0.025
	100	0.235	0.127	0.107	0.129	0.012	0.225	0.128	0.088	0.135	0.010
	200	0.211	0.096	0.064	0.108	0.008	0.202	0.092	0.052	0.103	−0.006
$\tau = 0.4$				$N = 5$					$N = 10$		
	T	Naive	Post	OS	OSw	DS	Naive	Post	OS	OSw	DS
	50	0.195	0.103	0.101	0.105	0.040	0.152	0.082	0.069	0.076	0.010
	100	0.180	0.063	0.050	0.068	−0.038	0.176	0.079	0.067	0.093	−0.009
	200	0.209	0.077	0.046	0.085	−0.012	0.195	0.076	0.039	0.084	−0.012
$\tau = 0.6$				$N = 5$					$N = 10$		
	T	Naive	Post	OS	OSw	DS	Naive	Post	OS	OSw	DS
	50	0.214	0.108	0.089	0.103	0.023	0.158	0.087	0.062	0.050	−0.037
	100	0.234	0.106	0.087	0.105	0.012	0.243	0.117	0.079	0.120	0.019
	200	0.230	0.097	0.067	0.098	0.016	0.0.258	0.103	0.069	0.112	0.018
$\tau = 0.8$				$N = 5$					$N = 10$		
	T	Naive	Post	OS	OSw	DS	Naive	Post	OS	OSw	DS
	50	0.270	0.127	0.115	0.124	0.020	0.252	0.114	0.086	0.100	0.010
	100	0.250	0.087	0.059	0.086	0.010	0.264	0.112	0.083	0.120	0.012
	200	0.205	0.039	0.023	0.039	0.001	0.196	0.045	0.025	0.044	0.002

D.3 Coverage for different quantiles

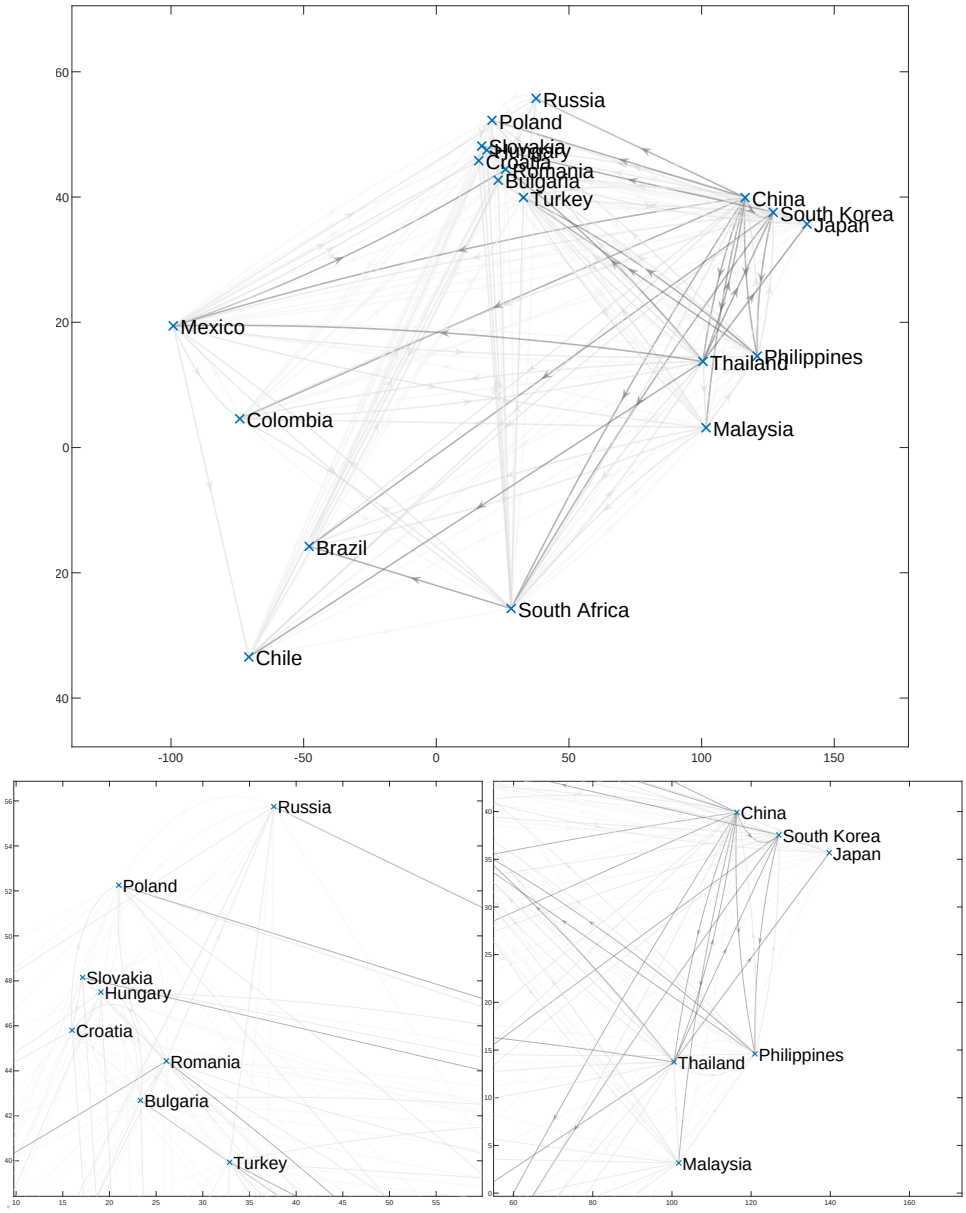
The remaining quantiles and the corresponding coverage and standard deviation of the estimators can be found in Table 11.

Table 11: Coverage and average standard deviation of confidence regions resulting from the post-ℓ_1, QR, and DS estimators of the treatment effect confounded by τ-dependent Erdos-Renyi networks under nominal coverage probability of 95%. For all settings, we used $R = 500$, adaptive penalization and cross-validation to select the penalty parameter.

Design			Coverage			SD		
τ	N	T	Post	QR	DS	Post	QR	DS
0.4	5	50	90.3%	96.0%	95.4%	0.357	0.502	0.470
0.4	5	100	88.3%	96.6%	96.6%	0.247	0.335	0.313
0.4	5	200	87.7%	98.6%	98.6%	0.171	0.222	0.207
0.4	10	50	90.9%	97.4%	96.0%	0.272	0.433	0.368
0.4	10	100	89.7%	98.0%	96.0%	0.184	0.277	0.243
0.4	10	200	83.7%	97.7%	98.2%	0.130	0.190	0.160
0.6	5	50	88.3%	96.3%	94.9%	0.376	0.506	0.480
0.6	5	100	85.3%	97.4%	95.4%	0.271	0.347	0.332
0.6	5	200	84.0%	97.1%	95.7%	0.196	0.242	0.230
0.6	10	50	88.6%	98.3%	96.0%	0.292	0.442	0.384
0.6	10	100	83.7%	97.4%	95.7%	0.212	0.296	0.270
0.6	10	200	84.6%	96.8%	96.0%	0.153	0.204	0.184
0.8	5	50	88.0%	99.1%	98.2%	0.297	0.434	0.393
0.8	5	100	86.9%	97.1%	96.0%	0.197	0.250	0.235
0.8	5	200	88.6%	96.9%	94.9%	0.145	0.167	0.161
0.8	10	50	84.0%	98.6%	96.0%	0.224	0.421	0.320
0.8	10	100	78.6%	98.9%	96.3%	0.144	0.226	0.184
0.8	10	200	87.4%	96.6%	95.1%	0.103	0.132	0.117

E Figures empirical example

Figure 2: Estimated network structure at $\tau = 0.9$.



Notes: The darkness of an edge represents the strength of the corresponding link: the darker the stronger. The global overview is in the top panel whereas details on eastern european and asian countries are in, respectively, the bottom left and right panels.