TIME SERIES APPROACH TO THE EVOLUTION OF NETWORKS: PREDICTION AND ESTIMATION

ANNA BYKHOVSKAYA

ABSTRACT. The paper analyzes non-negative multivariate time series which we interpret as weighted networks. We introduce a model where each coordinate of the time series represents a given edge across time. The number of time periods is treated as large compared to the size of the network. The model specifies the temporal evolution of a weighted network that combines classical autoregression with non-negativity, a positive probability of vanishing, and peer effect interactions between weights assigned to edges in the process. The main results provide criteria for stationarity vs. explosiveness of the network evolution process and techniques for estimation of the parameters of the model and for prediction of its future values.

Natural applications arise in networks of fixed number of agents, such as countries, large corporations, or small social communities. The paper provides an empirical implementation of the approach to monthly trade data in European Union. Overall, the results confirm that incorporating non-negativity of dependent variables into the model matters and incorporating peer effects leads to the improved prediction power.

KEYWORDS: multivariate time series, non-negative time series, networks, dynamic tobit, LAD.

Date: July 31, 2021.

The author would like to thank Donald Andrews, Vadim Gorin, Bruce Hansen, Peter Phillips, Jack Porter, and Larry Samuelson as well as anonymous referees for valuable comments and suggestions.

1. Introduction

1.1. Motivation. This paper studies non-negative time series. We are especially interested in multivariate time series, whose components can be interpreted as edges of some network (if a network is composed of only two vertices, then we are dealing with a basic one-dimensional time series). The novelty of our view on networks lies in the time series approach. Citing Diebold and Yılmaz (2014), "the network and multivariate time series literatures have much to learn from each other", and our paper makes a step in this direction. We develop a model and a consistent estimator of its parameters. The three main features of our model are non-negativity of weights associated with edges, positive probability of vanishing of each edge, and possibility for edges to affect each other in the multivariate setting.

Non-negative time series often arise in different socio-economic settings. For example, production decisions by firms cannot be negative. Similarly, trade and technology adoption by firms or countries is a non-negative variable, which evolves over time. Macroeconomic variables such as exchange rates and T-bill rates are non-negative by definition. Finally, the amount of time two people talk to each other at a given day is again either zero or positive.

One can note that in almost all of the above examples zero occurs with positive probability: there are times when firms decide not to produce at all, or two countries do not sell some good to each other, or a group of people do not call each other and so on. Thus, if one wants to find a suitable model for such non-negative processes, allowing it to take zero values with positive probability is an important condition. However, this feature cannot be captured by linear models or by their simplest modifications such as GARCH.

Moreover, such non-negative processes can be interdependent, leading us to consider multivariate time series, which we view as networks. E.g., trade flow between firms A and B can affect the trade between firms A and C: if a firm A is in search of a provider of some intermediate good, it may contact firm B, with which it has well-established trading relationship. Even if B cannot provide such intermediate good, it may recommend firm C, with which B in turn has a well-established relationship. Robinson and Stuart (2006) illustrate that the whole network of past alliances in the biotechnology industry affects the structure and size of alliance agreements between any two given firms in the industry in the future.

The availability of time series data for social and economic networks, such as call detail records and international and financial trade data has grown tremendously. In this paper we use such time-series data to estimate parameters of network evolution and to make predictions. In practice being able to produce a good forecast of time-varying networks and multivariate time-series helps in election and sales campaigns (e.g., whom to target with ads or other type of information). We also refer to Holme and Saramäki (2012) for an interdisciplinary review highlighting the importance of the temporal structure of networks.

Working with networks as a multivariate time series (when the number of time periods is large compared to the size of the network) and not as a cross-section or panel makes a difference between our paper and what is currently done in networks literature (see e.g., Bramoullé et al., eds (2016)). One natural example of applicability of our setting is when agents are countries or large companies (as there are not so many of them). For social communities this is also of relevance, as the dynamics of small groups of people might be very different from the large ones (see e.g., Palla et al. (2007)).

1.2. Results.

1.2.1. *Setting.* We deal with multivariate non-negative time series, which we interpret as weighted networks (alternative interpretations are possible and are mentioned in Section 2). We want to capture three essential aspects of networks. Those properties are non-negativity of weights, possibility not to have an edge between any two nodes with positive probability, and possibility for the past of the network to affect all edges today. These three features lead us to consider a non-linear process which builds on the dynamic Tobit model[1]

$$\mathbf{y}_t = [\boldsymbol{\alpha} + \boldsymbol{\beta} \circ \mathbf{y}_{t-1} + \boldsymbol{\gamma} \circ \mathbf{z}_{t-1} + \mathbf{u}_t]_+, \quad t = 1, \dots, T, \tag{1}$$

where $\circ$ stands for the Hadamard product, $[\cdot]_+$ stands for the coordinate-wise positive part, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are N-dimensional vectors of unknown coefficients, which need to be estimated, $\mathbf{u}_t$ is a random N-dimensional error, which is independent across time t, but may be correlated along its N coordinates, $\mathbf{y}_t$ is observed N-dimensional vector which evolves over time, and

[1]Eq. (1) can also be obtained as a solution to a utility maximization problem, see Supplementary Material.

$\mathbf{z}_{t-1}$ is a peer effect or an interaction term. That is, $\mathbf{z}_{t-1}$ is a function (assumed to be explicitly known) of past periods of the process $\mathbf{y}_t$. It allows various coordinates of the process to affect each other in the future in a potentially non-linear way. This gives us a lot of flexibility in how we model peer-effects. Moreover, the positive part per se creates nonlinearity, which leads to technical difficulties. One can add more regressors to Eq. (1) and most of our results continue to hold in the extended setting (see Remarks 1 and 2).

Our setting allows for arbitrary heteroskedastisity in the errors $\mathbf{u}_t$, that is, arbitrary correlation across coordinates or individuals. This differs a lot from what is feasible when one relies on cross-sectional variation instead of the variation across time. For example, in the networks setting where coordinates of $\mathbf{y}_t$ are coded by pairs (i,j), so that y_{ijt} is the weight of the edge $(i \to j)$ and $\mathbf{y}_t = (y_{ijt})_{i,j}$, the classical approach (N is large and $T = 1$) assumes either dyadic error structure where $Cov(u_{ijt}, u_{k\ell t}) = 0$ if $i \neq k, \ell$ and $j \neq k, \ell$ (e.g., Graham (2017), Graham (2019)) or other type of decreasing correlation, so that one can average over individuals and use law of large numbers and central limit theorem.

In principle, an alternative basic modelling approach could be to consider a latent censored model, see e.g., Wei (1999) and references therein. That is to assume that we observe a truncated process of the form $y_t^* = y_t \mathbf{1}(y_t > 0)$, where y_t is an unobserved underlying process. In our applications there need not exist such a process. Moreover, the censored structure makes the observed process non-Markovian. Our approach, in contrast, guarantees, a Markovian structure (in a state space taking into account several time lags). So it is easy to make predictions based on our model.

Let us describe our main findings and compare them with the existing literature on dynamic Tobit and non-linear time series.

1.2.2. *Mixing and stationarity.* Our first set of results describes asymptotic properties of Eq. (1). Those properties are crucial for the estimation of parameters of the model. In Theorem 2 we give[2] sufficient conditions under which the process $\mathbf{y}_t$ is strongly mixing and converges (when started from any initial conditions) to a stationary distribution as

[2]In this and other theorems we state the results in the networks notation $\mathbf{y}_t = (y_{ijt})_{i,j}$. When coordinates of $\mathbf{y}_t$ have other meaning, the same results follow by straightforward renaming of variables.

$t \to \infty$. As far as we know, those results for the model with interactions are new. Due to the presence of interactions $\mathbf{z}_t$, the model does not belong to a classical dynamic Tobit framework and requires a different treatment: Non-linearities created by $\mathbf{z}_t$ significantly complicate the setting, leading to a potentially non-Lipschitz dependence on the past and making asymptotic analysis challenging. For example, in the networks setting the triangular peer-effect function $z_{ijt} = \sum_k \frac{\sqrt{y_{ikt}y_{kjt}}}{n-2}$, which we use in the empirical example, is not Lipschitz. Thus, the classical arguments for stationarity, which are based on the contraction mapping (for a recent treatment see e.g., Debaly and Truquet (2021)), are not applicable and one has to come up with a different approach. Our solution is to develop a proof based on the large deviations principle for $\mathbf{y}_t$ and analyze the expected time it takes for the process to jump to zero. Then we apply renewal theorem to obtain mixing and convergence to stationarity independent of initial conditions.

For the special case of the model when there are no interactions ($\boldsymbol{\gamma} \equiv 0$ so that the model reduces to a separate equation of the form $y_t = [\alpha + \beta y_{t-1} + u_t]_+$ for each coordinate), we provide the full classification of asymptotic behavior of the process (Theorem 3): the change from stationary to explosive behavior is at $\alpha = 0$, $\beta = 1$. One distinction with the usual linear autoregression model is that the $\alpha < 0$, $\beta = 1$ case is stationary.

While the treatment of our most general case with peer effects was not addressed before, the particular setting when there are no interactions intersects with the dynamic Tobit literature. When $\beta < 1$ existence of a stationary solution for such setting was shown in de Jong and Herrera (2011) and Hahn and Kuersteiner (2010); the latter and Michel and de Jong (2018) also develop the strong mixing property of this solution. Compared to these papers, we additionally analyze the case of arbitrary initial conditions and show that the mixing properties continue to hold and the system eventually converges to the stationary regime. For the complementary $\beta \geq 1$ situation the only previous result that we are aware of is the analysis of Cavaliere (2004) equivalent to our $\alpha = 0$, $\beta = 1$ case.

Let us stress that in contrast to classical autoregression, where a continuous distribution of the errors is important to establish strong mixing (see Withers (1981) and Andrews (1984) for examples of non-strongly mixing AR processes with discrete errors), in our setting we

only need the errors to have support that is unbounded from below. Thus, the distribution is not required to be continuous. This also differs from Hahn and Kuersteiner (2010), who require a continuous distribution of the errors to obtain strong mixing.

1.2.3. *Estimation.* We discuss how to estimate the parameters of the model (Theorems 5-9). Due to censoring the ordinary least squares (OLS) estimator is inconsistent in our setting[3]. This matches a similar inconsistency for censored regression models in cross-section, cf. discussion at the end of Section 4.2 in Amemiya (1984). Our approach builds on and extends the least absolute deviations (LAD) method, which was used for Tobit models in cross-section in Powell (1984). Theorems from Powell (1984) are not applicable in our setting because the independence assumption from that paper does not hold. Additional complication is presented by the fact that while $\mathbf{y}_t$ might converge to a stationary process, it is not stationary per se. Thus, novel ideas are required to show LAD's consistency and asymptotic normality.

Some of our results also cover explosive cases and we show that while the large T asymptotics of the system changes drastically, the LAD estimator is still consistent. These findings are in line with the results on the consistency of OLS in the explosive autoregressive model (see White (1958), Anderson (1959), and Wang and Yu (2015)).

To our knowledge, the only previous text which addressed the asymptotic properties of LAD for dynamic Tobit models is de Jong and Herrera (2011). Restricted to our setting, the results of the above paper can be used to analyze the model in the stationary regime without peer effects ($\boldsymbol{\gamma} \equiv 0$). Our work goes beyond de Jong and Herrera (2011) by allowing peer effects, arbitrary initial conditions, and explosive and unit root type behaviors.

We remark that, in general, minimization of absolute deviations in models with a positive part is a non-convex problem and designing numerical algorithms requires special care (see e.g., Khan and Powell (2001)). Similarly, asymptotic normality in the censored cross-section model in Powell (1984) relies on certain continuity properties of a function of the true parameter value and the error distribution, which are hard to check. In contrast, we

[3]In Supplementary Material we show how to correct the OLS procedure to restore consistency. This requires throwing out observations and the accuracy of the estimation decreases. Further, in the special case when errors $\mathbf{u}_t$ are Gaussian we prove that the maximum likelihood estimator (MLE) is consistent. In contrast, our approach based on LAD does not require Gaussianity.

find that whenever interactions term has non-negative support and the true coefficients are all positive, neither of these problems exist in our setting.

Being able to consistently estimate $\boldsymbol{\gamma}$ allows us to analyze the importance of peer effects in the network formation process. This relates our paper to the vast literature on estimating peer effects (e.g., see Blume et al. (2011) for review). Several approaches rely on specific functional forms to ensure a consistent estimate (e.g., triangular peer effects in Graham (2016), which resemble our triangles example in Section 2.2), while others introduce model free notions for the strength of peer effects (such as connectedness in Diebold and Yılmaz (2015)). However, we have not seen in the literature estimation techniques based on variation across time in non-linear Markov evolution model, as in our work.

To illustrate our methodology, we apply the framework to monthly trade data between European Union countries for pharmaceutical products. In our experiments model-based estimation techniques improve predictions compared to the benchmark "tomorrow=today" approach. The addition of peer effects $\mathbf{z}_t$ (vs. setting $\boldsymbol{\gamma} \equiv 0$ in the model specification) also leads to improved prediction power. Results of the Diebold-Mariano test (Diebold and Mariano (1995)) and the tests for the significance of peer effects, which are reported in Section 6.3, support the above conclusions.

1.3. Outline of the paper. Section 2 presents the model and the main equation of interest. All assumptions are stated in that section, as well as sufficient conditions for stationarity. Section 3 discusses the special case of the model when there are no peer effects. Section 4 discusses estimation of the model, while Section 5 proposes a method to measure predictive power. Section 6 applies the model to the trade of pharmaceutical products in European Union. Finally, Section 7 concludes. All proofs, unless otherwise noted, are in Supplementary Material, which as well contains results on OLS and MLE estimation.

2. Model

2.1. Framework. We analyze a multivariate non-negative time series with potential interactions. At time $t = 1, \ldots, T$ we observe N-dimensional vector $\mathbf{y}_t$. In our asymptotic results we assume that N is fixed and T goes to infinity. The value of each coordinate of the

process at time t can be influenced by other coordinates in the past H periods via several peer effects/interactions functions $\mathbf{p}^\kappa : \mathbf{R}_+^{NH} \to \mathbf{R}^N$, $\kappa = 1, \dots, K$, which are assumed to be explicitly known. Formally, the data generating process is

$$\mathbf{y}_t = [\boldsymbol{\alpha} + \boldsymbol{\beta} \circ \mathbf{y}_{t-1} + \sum_{\kappa}^{K} \boldsymbol{\gamma}^\kappa \circ \mathbf{p}^\kappa (\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-H}) + \mathbf{u}_t]_+, \tag{2}$$

where $\mathbf{u}_t$ is a random N-dimensional error, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}^\kappa, \kappa = 1, \dots, K$, are N-dimensional vectors of unknown coefficients, $\circ$ stands for the Hadamard product[4], and $[\cdot]_+$ is the coordinate-wise positive part. The positive part in Eq. (2) creates nonlinearity and leads to a positive mass at zero.

The model is initialized at $t = 1 - H, \dots, 0$ by arbitrary values (possibly random). To be more precise, we assume that as T goes to infinity, H does not grow and $\mathbf{y}_t = O(1)$ for $t = 1 - H, \dots, 0$.

Functions $\mathbf{p}^\kappa$ serve as various aggregators of the past structure of the data in a way that affects the current state. Their dependence on more than one past period captures the fact that it may take multiple periods for the information about others to be transmitted. The multitude of peer-effect terms $\boldsymbol{\gamma}^\kappa \circ \mathbf{p}^\kappa$ gives a lot of flexibility and allows one to capture very general, diverse forms of interactions. A lot of features of the data can be captured through $\mathbf{p}^\kappa$. Additionally, extra lags of the autoregressive component can be modeled by $\mathbf{p}^\kappa$. E.g., by adding extra peer-effects $\sum_{\tau=2}^{P} \boldsymbol{\beta}^\tau \circ \mathbf{y}_{t-\tau}$, we get $\sum_{\tau=1}^{P} \boldsymbol{\beta}^\tau \circ \mathbf{y}_{t-\tau}$ instead of $\boldsymbol{\beta} \circ \mathbf{y}_{t-1}$.

There are many settings which could be modeled via Eq. (2). Let us mention some.

(1) Our main example is a time-varying network on n vertices. Undirected network has $N = n(n-1)/2$ coordinates and directed network has $N = n(n-1)$ coordinates. For the former case, we can think of a social weighted network, where nodes represent people and edges represent how much time they spend together. For the case of a directed network, we can think of firms or countries and trade between them. Looking separately at exports and imports, we get a directed network.

[4]For two matrices A and B of the same dimension $m_1 \times m_2$, the Hadamard product $A \circ B$ is a $m_1 \times m_2$ matrix with elements given by $(A \circ B)_{ij} = (A)_{ij}(B)_{ij}$.

In the case of networks it is convenient to write $\mathbf{y}_t = (y_{ijt})_{i,j=1,\dots,n}$, so that y_{ijt} can be interpreted as either how much i and j interact at time t or the amount of trade from i to j at time t.

(2) Our framework is suitable for non-negative panels where N is fixed and T is large and we expect some interactions across units. E.g., alcohol consumption by classmates or behavior of various financial markets. In this case we can view cross-section units (classmates or financial markets) as vertices in a network and analyze their evolution. That is, we can analyze how the alcohol consumption of one's peers affects one's decision to drink or how a crush in one markets affects the behavior of other markets. When one analyzes financial markets, we may expect that a crush in one of them pushes the others also to go down.

(3) Furthermore, our setting allows for joint analysis of the evolution of edges and vertices in a network. This is achieved by combining cases (1) and (2) together so that $N = n + n(n-1)$ for a directed and $N = n + n(n-1)/2$ for an undirected network.

(4) A special case of Eq. (2) with $H = 1$ and linear interactions corresponds to a non-negative VAR(1),

$$\mathbf{y}_t = [\boldsymbol{\alpha} + \mathbf{B}\mathbf{y}_{t-1} + \mathbf{u}_t]_+,$$

where $\mathbf{B}$ is $N \times N$ matrix of unknown coefficients. Such setting can be useful in modeling nominal interest rates or realized volatilities.

Through the rest of the paper we focus on the first application: evolution of a small weighted network. Results for alternative applications follow by straightforward renaming of variables. To shorten the notations, in the rest of the paper we focus on the special case of Eq. (2) where $K = 1$. As noted in the Remarks 1 and 2, our results still hold in the extended setting of Eq. (2) with $K > 1$.

With the networks notation $\mathbf{y}_t = (y_{ijt})_{i,j=1,\dots,n}$ and similarly for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$, and $\mathbf{p}(\cdot)$, we can rewrite our data generating process as

$$y_{ijt} = \left[\alpha_{ij} + \beta_{ij} y_{ijt-1} + \gamma_{ij} p_{ij}\left(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-H}\right) + u_{ijt}\right]_+ . \tag{3}$$

In Eq. (3) $\alpha_{ij} \in \mathbb{R}$ is a proxy for homophily, i.e. how similar i and j are. The larger α_{ij} is, the stronger is the link connecting i and j. We allow β_{ij} and γ_{ij} to be of any sign. The coefficient β_{ij} measures the dependence on the own past. The larger β_{ij} is, the more the link between i and j yesterday affects its weight today. The coefficient γ_{ij} captures the dependence on the peer effects/interactions p_{ij}.

2.2. Assumptions. We impose some assumptions on the error distribution and on the peer effects function.

Assumption 1. *The vector* $\{u_{ijt}\}_{i,j=1,\dots,n}$ *is i.i.d. over* t.

Assumption 1 allows errors to have arbitrary correlation across edges of the network and ensures that those correlations do not vary over time. This differs substantially from classical approach to networks, where they are modeled as cross-section or short panel. In those cases the correlation across individuals must either form a dyadic relationship or decrease with some measure of a distance, so that one can average over individuals and apply the law of large numbers. In our setting n is assumed to be fixed, and instead the variation across T is used to identify the network formation process. Thus, our setting does not rely on "decreasing across n correlations" and can accommodate any type of cross-sectional interdependence.

While we do not prove it here, we believe that Assumption 1 can be weakened and most of the results continue to hold for more general correlation across time (e.g., $MA(q)$ errors or $MA(\infty)$ with decaying fast enough mixing coefficients).[5]

Assumption 2. *The support of* $\{u_{ijt}\}_{i,j=1,\dots,n}$ *satisfies* $\mathbb{P}(u_{ijt} < -M \,\forall i, j) > 0$ *for all* $M > 0$.

Assumption 2 is used to show stationarity.[6] It implies that errors jointly take large negative values with positive probability. For example, it holds for normally distributed errors.

[5]The most challenging part of the extension is showing asymptotic normality of the LAD estimator. This would require a potentially different proof and the asymptotic variance would need to be updated. The much more complicated formula for the variance can be guessed by comparing with de Jong and Herrera (2011).

[6]Convergence to stationarity will be used to show consistency and asymptotic normality of our estimators of $\{\alpha_{ij}, \beta_{ij}, \gamma_{ij}\}_{i,j}$. Yet, it is plausible that one can relax Assumption 2, and still get a consistent estimator as long as several first moments of the network evolution process $\{y_{ijt} : i, j = 1 \dots, n\}_{t \geq 1}$ are uniformly bounded. We further discuss estimation without stationarity in Remark 3 and bounds on moments in Theorem 1.

In contrast, uniform distribution (or any other bounded distribution) does not satisfy Assumption 2. A consequence of Assumption 2 is that for any values of the process $\{y_{ijt}\}_{i,j}$ at time t, there is a positive probability that at time $t+1$ the process jumps to zero. This observation is crucial for mixing and stationarity properties in Theorems 2 and 3.

Eq. (3) has one degree of freedom, so we need to impose a normalization assumption on the error term u_{ijt}. We consider two different normalization assumptions: zero mean or zero median, which are stated below.

(1) *Normalization of the mean:* For all i, j, t, $\mathbb{E}u_{ijt} = 0$.

(2) *Normalization of the median:* For all i, j, t, $\text{med}(u_{ijt}) = 0$.

Alternative normalizations only lead to differences in α_{ij}:

$$\alpha_{ij}^{E} = \alpha_{ij}^{\text{med}} + \mathbb{E}u_{ijt}^{\text{med}},$$

where α_{ij}^{E} is an intercept under normalization of the mean and α_{ij}^{med} and $\mathbb{E}u_{ijt}^{\text{med}}$ are an intercept and a mean of the error under normalization of the median.

Assumption 3 (Peer effects do not grow faster than their maximal argument.)**.** *$p_{ij} : \mathbb{R}_{+}^{NH} \to \mathbb{R}$ is such that there exists a constant $\mathcal{A} \in \mathbb{R}$ for which*

$$|p_{ij}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-H})| := \left| p_{ij}\left(\left\{ y_{kls} \right\}_{\substack{k,l=1,\ldots,n \\ s=t-H,\ldots,t-1}} \right) \right| \leq \mathcal{A} + \max_{\substack{k,l=1,\ldots,n \\ s=t-H,\ldots,t-1}} y_{kls}.$$

Each function $p_{ij}(\cdot)$ depends on H observations of the N-dimensional process $\mathbf{y}_t$. Thus, the function $p_{ij}(\cdot)$ has NH arguments of the form y_{kls}. Assumption 3 requires the peer effect function to be bounded by a linear function of its maximal argument, $\max_{\substack{k,l=1,\ldots,n \\ s=t-H,\ldots,t-1}} y_{kls}$. The choice of constant 1 in front of $\max_{k,l,s} y_{kls}$ is without loss of generality. Writing instead $C \max_{k,l,s} y_{kls}$, where $C > 0$ is some constant, only leads to rescaling of the peer-effect function and its coefficient γ_{ij}.

Let us present some examples of possible peer effect functions.

- *Maximum*:

$$p_{ij}\left(\left\{ y_{kls} \right\}_{\substack{k,l=1,\ldots,n \\ s=t-H,\ldots,t-1}} \right) = \max_{\substack{(k,l)\neq(i,j) \\ s=t-H,\ldots,t-1}} y_{kls}.$$

This function represents the largest possible stimulus to increase trade or communication. This can be interpreted as a steadily expanding economy.

- *Minimum*:

$$p_{ij}\left(\left\{y_{kls}\right\}_{\substack{k,l=1,\dots,n\\ s=t-H,\dots,t-1}}\right) = \min_{\substack{(k,l)\neq(i,j)\\ s=t-H,\dots,t-1}} y_{kls}.$$

This function corresponds to the smallest, but still non-zero influence from others. That is, if some edge jumps to zero, it pushes the other edges in that direction. Alternatively, if all edges have positive weights, the peer effect term is still positive and helps to maintain a non-zero edge between i and j.

- *Linear*:

$$p_{ij}\left(\left\{y_{kls}\right\}_{\substack{k,l=1,\dots,n\\ s=t-H,\dots,t-1}}\right) = \sum_{\substack{k,l=1,\dots,n\\ r=1,\dots,H}} \lambda_{klr} y_{klt-r},$$

$$\text{where } \{\lambda_{klr}\}_{\substack{k,l=1,\dots,n\\ r=1,\dots,H}} \text{ are known and } \sum_{\substack{k,l=1,\dots,n\\ r=1,\dots,H}} |\lambda_{klr}| \leq 1.$$

The linear function represents an intermediate point between the previous examples.

- *Triangles*:

$$p_{ij}\left(\left\{y_{kls}\right\}_{\substack{k,l=1,\dots,n\\ s=t-H,\dots,t-1}}\right) = \sum_{k\neq i,j} \frac{\sqrt{y_{ikt-H} y_{kjt-H}}}{n-2}.$$

The square root makes the order of the peer effect term to be the same as the order of the autoregressive term, y_{ijt}. Thus, Assumption 3 is satisfied.

Triangular peer effects have the most interesting functional form, and we use it in our empirical application. The interpretation is that if i and k are strongly connected, and k and j are too, then there is a higher probability of a connection between i and j in the future. Note that the product makes it important that both connections are present. If k and j are not connected, we cannot expect k to "introduce j to i". Thus, we look at all triangles which have (i, j) as one of their legs. Two strong links in such triangles are expected to strengthen the third leg, (i, j). This can be summed up as "friend of my friend is my friend". Such peer effects may be present in social interactions, interactions between firms or countries, and so on.

The triangular peer effect function is related to the network formation process in Graham (2016), where the presence of an edge between two nodes depends on the number of triangules containig those two nodes in the past. In Graham (2016) edges do not have weights, so the number of triangles containing i and j at time t is $\sum_k D_{ikt}D_{kjt}$, where $D_{ikt} = 1$ when there is an edge between i and k at time t. A similar statistics $\max_k D_{ikt}D_{kjt}$ is used in Leung and Moon (2019) to generate network clustering.

2.3. Stationarity. Theorem 1 provides sufficient conditions under which the mean of the process $\mathbf{y}_t$ remains bounded. This finiteness does not guarantee convergence to a stationary distribution as the process may have cycles. If we additionally assume that $\{u_{ijt}\}_{i,j}$ has unbounded from below support (Assumption 2), then stationarity holds (Theorem 2).

Theorem 1. *Suppose that Assumptions 1 and 3 are satisfied,* $\mathbb{E}u_{ijt}$ *exists for all* i, j, t*, and there exists a constant* $C \in (0, 1)$ *such that* $\max(0, \beta_{ij}) + |\gamma_{ij}| < C$ *for all* i, j*. Then there exists a constant* C_1 *such that* $\mathbb{E}y_{ijt} < C_1 < \infty$ *for all* i, j, t*.*

Definition. *Let* $\mathbf{y}_t := \{y_{ijt}\}_{i,j}$ *and let* $\mathcal{Y}_{t,s}$ *be the* σ*-algebra generated by* $\mathbf{y}_t, \dots, \mathbf{y}_s$*. The process* $\mathbf{y}_t$ *is strongly mixing if*

$$\lim_{t\to\infty} \sup_s \sup_{\substack{\Delta_1\in\mathcal{Y}_{0,s},\\ \Delta_2\in\mathcal{Y}_{t+s,\infty}}} |\mathbb{P}(\Delta_1 \cap \Delta_2) - \mathbb{P}(\Delta_1)\mathbb{P}(\Delta_2)| = 0.$$

Theorem 2. *Suppose that Assumptions 1, 2, and 3 are satisfied,* $\mathbb{E}u_{ijt}^4 < \infty$ *for all* i, j, t*, and there exists a constant* $C \in (0, 1)$ *such that* $\max(0, \beta_{ij}) + |\gamma_{ij}| < C$ *for all* i, j*. Then the multivariate process* $\mathbf{y}_t$ *is strongly mixing and converges to a stationary process.*[7]

Remark 1. In the case of multiple peer effects (extended setting of Eq. (2)), Assumption 3 should hold for each p_{ij}^κ, and $|\gamma_{ij}|$ in Theorem 2 should be replaced by the sum $\sum_{\kappa=1}^{K} |\gamma_{ij}^\kappa|$.

A striking feature of Theorem 2 is that we do not need the error distribution to be continuous to get strong mixing. This differs from the linear case (see Withers (1981) and

[7]Formally this means that the finite-dimensional distributions of the process $\{y_{ijt+\tau} : i, j = 1 \dots, n\}_{\tau\in\mathbb{Z}}$, converge to those of a stationary in τ process as $t \to \infty$.

Andrews (1984) for examples of $AR(1)$ processes which are not strongly mixing). The reason is that in our setting the expected value of the time until the process jumps to be identically zero ($y_{ijt} = 0$ for all i, j) is finite. After this time the precise value of the initial condition becomes irrelevant.

Example 1. The fact that the peer effects function p_{ij} can depend only on a fixed number of time periods is crucial. For example, suppose that we have only one equation ($n = 2$) which is initialized at $y_0 = 0$, and the error process u_t has unbounded support from above. Further suppose $\alpha = \beta = 0$, $\gamma = 0.5$ and $z_t := p_{ij}(y_t, \dots, y_0) = \max(y_t, \dots, y_0)$, so that Assumption 3 is satisfied. Then $\beta + |\gamma| = 0.5 < 1$, but the process y_t is explosive. To see this, let us analyze the behavior of y_t and z_t.

By definition, $z_0 = y_0 = 0$, so that $y_1 = [u_1]_+$ and $z_1 = \max(0, u_1) \geq 0$. Thus, $y_2 = [0.5z_1 + u_2]_+ \geq [u_2]_+$ and $z_2 = \max(0, y_1, y_2) \geq \max(0, u_1, u_2)$. Similarly, $y_3 = [0.5z_2 + u_3]_+ \geq [u_3]_+$ so that $z_3 \geq \max(0, u_1, u_2, u_3)$. Applying induction, for any t we get $z_t \geq \max(0, u_1, \dots, u_t)$. Therefore, $z_t \xrightarrow[t\to\infty]{a.s.} \infty$, as the support of u_t is unbounded from above and the maximum of an infinite number of random variables with unbounded support diverges. Because $y_t = [0.5z_{t-1} + u_t]_+$, y_t also goes to infinity almost surely.

3. Special case: No interactions

In this section we consider a special case, where the weight of the edge between i and j at time t depends only on its own past. That is, past interactions between $k \neq i, j$ and l do not influence y_{ijt}. Thus, the model reduces to $\frac{n(n-1)}{2}$ separate equations of the form

$$y_t = [\alpha + \beta y_{t-1} + u_t]_+, \quad u_t \sim i.i.d.(0, \sigma^2). \tag{4}$$

A typical sample path for Eq. (4) is shown in Figure 1. When the process hits zero, it stays at zero for some time, then goes to an "AR(1)-excursion", until it becomes negative. Positive part in (4) then forces y_t to become zero instead, and everything starts again.

The following theorem provides a full classification of stationary/explosive behavior in the case of no interactions. In contrast to classical autoregression which has no positive part, when $\beta = 1$ the process still converges to a stationary distribution if $\alpha < 0$.

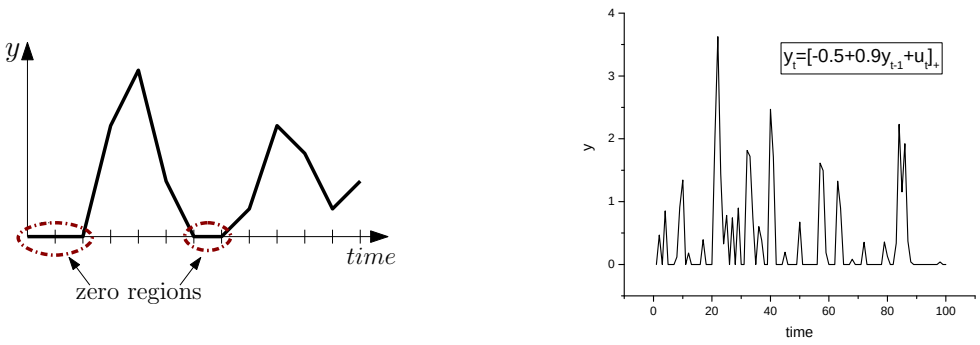

FIGURE 1. A typical sample path for $y_t = [\alpha + \beta y_{t-1} + u_t]_+$, $u_t \sim i.i.d.(0, \sigma^2)$.

Theorem 3. *(Classification Theorem) Let Assumptions 1 and 2 hold. Assume further that* $\mathbb{P}(u_t > -\alpha) > 0$, $\mathbb{E}u_t = 0$, *and* $\mathbb{E}u_t^4 < \infty$,

- *If* $\beta < 1$, *then* y_t *is strongly mixing and converges to a stationary process;*
- *If* $\beta = 1$, $\alpha < 0$, *then* y_t *is strongly mixing and converges to a stationary process;*
- *If* $\beta > 1$, *then* y_t *is divergent:* y_t *converges to* ∞ *almost surely;*
- *If* $\beta = 1$, $\alpha > 0$, *then* y_t *is divergent:* y_t *converges to* ∞ *almost surely;*
- *If* $\beta = 1$, $\alpha = 0$, *then* y_t *is mean-divergent:* $\mathbb{E}y_t \to \infty$. *The proper scaling limit is*

$$\frac{1}{\sqrt{T}} y_{\lfloor Tr \rfloor} \xrightarrow{d} \sigma |W(r)|, \; r \in [0,1], \;\; as \; T \to \infty,$$

where $W(\cdot)$ *is a standard Brownian motion and* $\mathbb{E}u_t^2 = \sigma^2$.

A visual summary of the results in Theorem 3 is shown in Figure 2, where the evolution of y_t is illustrated for different values of α and β.

Theorem 3 is proved in the Section D of the Supplementary Material (Theorems D.4, D.5, D.6, D.7, and D.9). The stationarity part of the proof relies on the large deviations principle and the renewal theorem. The idea is to show that the expected time until the process reaches zero is finite. Then one can apply the renewal theorem to get the limiting distribution. Interestingly, just the knowledge that the process hits zero with probability one is not enough. In particular, if $\alpha = 0, \beta = 1$ we get a standard unit-root process for time periods when y_t is positive. This process always hits zero. However, it does not converge to a stationary distribution, and as Theorem 3 shows, the mean of the process $y_t = [y_{t-1} + u_t]_+$ goes to infinity. Thus, there is a discontinuity between the stationary and explosive regions.

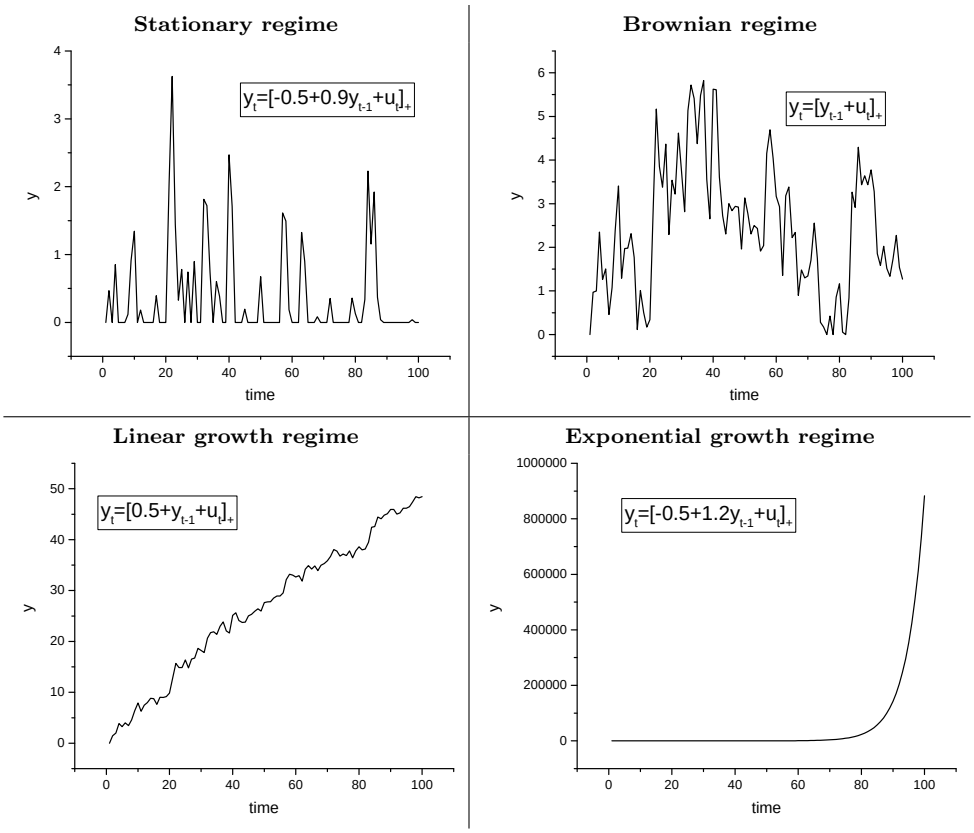

FIGURE 2. Illustration of Theorem 3.

This is similar to what is observed in the classical linear autoregressive case. However, while in the classical case the re-scaled unit root process converges to the Brownian motion, in our truncated setting we get the absolute value of the Brownian motion, see Theorem D.9 in the Supplementary Material and Cavaliere (2004).

The limiting distribution in the stationary case of Theorem 3 is complicated and cannot be written explicitly as a function of α, β, and the distribution of u. However, what can be calculated is the expected time the process y_t spends at zero once it hits zero. In particular, the average length of an interval of zeros is decreasing in α.

Lemma 4. *Let Assumption 1 hold. Once the process y_t hits zero, the expected time it spends until finally jumping to a positive value is $\frac{1}{1-F_u(-\alpha)}$, where F_u is the cdf of u_t.*

Proof. If $y_t = 0$, then $y_{t+1} > 0$ when $u_{t+1} > -\alpha$. Therefore, after each zero with probability $F_u(-\alpha)$ the process remains at zero and with the remaining probability it becomes positive. Thus, we get a sequence of Bernoulli random variables, where the expected time until the

first tail observation is

$$1 \cdot (1 - F_u(-\alpha)) + 2 \cdot F_u(-\alpha)(1 - F_u(-\alpha)) + 3 \cdot F_u^2(-\alpha)(1 - F_u(-\alpha)) + \ldots = \tfrac{1}{1-F_u(-\alpha)}. \quad \square$$

4. Estimation

We now return back to the general model summarized in Eq. (3). To simplify notation, for each pair (i, j) we define

$$z_{ijt-1} = p_{ij}\left(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-H}\right).$$

The two main difficulties in our model are the interactions between the outcome variables and the non-separable errors. We overcome them by using the least absolute deviations (LAD) estimator. The LAD estimates are obtained by solving the following minimization problem

$$\underset{\{a_{ij}\}_{ij},\{b_{ij}\}_{ij},\{c_{ij}\}_{ij}}{\arg\min} \sum_{i,j,t} \left| y_{ijt} - [a_{ij} + b_{ij} y_{ijt-1} + c_{ij} z_{ijt-1}]_+ \right|. \tag{5}$$

Notice that the minimization (5) breaks into the separate procedure for each edge

$$\underset{a_{ij},b_{ij},c_{ij}}{\arg\min} \sum_{t} \left| y_{ijt} - [a_{ij} + b_{ij} y_{ijt-1} + c_{ij} z_{ijt-1}]_+ \right|.$$

Thus, we are left with analyzing the LAD estimation of each edge separately. Let us fix an edge $(i \to j)$ and define $y_t = y_{ijt}$, $z_t = z_{ijt}$. In the following subsections we discuss the properties of the LAD estimator for a fixed edge $(i \to j)$ in our model.

Remark 2. All of the results in this section can be straightforwardly generalized to the case of multiple peer effects terms. I.e., to the model with K regressors $z_{t-1}^1, \ldots, z_{t-1}^K$

$$y_t = [\alpha + \beta y_{t-1} + \sum_{k=1}^{K} \gamma^k z_{t-1}^k + u_t]_+,$$

where coefficients α, β, $\{\gamma^k\}_{k=1}^K$ are unknown and have to be estimated. For example, for the above model, the analogue of the matrix M_R defined in Eq. (7) and used in Theorems 5 and 6 is a $K+2$ by $K+2$ matrix composed of all second moments of the vector $(1, y_t, z_t^1, \ldots, z_t^K)\mathbf{1}(\alpha + \beta y_t + \sum_{k=1}^K \gamma^k z_t^k \geq R)$.

4.1. Edge Estimation. The LAD estimator for a fixed edge solves

$$\min_{a,b,c} \sum_{t=1}^{T} |y_t - [a + b y_{t-1} + c z_{t-1}]_+|. \tag{6}$$

It turns out, as Theorems 5 and 6 show, the LAD estimators are consistent and asymptotically normal under mild conditions.

Let $R \geq 0$ and define the matrix M_R as

$$M_R = \mathbb{E}\left[\begin{pmatrix} 1 & y & z \\ y & y^2 & yz \\ z & yz & z^2 \end{pmatrix} \mathbf{1}(\alpha + \beta y + \gamma z \geq R) \right], \tag{7}$$

where (y, z) denotes the distributional limit of (y_t, z_t) as $t \to \infty$.

Note that if $R_1 > R_2$ and M_{R_1} is nonsingular, then M_{R_2} is also nonsingular. Equivalently, if M_{R_2} is singular, then so is M_{R_1}. The reason is that if there exists a non-zero vector $(\lambda_1, \lambda_2, \lambda_3)$ such that $\lambda_1 + \lambda_2 y + \lambda_3 z = 0$ when $\alpha + \beta y + \gamma z \geq R_2$, then the same holds for $\alpha + \beta y + \gamma z \geq R_1$, as $R_1 > R_2$. Thus, there is a bound $\bar{R} \in \mathbb{R} \cup \{+\infty\}$ such that M_R is nonsingular for any $R < \bar{R}$ and singular for any $R > \bar{R}$.

Theorem 5. *Assume that y_t is strongly mixing and converges to a stationary distribution and $(\alpha, \beta, \gamma) \in \Theta$, where Θ is some compact set in $\mathbb{R}^3$. If u_t is i.i.d. with a continuous and positive density at 0, med$(u_t) = 0$, and M_R is nonsingular for some $R > 0$ at the true parameter values, then the LAD estimator is consistent:*

$$\left(\hat{\alpha}_{LAD},\, \hat{\beta}_{LAD},\, \hat{\gamma}_{LAD}\right) \xrightarrow[T\to\infty]{\mathbb{P}} (\alpha,\, \beta,\, \gamma)\,.$$

Role of the assumptions in Theorem 5:

- The condition on the error density $f_u(0) > 0$ can be weakened to

$$\mathbb{P}(u \in [-\varepsilon, 0)) > 0,\ \mathbb{P}(u \in (0, \varepsilon]) > 0 \text{ for any } \varepsilon > 0.$$

That is, u_t must be in the left and right neighbourhoods of zero with positive probability. From the proof of Theorem 5, we only need $\min\left(\int_{-\tau}^{0}(\tau+u)dF_u(u), \int_{0}^{\tau}(\tau-u)dF_u(u)\right)$ to be positive for any $\tau > 0$, which is satisfied in this case.

On the other hand, for Theorems 6 and 8 such weakening of the condition $f_u(0) > 0$ leads to a change in the asymptotic theory, and we do not address it in the present paper.

- The singularity of the matrix M_R is a rare event. The matrix M_R is singular if $\beta = \gamma = 0$, $\alpha < R$. In this case the indicator equals zero, so the matrix is identically zero. The other case is if $\beta = 0$, $\gamma > 0$ and z never takes values above $\frac{R-\alpha}{\gamma}$, so that again the matrix M_R is identically zero.

If $\beta > 0$, the peer effect function $p_{ij}(\cdot)$ does not depend on y_{ijt}, and the random variable z is non-constant, then the matrix is nonsingular. For instance, the triangular peer effect functions p_{ij} do not depend on y_{ijt}. Minimum and maximum functions, if taken over all edges except the given edge $(i \to j)$, also do not depend on y_{ijt}. Similarly linear functions satisfy this as long as the corresponding weight $\lambda_{ijt} = 0$. Thus, in all those examples M_R is non-singular for $\beta > 0$.

The idea of the above argument is that if M_R is singular, then there exists a non-zero vector $(\lambda_1, \lambda_2, \lambda_3)$ such that $\lambda_1 \mathbf{1}(\alpha + \beta y + \gamma z \geq R) + \lambda_2 y \mathbf{1}(\alpha + \beta y + \gamma z \geq R) + \lambda_3 z \mathbf{1}(\alpha + \beta y + \gamma z \geq R) \equiv 0$. That is, when $\alpha + \beta y + \gamma z \geq R$, we must have $\lambda_1 + \lambda_2 y + \lambda_3 z = 0$. However, as $\beta > 0$ and z does not depend on y, we can perturb y a bit and get $y' = y + \varepsilon$, $\varepsilon > 0$, in which case the indicator is still non-zero, but the second equality fails unless $\lambda_2 = 0$. If $\lambda_2 = 0$, then we must have $z = -\lambda_1/\lambda_3$ whenever the indicator equals one. This again is impossible, when z is not a fixed constant.

Theorem 6. *Assume that y_t is strongly mixing and converges to a stationary distribution and $(\alpha, \beta, \gamma) \in \Theta$, where Θ is some compact set in $\mathbb{R}^3$. If u_t is i.i.d. with a continuous and positive density at 0, med$(u_t) = 0$, M_0 is nonsingular at the true parameter values, and the random function $\Delta \mapsto \mathbf{1}(\alpha + \beta y + \gamma z + \Delta > 0)$ is continuous with probability one at the true parameter values and at $\Delta = 0$, then the LAD estimator is asymptotically normal:*

$$\sqrt{T}\begin{pmatrix} \hat{\alpha}_{LAD} - \alpha \\ \hat{\beta}_{LAD} - \beta \\ \hat{\gamma}_{LAD} - \gamma \end{pmatrix} \xrightarrow[T\to\infty]{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \frac{1}{4f_u^2(0)} M_0^{-1}\right).$$

Role of the assumptions in Theorem 6:

- The condition that the random function $\Delta \mapsto \mathbf{1}(\alpha + \beta y + \gamma z + \Delta > 0)$ is continuous with probability one at the true parameter values and at $\Delta = 0$ is generally hard to violate, since the violation requires $\alpha+\beta y+\gamma z$ to equal zero with positive probability. This can happen when $\alpha = 0$, because $y = 0$ with positive probability. If z is also zero with positive probability, then the indicator jumps depending on the sign of Δ. On the other hand if $\alpha > 0$, $\beta, \gamma \geq 0$ and $z \geq 0$ with probability 1, then the indicator is identical 1. This corresponds to the special case of Theorem 7.
- We do not impose a condition on the existence of nonsingular M_R for some $R > 0$ in Theorem 6. This is because non-degeneracy of M_0 together with continuity of $\mathbf{1}(\alpha+\beta y+\gamma z+\Delta > 0)$ imply the existence of $R > 0$, for which M_R is also nonsingular.

In the special "non-negative" case when $a > 0$, $b \geq 0$, $c \geq 0$, and $z \geq 0$ the minimization problem (6) is convex. Thus, a local minimum is a global minimum when the true $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$, which is useful in computer implementations. Moreover, the conditions of Theorems 5 and 6 can be simplified: $[a + by_{t-1} + cz_{t-1}]_+ \equiv a + by_{t-1} + cz_{t-1}$ when a is positive and b, c and z_{t-1} are non-negative (y_{t-1} is non-negative by assumption). Thus, in this case the positive part in the minimization problem never binds, and the additional condition on the stationary distribution and true parameter does not arise. This special case is illustrated in Theorems 7 and 8.

Theorem 7. *Assume that y_t is strongly mixing and converges to a stationary distribution. If $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$, the peer effect function is non-negative (i.e., $z_t \geq 0$ for all t), u_t is i.i.d. with a continuous and positive density at 0, $med(u_t) = 0$, and the random variables $1, y, z$ are linearly independent*[8]*, then the LAD estimator is consistent:*

$$\left(\hat{\alpha}_{LAD},\, \hat{\beta}_{LAD},\, \hat{\gamma}_{LAD}\right) \xrightarrow{\mathbb{P}} (\alpha,\, \beta,\, \gamma) \;\; as\; T \to \infty.$$

Theorem 8. *Assume that y_t is strongly mixing and converges to a stationary distribution. If $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$, the peer effect function is non-negative (i.e., $z_t \geq 0$ for all t), u_t is i.i.d. with a continuous and positive density at 0, $med(u_t) = 0$, and the random variables $1, y, z$ are linearly independent, then the LAD estimator is asymptotically normal:*

$$\sqrt{T}\begin{pmatrix} \hat{\alpha}_{LAD} - \alpha \\ \hat{\beta}_{LAD} - \beta \\ \hat{\gamma}_{LAD} - \gamma \end{pmatrix} \xrightarrow[T\to\infty]{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \frac{1}{4f_u^2(0)}\begin{pmatrix} 1 & \mathbb{E}y & \mathbb{E}z \\ \mathbb{E}y & \mathbb{E}y^2 & \mathbb{E}yz \\ \mathbb{E}z & \mathbb{E}yz & \mathbb{E}z^2 \end{pmatrix}^{-1}\right).$$

The linear independence condition in Theorems 7 and 8 holds automatically whenever the peer effect function $p_{ij}(\cdot)$ depends on an argument other that y_{ijt-1} in a non-degenerate way. Indeed, in this case y_t contains only error u_{ijt}, while z_t also contains errors u_{klt} for $(k, l) \neq (i, j)$ (if z_t only depends on lagged peer effects, then it is independent of u_{ijt}). Thus, there is uncertainty, which cannot be removed by taking linear combinations.

Remark 3. Theorems 5–8 assume that y_t is strongly mixing and converges to a stationary distribution. This is guaranteed by Theorems 2 and 3 under Assumption 2 (unbounded from below support of errors). We conjecture that if y_t does not converge to a stationary distribution, but has uniformly bounded several first moments[9] and some form of mixing, consistency and asymptotic normality still hold, as long as we replace M_0 with its finite

[8]Linear independence of $1, y, z$ means that there do not exist constants $\lambda_1, \lambda_2, \lambda_3$ such that $\lambda_1 + \lambda_2 y + \lambda_3 z = 0$ almost surely (1 represents a constant random variable).

[9]E.g., see Theorem 1 which guarantees uniformly bounded first moment.

sample version

$$\widehat{M}_0 = \frac{1}{T}\sum_t \begin{pmatrix} 1 & y_t & z_t \\ y_t & y_t^2 & y_t z_t \\ z_t & y_t z_t & z_t^2 \end{pmatrix} \mathbf{1}(\alpha + \beta y_t + \gamma z_t \geq 0).$$

That is, we conjecture that

$$\sqrt{T}\widehat{M}_0^{1/2}\left(\hat{\alpha}_{LAD} - \alpha,\, \hat{\beta}_{LAD} - \beta,\, \hat{\gamma}_{LAD} - \gamma\right)' \xrightarrow[T\to\infty]{d} \frac{1}{2f_u(0)}\mathcal{N}(0, I).$$

4.2. LAD in the explosive case. For the model without peer effects Theorem 3 provides a full classification of the asymptotic behavior of y_t. Using this theorem it is possible to establish consistency of the LAD estimator in the model without peer effects not only under a stationarity assumption, but also for explosive and mean-explosive scenarios. This corresponds to cases when $\beta > 1$ or $\{\beta = 1,\, \alpha + \mathbb{E}u_t \geq 0\}$.[10]

Theorem 9. *Suppose $\gamma = 0$ (no peer effects) and suppose that $(\alpha, \beta) \in \Theta$, where Θ is some compact set in $\mathbb{R}^2$. If u_t is i.i.d. with a continuous and positive density at 0, med$(u_t) = 0$, then the LAD estimator is consistent for $\{\beta = 1,\, \alpha + \mathbb{E}u_t \geq 0\}$ and $\beta > 1$:*

$$\left(\hat{\alpha}_{LAD},\, \hat{\beta}_{LAD},\, \right) \xrightarrow[T\to\infty]{\mathbb{P}} (\alpha,\, \beta)\,.$$

We conjecture that consistency holds for any value of γ. Because the proof of Theorem 9 uses a separate argument depending on the asymptotic behavior of y_t, we cannot extend it to the model with peer effects. To be more precise, when $\gamma = 0$ the model has three types of explosive behavior: exponential growth regime ($\beta > 1$), linear growth regime ($\beta = 1, \alpha + \mathbb{E}u_t > 0$), and Brownian regime ($\beta = 1, \alpha + \mathbb{E}u_t = 0$). Yet, it is unclear what the analogues of these regimes are when $\gamma \neq 0$.

4.3. Discussion: LAD vs OLS. Interestingly, although the model is non-linear we still get consistent L_1 or LAD estimates. Yet, the same approach does not work with L_2 or OLS. The following example illustrates that OLS leads to inconsistency bias, while LAD does not.

[10]Notice that in Theorem 3 we had α instead of $\alpha + \mathbb{E}u_t$. This is due to the fact that we assumed $\mathbb{E}u_t = 0$ in Theorem 3, while now we use a different normalization, med$(u_t) = 0$.

Example 2. Let $y_t = [\alpha + u_t]_+$, $a \geq 0$, $\operatorname{med}(u_t) = \mathbb{E}u_t = 0$. Then the LAD estimate solves

$$\min_{\hat{\alpha}} \sum_t |y_t - \hat{\alpha}|.$$

The solution to the minimization problem is the sample median, that is $\hat{\alpha}_{LAD} = \operatorname{med}(y_1, \ldots, y_T)$. As $T \to \infty$ the sample median converges to the median of the stationary distribution of y_t. The median of y_t equals α, because with probability 0.5 the error u_t is positive, so that $y_t = \alpha + u_t \geq \alpha$, and with probability 0.5 the error u_t is negative, so that either $y_t = 0 \leq \alpha$ or $y_t = \alpha + u_t \leq \alpha$. Thus, with probability 0.5 $y_t \geq \alpha$ and with probability 0.5 $y_t \leq \alpha$. So $\hat{\alpha}_{LAD} \xrightarrow[T\to\infty]{\mathbb{P}} \operatorname{med}(y) = \alpha$.

However, the results are different if we minimize an L_2 norm instead of an L_1 norm ignoring positive part. The solution to

$$\min_{\hat{\alpha}} \sum_t (y_t - \hat{\alpha})^2$$

is the sample mean, $\hat{\alpha}_{OLS} = \frac{1}{T}\sum_t y_t$. As $T \to \infty$ the sample mean converges to the actual mean of y_t, so that

$$\hat{\alpha}_{OLS} \xrightarrow[T\to\infty]{\mathbb{P}} \mathbb{E}y_t = \mathbb{E}[\alpha + u_t]_+ = \int_{-\alpha}^{\infty} (\alpha + u) f_u(u) du = \alpha(1 - F(-\alpha)) + \int_{-\alpha}^{\infty} u f_u(u) du \neq \alpha.$$

The intuition is that median is more robust to truncation at zero: if the median of a process is positive, it does not matter if we replace negative values with zero and vice versa. Yet, the mean is significantly shifted by such procedure.

In the Supplementary Material we show how to correct the OLS procedure to get consistent estimates. The idea is to use "identification from infinity" approach, cf. Chamberlain (1986).

5. Prediction

As in Section 4, we continue to work separately with each individual edge. After the parameters of the model are estimated, one can do predictions. The model is Markovian (in a state space taking into account H time lags), thus, can be easily used for predictions. This leads us to propose a prediction of the form $\hat{y}_{T+1}^{LAD} = [\hat{\alpha}_{LAD} + \hat{\beta}_{LAD} y_T + \hat{\gamma}_{LAD} z_T]_+$.

We use a rolling window approach to measure predictive power: we choose some number $T' < T$, and estimate the model based on observations $t, \dots, t + T' - 1$. Then we form a forecast $\hat{y}_{t+T'}$ based on the obtained estimates and observed values $\{y_t, \dots, y_{t+T'-1}\}$. For each $t = 1, \dots, T - T'$ we calculate the difference between the predicted out-of-sample value $\hat{y}_{t+T'}$ and the actual value $y_{t+T'}$. We get a measure of how well we can predict the data:

$$R_{abs} = \frac{1}{T - T'} \sum_{t=1}^{T-T'} |y_{t+T'} - \hat{y}_{t+T'}|. \tag{8}$$

The smaller R_{abs} is, the better predictions we have on average. Similarly, we can also sum over all pairs (i, j) to get a prediction measure over the whole network.

There are two reasons why we use absolute deviations, i.e. the L_1 norm and not the more usual L_2 norm. First, as the estimation relies on minimizing the L_1 norm, it is more consistent to also use the same norm to evaluate predictions. Second, the 1-step-ahead optimal prediction in L_2 norm when α, β, and γ are known is $\int_{-\alpha-\beta y_{t-1}-\gamma z_{t-1}}^{\infty} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u) f_u(u) du$. Thus, it crucially depends on the distribution of the error term, f_u, even when there is no parameter uncertainty. However, the 1-step-ahead optimal prediction in L_1 when α, β, and γ are known is $[\alpha + \beta y_{t-1} + \gamma z_{t-1}]_+$. That is, it does not depend on the distribution of the error and is more convenient to work with.[11]

[11]Suppose that α, β, and γ are known. To see that the optimal 1-step-ahead prediction in L_1 is $[\alpha + \beta y_{t-1} + \gamma z_{t-1}]_+$, define $\Delta\hat{y}_{t+1} = \hat{y}_{t+1} - \alpha - \beta y_t - \gamma z_t$ and write

$$\begin{aligned} \int |y_{t+1} - \hat{y}_{t+1}| f(u) du &= \int |[\alpha + \beta y_t + \gamma z_t + u]_+ - \hat{y}_{t+1}| f(u) du \\ &= \int |\max(u, -\alpha - \beta y_t - \gamma z_t) - \Delta\hat{y}_{t+1}| f(u) du. \end{aligned} \tag{9}$$

$$\text{Notice that } \max(u, -\alpha - \beta y_t - \gamma z_t) = \begin{cases} u, & u \geq -\alpha - \beta y_t - \gamma z_t, \\ -\alpha - \beta y_t - \gamma z_t, & u < -\alpha - \beta y_t - \gamma z_t, \end{cases},$$

$$\text{so that } \operatorname{med}(\max(u, -\alpha - \beta y_t - \gamma z_t)) = \begin{cases} 0, & \alpha + \beta y_t + \gamma z_t \geq 0, \\ -\alpha - \beta y_t - \gamma z_t, & \alpha + \beta y_t + \gamma z_t < 0. \end{cases}$$

Because $\arg\min_C \mathbb{E}|v - C| = \operatorname{med}(v)$, minimizing Eq. (9) gives

$$\Delta\hat{y}_{t+1} = \begin{cases} 0, & \alpha + \beta y_t + \gamma z_t \geq 0, \\ -\alpha - \beta y_t - \gamma z_t, & \alpha + \beta y_t + \gamma z_t < 0 \end{cases}$$

and $\hat{y}_{t+1} = [\alpha + \beta y_t + \gamma z_t]_+$.

6. Empirical Application

It is important to understand the process of the formation of international trade networks of various goods. A good forecast about the future amount of trade is often crucial for numerous policy decisions such as, for example, how much ships/planes/etc. to allocate to the transportation of a given good or whether a country should support the production of a given good. One of the key features of trade data is that trade flows between countries have a large number of zeros (i.e., not all countries trade with each other at a given time, see e.g. Table 1 in Dueñas and Fagiolo (2013)). Helpman et al. (2008, Figure I) report that every year between 1970 and 1997 among 158 countries around half of pairs were not trading. Moreover, the more we disaggregate the data on the product level, the more zeros appear. Yet dealing with zeros provides a challenge. Simply ignoring pairs with zero trade flows leads to biased estimates. Thus, a number of approaches have been considered to deal with zeros in trade (see Head and Mayer (2014, Section 5.2) and references therein). One of the common ways is to assume that the data comes from conditional Poisson distribution and apply Pseudo Poisson Maximum Likelihood, PPML (Santos Silva and Tenreyro (2006)). Yet, when one goes to time series, this method becomes inappropriate. To be more specific, one of the forms of PPML assumes that (in our notation)

$$y_t = \exp\left(\alpha + \beta y_{t-1} + \gamma z_{t-1}\right) u_t, \tag{10}$$

where u_t is a multiplicative non-negative error. Thus, due to the presence of the exponent of βy_{t-1} on the right-hand side of Eq. (10), $\beta > 0$ would lead the process to diverge to infinity extremely fast as time goes to infinity. Thus, a different approach is needed to accommodate time-series (large T) data. This motivates us to use trade data to analyze the performance of the techniques developed in the paper.

6.1. Data. We apply our model to monthly exports of pharmaceutical products. The data is obtained from the Eurostat COMEXT database (European Commission (accessed May 1, 2018)). Pharmaceutical industry represents one of the largest industrial sectors in the EU and provides a sizable, positive contribution to the EU trade balance (e.g. see Section II in

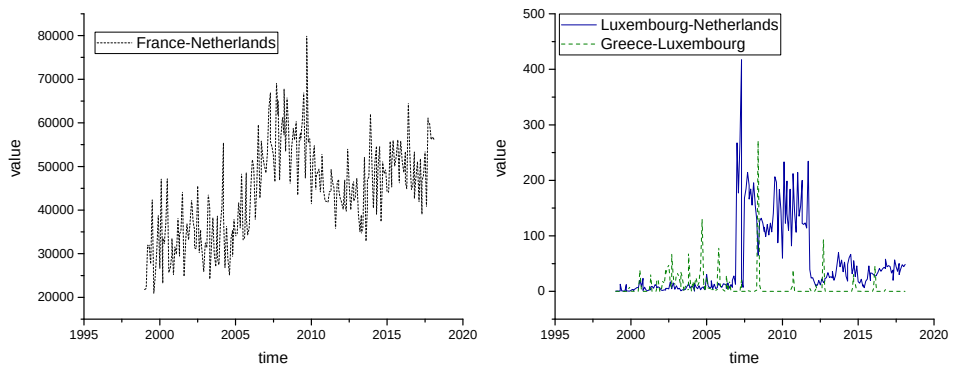

FIGURE 3. Exports of pharmaceutical products between EU countries. Value is in 1000 Euro.

Gambardella et al. (2000) for the discussion based on the Eurostat data). Pharmaceutical industry has a two-digit code 30 as labeled by the Harmonized Commodity Description and Coding Systems (HS). The time period of observations is from January of 1999 until February of 2018, so that $T = 230$. We choose 12 European Union (EU) countries. Those are the countries which joined the EU first. Thus, each pair of countries represents an edge in the network, and we apply the techniques from previous sections to each such edge.

Figure 3 shows exports of pharmaceutical products between France and Netherlands, Luxembourg and Netherlands, and Greece and Luxembourg. We can see that although the first graph does not have zeros, the second has periods of no trade for both pairs of countries. Zero trade is consistent with our model. The longer length of zero intervals (green dashed curve vs. blue straight one in Figure 3) could be explained by more negative α_{ij}.

6.2. Predictions. To measure predictive power we use a rolling window of size 200. Our consistency results rely on $T \to \infty$, thus, we want to have as large as possible window size. Additionally, we need more than one window to see how good are our predictions. This leads us to the choice of window size $T' = 200$, which is large enough and still can be rolled enough times[12]. The increment between successive rolling windows is 1 period. For each window we make a 1-period-ahead forecast. That is, based on observations from t to $t + 199$ we predict what happens at $t + 200$ for $t = 1, \dots, 30$. This gives us 30 overlapping windows and, thus, 30 forecasts. We report the average predictive absolute error over them. For testing in the

[12] $T' = 195$ or $T' = 205$ would be as good and the precise value of T is ad hoc in this sense.

Model and method	R_{abs} in billion euros	
	window size $T' = 200$	window size $T' = 115$
LAD w. p.e.	1.8936	1.8315
MLE w. p.e.	1.9215	1.8274
OLS w. p.e.	1.9217	1.8279
LAD w/o p.e.	1.9865	1.8977
MLE w/o p.e.	2.0036	1.8847
OLS w/o p.e.	2.0037	1.8848
"today"	2.0316	1.9096

TABLE 1. Prediction errors with and without peer effects under different estimation techniques. Smaller numbers mean better quality of prediction. $R_{abs} = \frac{1}{T-T'} \sum_{i,j} \sum_{t=1}^{T-T'} |y_{ij,t+T'} - \hat{y}_{ij,t+T'}|$, where $y_{ij,t+T'}$ is the true value and $\hat{y}_{ij,t+T'}$ is the out-of-sample prediction.

next subsection we need more windows and with that aim in mind we also report results for a rolling window of size 115, which gives 115 overlappind windows (from t to $t + 114$ for $t = 1, \ldots, 115$). The reason is that successive windows are very correlated, and to do testing we want to diminish the effect from correlation. Thus, we need more windows over which the test is going to average. In other words, there is a trade-off between maximizing the window size to improve the quality of estimates and maximizing the number of windows (hence, decreasing the size of each) to improve the statistical significance of testing. Without any additional information on this trade-off, we choose the values of two parameters to be equal, leading to window size $T' = 115$. It is impossible to efficiently present the results for all 132 edges individually, so instead we sum over all edges and report the total result.

We consider three alternative benchmark predictions to evaluate the performance of our model and estimation and forecasting techniques. The first benchmark is the naïve forecast "today equals tomorrow", i.e., $\hat{y}_{t+1} = y_t$. The second alternative is to ignore positivity and treat the model as linear, i.e., $y_t = \alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t$ (OLS in Table 1). Finally, the third competing option is to stick to our model, but to use Gaussian likelihood instead of the LAD (MLE in Table 1). As is shown in the Supplementary Material, in our setting with positive part MLE is consistent when errors are normally distributed (Theorem C.1), yet when the error distribution is far from normal, MLE can perform poorly.

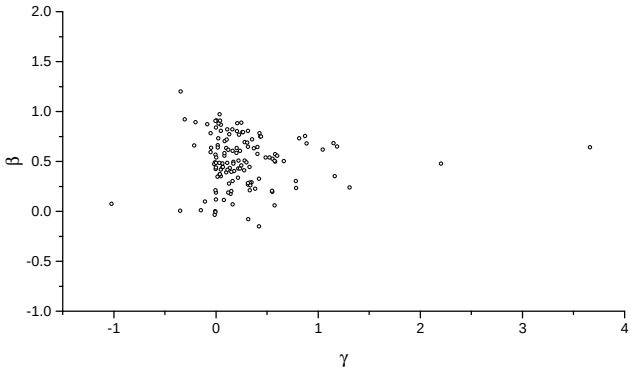

FIGURE 4. LAD estimates for 132 country pairs for the model $y_t = [\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t]_+$.

We compare the above three approaches with the LAD estimator with and without peer effects. Thus, both different models and different estimation techniques are compared. Results are shown in Table 1, where "w. p.e" stays for the model with peer effects (regressor z_{t-1} is included) and "w/o p.e" stays for the model without peer effects (regressor z_{t-1} is excluded). Let us concentrate on the first column corresponding to the larger rolling window $T' = 200$. We can see that even if we do not incorporate peer effects (last four rows) model-based predictors outperform all alternatives. The fact that LAD performs significantly better than the MLE suggests that the error distribution may be far from normal.

When we incorporate peer effects, we use the triangular peer effect function: $z_{ijt-1} = p_{ij}\left(\{y_{k\ell t-1}\}_{k,\ell}\right) = \sum_k \frac{\sqrt{y_{ikt-1}y_{kjt-1}}}{n-2}$. The first column of Table 1 shows that adding peer effects reduces the prediction error both under LAD and MLE estimation approaches. Same happens if we ignore positivity, assume linear model, and estimate it by OLS (first three rows vs. next three rows). This suggests the presence of peer effects in the data. The scatter plot of the LAD estimates of $(\beta_{ij}, \gamma_{ij})$ is shown in Figure 4. The mean value of β is 0.5037 and the mean value of γ is 0.2652.

In general, including irrelevant regressors in the model leads to more noise when one does prediction. Yet, one can expect that the use of additional relevant regressors may improve the prediction results reported in Table 1. As both columns of Table 2 show, using four lags and one peer effect regressor or using five lags gives the best results ($z_{ijt-1} = \sum_k \frac{\sqrt{y_{ikt-4}y_{kjt-4}}}{n-2}$). Also using peer effects evaluated at $t-4$ leads to better performance than using peer effects

Estimated model	R_{abs} in billion euros	
	window size $T' = 200$	window size $T' = 115$
5 lags	1.6731	1.6650
4 lags+p.e. at $t-4$	1.6733	1.6437
4 lags+p.e. at $t-1$	1.6797	1.6511
1 lag+p.e. at $t-4$	1.8569	1.8172
1 lag+p.e. at $t-1$	1.8936	1.8315
1 lag	1.9865	1.8977
"today"	2.0316	1.9096

TABLE 2. Prediction errors with different lags and peer effects. Smaller numbers mean better quality of prediction. $R_{abs} = \frac{1}{T-T'} \sum_{i,j} \sum_{t=1}^{T-T'} |y_{ij,t+T'} - \hat{y}_{ij,t+T'}|$, where $y_{ij,t+T'}$ is the true value and $\hat{y}_{ij,t+T'}$ is the out-of-sample prediction.

evaluated at $t-1$. This suggests that peer effects can help to predict the future, though the optimal functional form of the peer effect function is unclear.

Finally, let us discuss the last columns in Tables 1 and 2 corresponding to the smaller window size $T' = 115$. We see that all methods behave much worse than for $T' = 200$: the improvement of R_{abs} upon the prediction "today equals tomorrow" shrinks. This suggests that the available T is at the border of applicability of our methods, yet the importance of peer effects remains visible. To see the value of peer effects better, we also report the relative performance of LAD with and without peer effects (at $t-4$) for each pair of countries in Table 3.[13] For more than two thirds of all edges the prediction improves because of peer effects. We reinforce the validity of this result in the next subsection.

6.3. Tests. When we were analyzing the performance of our techniques and comparing predictive errors (Tables 1 and 2), we averaged over all edges of the networks. In the current subsection, we look at how good the prediction is for each individual edge. For each edge we compute the Diebold-Mariano test (Diebold and Mariano (1995)) to compare the predictive accuracy under different model specifications. Here we use the rolling window of size 115, as the test requires a large number of predictions. This gives us 115 overlapping windows (from t to $t+114$ for $t = 1, \ldots, 115$). We consider 1-period-ahead forecasts under various models. We compare the model without peer effects against the benchmark prediction

[13]See Supplementary Material for a similar table with peer effects at $t-1$.

$\frac{R_{w.p.e.}-R_{w/o\ p.e.}}{R_{w/o\ p.e.}}$

From \ To	FR	NL	DE	IT	UK	IE	DK	EL	PT	ES	BE	LU
FR	·	0.08	-0.02	0.01	0.01	-0.03	0.00	0.02	0.05	0.06	-0.01	0.01
NL	-0.16	·	-0.39	-0.07	-0.08	0.02	-0.03	-0.11	-0.15	-0.12	-0.03	0.02
DE	-0.02	-0.02	·	-0.23	-0.06	-0.01	-0.08	-0.09	-0.15	-0.08	0.16	-0.02
IT	-0.12	-0.07	-0.13	·	-0.04	-0.04	-0.17	-0.16	-0.03	-0.09	-0.03	-0.08
UK	-0.01	-0.05	-0.19	-0.01	·	0.06	0.01	-0.03	0.04	0.01	-0.05	0.05
IE	-0.00	-0.15	0.00	-0.22	-0.12	·	-0.06	-0.10	-0.10	0.03	-0.01	0.04
DK	-0.13	-0.13	-0.06	-0.12	-0.05	-0.03	·	-0.07	-0.02	-0.14	0.04	0.09
EL	-0.12	0.09	-0.06	-0.04	-0.04	0.02	-0.02	·	-0.05	-0.00	-0.01	-0.01
PT	-0.11	-0.23	-0.08	-0.14	-0.14	0.02	0.03	-0.02	·	-0.06	-0.10	-0.13
ES	-0.06	-0.19	-0.05	-0.16	-0.01	-0.08	0.02	-0.17	-0.10	·	-0.00	-0.01
BE	-0.04	0.06	0.18	0.00	-0.11	0.05	0.01	-0.05	-0.10	0.02	·	-0.03
LU	0.01	0.16	0.03	0.06	-0.17	0.24	-0.16	0.04	-0.07	-0.04	0.01	·

TABLE 3. Performance of LAD with peer effects (at $t-4$) relative to LAD without peer effects for all country pairs. Negative numbers mean LAD with peer effects is better, positive numbers mean LAD without peer effects is better. For each pair of countries (i,j) we calculate $R_{w.(w/o)p.e.} = \frac{1}{T-T'}\sum_{t=1}^{T-T'}|y_{ij,t+T'}-\hat{y}_{ij,t+T'}|$, where $\hat{y}_{ij,t+T'}$ is the prediction based on LAD with (without) peer effects and window size is $T'=115$.

"today=tomorrow" and against the model with peer effects. There are 132 edges in total, and generally for between one third and one half of them we reject the null of identical predictive accuracy. This is shown in Table 4. Similarly, Table 5 shows the results of the one-sided Diebold-Mariano tests for different model specifications. Almost all of the rejections in the former table correspond to rejections in the latter table.[14] That is, when we compare two possible model specifications, the one with the positive part and peer effects dominates. This strengthens the importance of incorporating positivity and peer effects. On the other hand, testing the model with 5 lags versus the model with peer effects seems to show much less difference in predictability.

Additionally, for the model with 1 lag and peer effects evaluated at time $t-1$ we calculate t-statistics to analyze significance of the peer effects for each edge. That is, for each edge $(i \to j)$ we calculate $\frac{\sqrt{T}\hat{\gamma}_{ij}}{\widehat{std}(\gamma_{ij})}$. We use asymptotic variance from Theorem 6 to estimate $\widehat{std}(\gamma_{ij})$.

[14]One needs to compare the 10% column of the two-sided test with the 5% column of the one-sided test.

	Number of rejections	
	5%	10%
"today" vs. 1 lag	38	47
"today" vs. 1 lag+p.e. at $t-1$	45	59
"today" vs. 1 lag+p.e. at $t-4$	53	60
1 lag vs. 1 lag+p.e. at $t-1$	39	51
1 lag vs. 1 lag+p.e. at $t-4$	49	60
5 lags vs. 4 lags+p.e. at $t-1$	16	22
5 lags vs. 4 lags+p.e. at $t-4$	15	27

TABLE 4. Two-sided Diebold-Mariano test of equal predictability.

	Number of rejections	
	5%	10%
"today" worse than 1 lag	41	51
"today" worse than 1 lag+p.e. at $t-1$	52	62
"today" worse than 1 lag+p.e. at $t-4$	57	65
1 lag worse than 1 lag+p.e. at $t-1$	38	46
1 lag worse than 1 lag+p.e. at $t-4$	50	57
5 lags worse than 4 lags+p.e. at $t-1$	11	21
5 lags worse than 4 lags+p.e. at $t-4$	21	23

TABLE 5. One-sided Diebold-Mariano test.

The matrix of second moments M_0 is estimated by its sample analogue, while density at zero, $f_{u_{ij}}(0)$, is estimated by uniform kernel. The bandwidth is chosen so that approximately half of the observations are used. On 5% level peer effects are significant for approximately two thirds of edges. If we instead consider peer effects evaluated at time $t-4$, the number of significant coefficients goes slightly up (92 out of 132 versus 83 out of 132). This reinforces results of the Diebold-Mariano tests reported in Table 5.

7. Conclusion

This paper presents a novel approach to modeling and estimating non-negative time series with interactions (peer effects). Estimation does not require knowledge of the error distribution, thereby making the whole process more attractive to use. In our leading example, weighted networks, we treat them as multivariate time series. Instead of using the variation across individuals, we use variation across time to identify parameters of the network. The

main advantage is that we allow the realization of each edge today to depend on the whole structure of the graph in the previous time periods, and not only on the properties of two nodes, which are connected by that edge. Moreover, the Markov form of the equations makes them convenient for doing predictions. As the empirical example suggests, our model does, indeed, help to predict the future. Overall, the results confirm that incorporating non-negativity of the dependent variables into the model matters and incorporating peer effects leads to the improved predictive power.

In the future it would be interesting to apply the model to different data sets. Phone call data for a small group of individuals and technology adoption by countries from one another seem like natural candidates.

From a theoretical point of view, it would be interesting to investigate in more depth the discontinuity in the asymptotics in the model without peer effects. The behaviour of the process $y_t = [y_{t-1} + u_t]_+$ differs dramatically from what one gets by shifting α from zero or β from one slightly. Thus, finding a way to unify the cases in the neighbourhood of the point $(\alpha = 0, \beta = 1)$ in the spirit of Phillips (1987) may be helpful from a practical point of view. Yet, it is a challenging problem. The proper scaling limit of such a process is complicated, because it involves the computation of the time the process spends at zero in the limit.

8. Supplementary Material

See "Evolution of Networks: Supplementary Material" for all proofs, additional tables, game theoretical justification of Eq. (3), and results on OLS and MLE estimation.

References

Amemiya, T., "Tobit models: A survey," *Journal of Econometrics*, 1984, *24* (1–2), 3–61.

Anderson, T.W., "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations," *The Annals of Mathematical Statistics*, 1959, *30* (3), 676–678.

Andrews, D.W.K., "Non-Strong Mixing Autoregressive Processes," *Journal of Applied Probability*, 1984, *21* (4), 930–934.

Blume, L.E., W.A. Brock, S.N. Durlauf, and Y.M. Ioannides, *Identification of social interactions*, Vol. 1, North-Holland, 2011. In Handbook of social economics.

Bramoullé, Y., A. Galeotti, and B. Rogers, eds, *The Oxford handbook of the economics of networks*, Oxford University Press, 2016.

Cavaliere, G., "The Asymptotic Distribution of the Dickey–Fuller Statistic under Nonnegativity Constraint," *Econometric Theory*, 2004, *20* (4), 808–810.

Chamberlain, G., "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 1986, *32* (2), 189–218.

de Jong, R. and A.M. Herrera, "Dynamic censored regression and the Open Market Desk reaction function," *Journal of Business & Economic Statistics*, 2011, *29* (2), 228–237.

Debaly, Z.M. and L. Truquet, "Iterations of dependent random maps and exogeneity in nonlinear dynamics," *Econometric Theory*, 2021.

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

_____ **and** _____, *Financial and macroeconomic connectedness: A network approach to measurement and monitoring*, Oxford University Press, 2015.

_____ **and R.S. Mariano**, "Comparing predictive accuracy," *Journal of Business & economic statistics*, 1995, *13* (3), 253–263.

Dueñas, M. and G. Fagiolo, "Modeling the international-trade network: a gravity approach," *Journal of Economic Interaction and Coordination*, 2013, *8* (1), 155–178.

European Commission, "Eurostat," `https://ec.europa.eu/eurostat/` accessed May 1, 2018.

Gambardella, A., L. Orsenigo, and F. Pammolli, "Global competitiveness in pharmaceuticals: a European perspective," *MPRA Paper No.* 15965, 2000.

Graham, B.S., "Homophily and transitivity in dynamic network formation," *Working paper*, 2016.

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

———, "Dyadic regression," *Working paper*, 2019.

Hahn, J. and G. Kuersteiner, "Stationarity and mixing properties of the dynamic Tobit model," *Economics Letters*, 2010, *107* (2), 105–111.

Head, K. and T. Mayer, *Gravity equations: Workhorse, toolkit, and cookbook*, Vol. 4, North Holland, 2014. In Handbook of international economics.

Helpman, E., M. Melitz, and Y. Rubinstein, "Estimating trade flows: Trading partners and trading volumes," *The quarterly journal of economics*, 2008, *123* (2), 441–487.

Holme, P. and J. Saramäki, "Temporal networks," *Physics reports*, 2012, *519* (3), 97–125.

Khan, S. and J.L. Powell, "Two-step estimation of semiparametric censored regression models," *Journal of Econometrics*, 2001, *103* (1–2), 73–110.

Leung, M.P. and H.R. Moon, "Normal Approximation in Large Network Models," *arXiv preprint arXiv:*1904.11060, 2019.

Michel, J. and R.M. de Jong, "Mixing properties of the dynamic Tobit model with mixing errors," *Economics Letters*, 2018, *162*, 112–115.

Palla, G., A.L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, 2007, *446* (7136), 664–667.

Phillips, P.C.B., "Towards a unified asymptotic theory for autoregression," *Biometrika*, 1987, *74* (3), 535–547.

Powell, J.L., "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, 1984, *25* (3), 303–325.

Robinson, D.T. and T.E. Stuart, "Network effects in the governance of strategic alliances," *The Journal of Law, Economics, & Organization*, 2006, *23* (1), 242–273.

Silva, J.M.C. Santos and S. Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Wang, X. and J. Yu, "Limit theory for an explosive autoregressive process," *Economics Letters*, 2015, *126*, 176–180.

Wei, S.X., "A Bayesian approach to dynamic Tobit models," *Econometric Reviews*, 1999, *18* (4), 417–439.

White, J.S., "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *The Annals of Mathematical Statistics*, 1958, *29* (4), 1188–1197.

Withers, C.S., "Central limit theorems for dependent variables. *I*," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1981, *57* (4), 509–534.

(Anna Bykhovskaya) University of Wisconsin-Madison

E-mail address: anna.bykhovskaya@wisc.edu