EVOLUTION OF NETWORKS: PREDICTION AND ESTIMATION

ANNA BYKHOVSKAYA

Abstract. The paper introduces a model for the evolution of weighted networks. It views a network as a multivariate time series, where each coordinate represents a given edge across time. The number of time periods is treated as large compared to the size of the network. The model is nonparametric with respect to the distribution of the errors and specifies the temporal evolution of a weighted graph that combines classical autoregression with non-negativity, a positive probability of vanishing, and peer effect interactions between weights assigned to edges in the process. As a result, the model is nonlinear in multiple ways.

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. The asymptotic theory is more complex than that of the classical autoregression.

The paper provides an empirical implementation of the approach to monthly trade data in European Union.

Date: November 10, 2018.

The author would like to thank Don Andrews, Vadim Gorin, Peter Phillips, and Larry Samuelson for valuable comments, suggestions, and encouragement throughout the duration of the project.

Contents

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 edges, then we are dealing with a basic one-dimensional time series). 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, technology adoption by firms or countries is a non-negative variable, which evolves over time. Macroeconomic variables such as exchange rates, T-bill rates and so on are non-negative by definition. Finally, if we look at the number of passengers travelling from country A to country B at a given day, it 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 can not be captured by linear models such as GARCH.

Moreover, we look not only at a one-dimensional time series, but also at a multivariate ones. For instance, instead of considering the traffic flow from one given country to another, we can track the whole set of countries, and get a non-negative weighted network, which governs the amount of travel. In the same spirit, we can construct a time-varying network of trade between different countries or firms or a social network, which measures the amount of communication between people. In this paper we focus on networks (or weighted graphs) as a main example of a non-negative structure. In contrast to other cases, networks may potentially involve interactions between the coordinates corresponding to their edges. For example, 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 can not 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. (For more examples on the role of social and economic networks in real life see, for example, Jackson (2010).)

Often one may need to be able to predict a structure of a network as it evolves. It may be important to know whom is better to target with ads, election campaigns, or other type of information. Countries may need to know the future trading patterns to decide production of which types of goods to support. We also refer to Holme and Saramäki (2012) for an interdisciplinary review highlighting the importance of the temporal structure of networks. 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. When one goes to the data, 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. In this paper we deal with weighted networks, so that each edge is associated with a number, which evolves with time. 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 modelling networks as a multivariate non-linear time series, i.e. denoting by y_{ijt} the weight of edge $(i \to j)$,

$$y_{ijt} = [\alpha_{ij} + \beta_{ij} y_{ijt-1} + \gamma_{ij} z_{ijt-1} + u_{ijt}]_+, \tag{1}$$

where $[\cdot]_+$ stands for positive part, α_{ij}, β_{ij}, γ_{ij} are unknown coefficients, which need to be estimated, u_{ijt} is a random error, which is independent across time t, but may be correlated along edges (i, j), and z_{ijt-1} is a peer effect or an interaction term. That is, z_{ijt-1} is a function (assumed to be explicitly known) of past periods of the network and it allows edges to affect each other in the future. Note that we do not impose linearity on z. It may be a non-linear function. Moreover, the positive part per se creates nonlinearity, which leads to technical difficulties. The equation (1) can be obtained as a solution to a certain utility maximization problem, as we outline in Section 2. One can add more regressors to (1) and most of our following results continue to hold in the extended setting (see Remarks 1, 2, and 4). We note that there is a large number of much more sophisticated models for the network evolution in the literature (e.g. Pin and Rogers (2015) where agents play a prisoner's dilemma every period and choose connections based on the observed outcomes), yet the estimation procedures become more advanced and model-specific with each layer of complexity. It becomes preferable to address the basic case, so that the general principles and challenges can be identified.

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

Let us describe our main findings. First, we show a sufficient condition for the model to be stationary. We normalize z_{ijt} in a way which insures that peer effects do not grow faster than their maximal argument (see Section 2.3 for more details). We prove that if $\max(0, \beta_{ij}) + |\gamma_{ij}|$ is uniformly over i, j bounded by some constant smaller than one, then the process is strongly mixing and converges (when started from any initial conditions) to a stationary distribution as $t \to \infty$. For the special case of the model where $\gamma_{ij} \equiv 0$, we provide the full classification of asymptotic behavior of the process: the change from stationary to explosive behavior is at $\alpha_{ij} = 0$, $\beta_{ij} = 1$. At that boundary case, the process rescaled by $\sqrt{T}$ converges to the absolute value of a Brownian motion as $T \to \infty$. One distinction with the usual linear autoregression model is that the $\alpha_{ij} < 0$, $\beta_{ij} = 1$ case is stationary due to presence of positive parts. We remark that the non-linearities created by $[\cdot]_+$ and by z_{ijt} leads to technical difficulties. In particular, we can not use the contraction mapping theorem directly and need to rely on renewal theory. As far as we know, these results are new; the only relevant papers seems to be Jong and Herrera (2011), Hahn and Kuersteiner (2010), and Michel and de Jong (2018). In these papers sufficient conditions for the existence of a stationary solution to (1) are found for the one-dimensional case with z linearly depending on the past periods, and for possibly correlated across time errors u_t. This particular stationary solution is further shown to be strongly mixing.

Interestingly, 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.

Second, we discuss how to estimate the parameters of the model. We start by showing that ordinary least squares (OLS) estimators are not consistent in our setting. This matches a similar inconsistency for censored regression models, c.f. discussion at the end of Section 4.2 in Amemiya (1984). We also show how to correct the OLS procedure to restore consistency, yet in doing so one needs to ignore a lot of observations, and, thus, the accuracy of the estimation decreases significantly. Further, if we assume the errors u_{ijt} are Gaussian, then we can explicitly write down the likelihood function. We prove that in this case the maximum likelihood estimator (MLE) is consistent. For the semiparametric case when the distribution of u_{ijt} is not specified, there is no guarantee that the Gaussian MLE is consistent and we need to proceed differently. Our approach is then related to the least absolute deviations (LAD) method, which was used in the context of censored regression in Powell (1984). We obtain an estimator by minimizing with respect to α_{ij}, β_{ij}, γ_{ij} the sum of the absolute differences between y_{ijt} and $[\alpha_{ij} + \beta_{ij}y_{ijt-1} + \gamma_{ij}z_{ijt-1}]_+$. Theorems from Powell (1984) are not applicable in our setting because the independence assumptions from that paper do not hold. However, by use of the martingale central limit theorem we establish show consistency and asymptotic normality of the estimator. Some of our results even cover explosive cases and we show that while the large T asymptotics of the system changes drastically, the LAD estimator still provides consistent estimators. This is in line with results on the consistency of OLS in the explosive autoregressive model (see White (1958) and Anderson (1959) for the model without a constant and Wang and Yu (2015) for the model with an intercept.)

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 find that whenever z has non-negative support and the true α_{ij}, β_{ij}, γ_{ij} are all positive, neither of these problems exist in our setting: the optimization problem is convex, asymptotic normality does not rely on any additional conditions and the asymptotic variance of the LAD estimator is given by a simple formula.

We propose an objective quality measurement of the model by looking at the absolute value of the prediction error at time step $t+1$ for the model estimated from the data up to time t. Let us emphasize that the use of absolute values instead of the more standard squared differences is important here: for the latter the optimal prediction (even if α_{ij}, β_{ij}, γ_{ij} are exactly known) depends on the distribution of errors, while for the former we can speak about model-independent prediction power.

We apply our framework to monthly trade data between European Union countries for pharmaceutical products. We use the basic prediction "tomorrow=today" as a benchmark for comparing the prediction power, and the interactions (peer effects) are modelled by a function based on the "friend of my friend is my friend" principle. In our experiments, any model-based estimation techniques lead to an improvement of the power over benchmark case. The basic linear OLS procedure, which ignores the positivity of weights, leads to the worst results among model-based estimations. The MLE estimator performs better, and the LAD estimator leads to the best results. The addition of peer effects z_{ijt} (vs. setting $\gamma_{ij} = 0$ in the model specification) also leads to improved prediction power. Results of the Diebold-Mariano test (Diebold and Mariano (1995)), which are reported in Section 6, 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, so that evolution of each edge is a separate process. The full classification in terms of stationary/explosive behavior is established in this case, as well as identification results. Section 4 discusses estimation of the model, while Section 5 proposes the 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 the Appendix.

2. Model

2.1. Set up. Our model allows both for undirected and directed networks. The main example of the former is a social weighted network, where nodes represent people and edges represent how much time they spend together. E.g., one can use phone call data as a measure of friendship (the more two people text or talk to each other, the closer their relationship is). For the case of a directed network, the applications are mostly for firms or countries and trade between them. Looking separately at exports and imports, we get a directed network.

The equation of interest is

$$y_{ijt} = \left[\alpha_{ij} + \beta_{ij} y_{ijt-1} + \gamma_{ij} p_{ij} \left(\left\{ y_{kls} \right\}_{\substack{\text{k,l=1,...,n} \\ \text{s=t-H,...,t-1}}} \right) + u_{ijt} \right]_+ , \tag{2}$$

where u_{ijt} is a random error, $t = 1, \ldots, T$ stands for time, and $i, j = 1, \ldots, n$ stand for agents (in the undirected case $i < j$, in the directed case $i \neq j$). So that y_{ijt} can be interpreted as either how much i and j talk at time t or the amount of trade from i to j.

In Eq. (2) $\alpha_{ij} \in \mathbb{R}$ represents 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 both 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. We coefficient γ_{ij} captures the dependence on the peer effects/interactions p_{ij}. The function p_{ij} serves as an aggregator of the past structure of the network in a way that affects the current state. Finally, the positive part in Eq. (2) creates nonlinearity and leads to a positive mass at zero.

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

Remark 1. One can replace the coefficient γ_{ij} and the scalar function p_{ij} in Eq. (2) by K–dimensional vectors. Our results will continue to hold in this extended setting, see Remarks 2 and 4.

2.2. Maximization problem. In this subsection we present a stylized game theoretical model that leads to our equation of interest, Eq. (2). It is another justification of Eq. (2) along with the primary one, which is to capture a number of essential properties of networks (non-negativity of edges, positive probability of vanishing of each edge, and interactions between edges, which affect the whole network).

Consider a world with n myopic agents (people/firms/countries/etc.) with quadratic adjustment costs. Agents can interact with each other over time. Time is discrete and goes from 1 to T. Every period, each agent i chooses how much time to spend with or how much to trade with each other agent j. The decision is based on two components: costs and benefits.

Benefits are characterized by a per unit gain of $\alpha_{ij} + u_{ijt}$. Here α_{ij} is a constant, while u_{ijt} is random component that is independent across time. Thus, y units of communication/trade leads to a benefit of $y(\alpha_{ij} + u_{ijt})$.

The second component is a quadratic adjustment cost function. Agents get disutility whenever there are deviations from some target expected level of communication/trade. The target is composed from an own past and a peer-effect or interactions component. The interaction term aggregates the whole structure of the network for up to H periods. That is, we assume that agent i by choosing to devote y units to agent j has to pay

$$\frac{1}{2}\left(y-\beta_{ij}y_{ijt-1}-\gamma_{ij}p_{ij}\left(\left\{y_{kls}\right\}_{\substack{\text{k,l=1,...,n}\\ \text{s=t-H,...,t-1}}}\right)\right)^2,$$

where

$$p_{ij}\left(\left\{y_{kls}\right\}_{\substack{\text{k,l=1,...,n}\\ \text{s=t-H,...,t-1}}}\right):\mathbb{R}_+^{n^2H}\to\mathbb{R}$$

represents peer effects/interactions function, which depends on H previous periods.

The interpretation is that β_{ij} represents a rate at which stock/relationship depreciates/appreciates. If $\beta_{ij}<1$, then the agent is introverted and tends to decrease communication; while if $\beta_{ij}>1$, the agent is an extrovert, who tends to expand communication. The interpretation of β_{ij} for firms corresponds to production depreciation (i.e. technology wears out) or production appreciation (better technology management over time increases production). The coefficient γ_{ij} indexes the sensitivity of the reference level with respect to peer effects/interactions, which are, in turn, represented by the function p_{ij}. The peer effects function depends on H past periods of the whole network, and, thus, captures interactions between different edges y_{kls} across time.

Agents have separate maximization problems for each time period t and with each peer j. Agent i solves the following maximization problem at day t with respect to agent j:

$$\max_{y\geq 0}\left[y(\alpha_{ij}+u_{ijt})-\frac{1}{2}\left(y-\beta_{ij}y_{ijt-1}-\gamma_{ij}p_{ij}\left(\left\{y_{kls}\right\}_{\substack{\text{k,l=1,...,n}\\ \text{s=t-H,...,t-1}}}\right)\right)^2\right]. \tag{3}$$

The solution to the maximization problem (3) is

$$y^*_{ijt}=\left[\alpha_{ij}+\beta_{ij}y_{ijt-1}+\gamma_{ij}p_{ij}\left(\left\{y_{kls}\right\}_{\substack{\text{k,l=1,...,n}\\ \text{s=t-H,...,t-1}}}\right)+u_{ijt}\right]_+,$$

which leads to the network evolution process described by Eq. (2).

2.3. Assumptions. We need to impose some assumptions on the error distribution and on the peer effects function.

Assumption 1. *u_{ijt} is i.i.d. over t for fixed i, j.*

Assumption 2. *The vector $\{u_{ijt}\}_{i,j=1,\ldots,n}$ has support such that $\mathbb{P}(u_{ijt} < -M \;\; \forall i, j) > 0$ for all $M > 0$.*

Assumption 3. *The vector $\{u_{ijt}\}_{i,j=1,\ldots,n}$ has support such that $\mathbb{P}(u_{ijt} > M \;\; \forall i, j) > 0$ for all $M > 0$.*

Assumption 2 implies that the errors are required to jointly take large negative values with positive probability, while Assumption 3 requires them to jointly take large positive values. The former is used to show stationarity, while the latter is used in the identification Theorems 5 and 6.

Eq. (2) 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.

Assumption 4 (Normalization of the mean.)**.** *For all i, j, t, $\mathbb{E}u_{ijt} = 0$.*

Assumption 5 (Normalization of the median.)**.** *For all i, j, t, med$(u_{ijt}) = 0$.*

Alternative Assumptions 4 and 5 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 Assumption 4 and α_{ij}^{med} and $\mathbb{E}u_{ijt}^{\text{med}}$ are an intercept and a mean of the error under Assumption 5.

Assumption 6 (Peer effects do not grow faster than their maximal argument.)**.** *$p : \mathbb{R}_{+}^{n^2 H} \to \mathbb{R}$ is such that there exists a constant $A \in \mathbb{R}$ for which*

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

Let us present some examples of possible peer effect functions.

- *Maximum*:

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

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

- *Minimum*:

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

  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*:

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

$$\text{where } \{\lambda_{klr}\}_{\substack{k,l=1,\dots,n\\ r=1,\dots,H}} \text{ are known and } \lambda_{klr}\geq 0, \sum_{\substack{\text{k,l=1,...,n}\\ \text{r=1,...,H}}} \lambda_{klr} \leq 1.$$

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

- *Triangles*:

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

  This is 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 can not 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, interactions between countries and so on.

2.4. Stationarity. Theorem 1, which is proved in the Appendix, provides sufficient conditions, under which the network does not explode. Non-exploding does not guarantee convergence to a stationary distribution, as formally the process may have cycles. Yet, it is enough for our estimation and prediction approaches to work. Moreover, if we additionally assume that $\{u_{ijt}\}_{i,j}$ has unbounded from below support (Assumption 2), then we can show stationarity (Theorem 2).

Theorem 1. *Suppose that Assumptions 1 and 6 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 the multivariate process $\{y_{ijt} : i, j = 1 \ldots, n\}_{t \geq 1}$ does not explode (i.e. there exists a constant C_1 such that $\mathbb{E}y_{ijt} < C_1 < \infty$ for all i, j, t).*

Definition. The process $\bar{y}_t = \{y_{ijt}\}_{i,j}$ is strongly mixing if for arbitrary Borel sets Δ_1 and Δ_2

$$\lim_{t \to \infty} |\mathbb{P}(\bar{y}_s \in \Delta_1, \bar{y}_{t+s} \in \Delta_2) - \mathbb{P}(\bar{y}_s \in \Delta_1)\mathbb{P}(\bar{y}_{t+s} \in \Delta_2)| = 0.$$

Theorem 2. *Suppose that Assumptions 1, 2, and 6 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 $\{y_{ijt} : i, j = 1 \ldots, n\}_{t \geq 1}$ is strongly mixing and converges to a stationary process.*

Remark 2. In the extended setting of Remark 1, Assumption 6 should hold for each component of the vector p_{ij} and $|\gamma_{ij}|$ in Theorem 2 should be replaced by the sum of the absolute values of the coordinates of γ_{ij}.

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

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 $\beta = 0$, $\gamma = 0.5$ and $z_t := p_{ij}(y_t, \ldots, y_0) = \max(y_t, \ldots, y_0)$, so that Assumption 6 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 $z_t \geq \max(0, u_1, \ldots, u_t)$. Therefore, $z_t \xrightarrow[t \to \infty]{a.s.} \infty$, as the support of u_t is unbounded

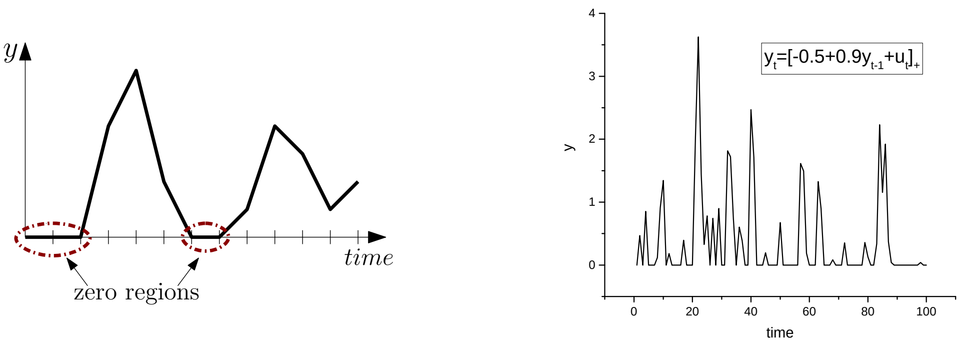

FIGURE 1. A typical sample path for $y_t = [\alpha + \beta y_{t-1} + u_t]_+$, $u_t \sim i.i.d.(0, \sigma^2)$.

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 connection between i and j at time t depends only on its 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

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

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 when $\alpha < 0$.

Theorem 3. *(Classification Theorem) Let Assumption 1 hold. Under the assumptions that u_t has unbounded support from below, $\mathbb{E}u_t = 0$, and $\mathbb{E}u_t^4 < \infty$,*

- *If $\beta < 1$, then y_t is strongly mixing and converges to a stationary process*[1]*;*
- *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: $y_t \xrightarrow{a.s.} \infty$;*
- *If $\beta = 1$, $\alpha > 0$, then y_t is divergent: $y_t \xrightarrow{a.s.} \infty$;*

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

- *If $\beta = 1$, $\alpha = 0$, then y_t is mean-divergent: $\mathbb{E}y_t \to \infty$. The proper scaling limit is*
$$\frac{1}{\sqrt{T}} y_{\lfloor Tr \rfloor} \xrightarrow{d} \sigma |W(r)|,\ r \in [0,1],\ \text{ as } T \to \infty,$$
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.

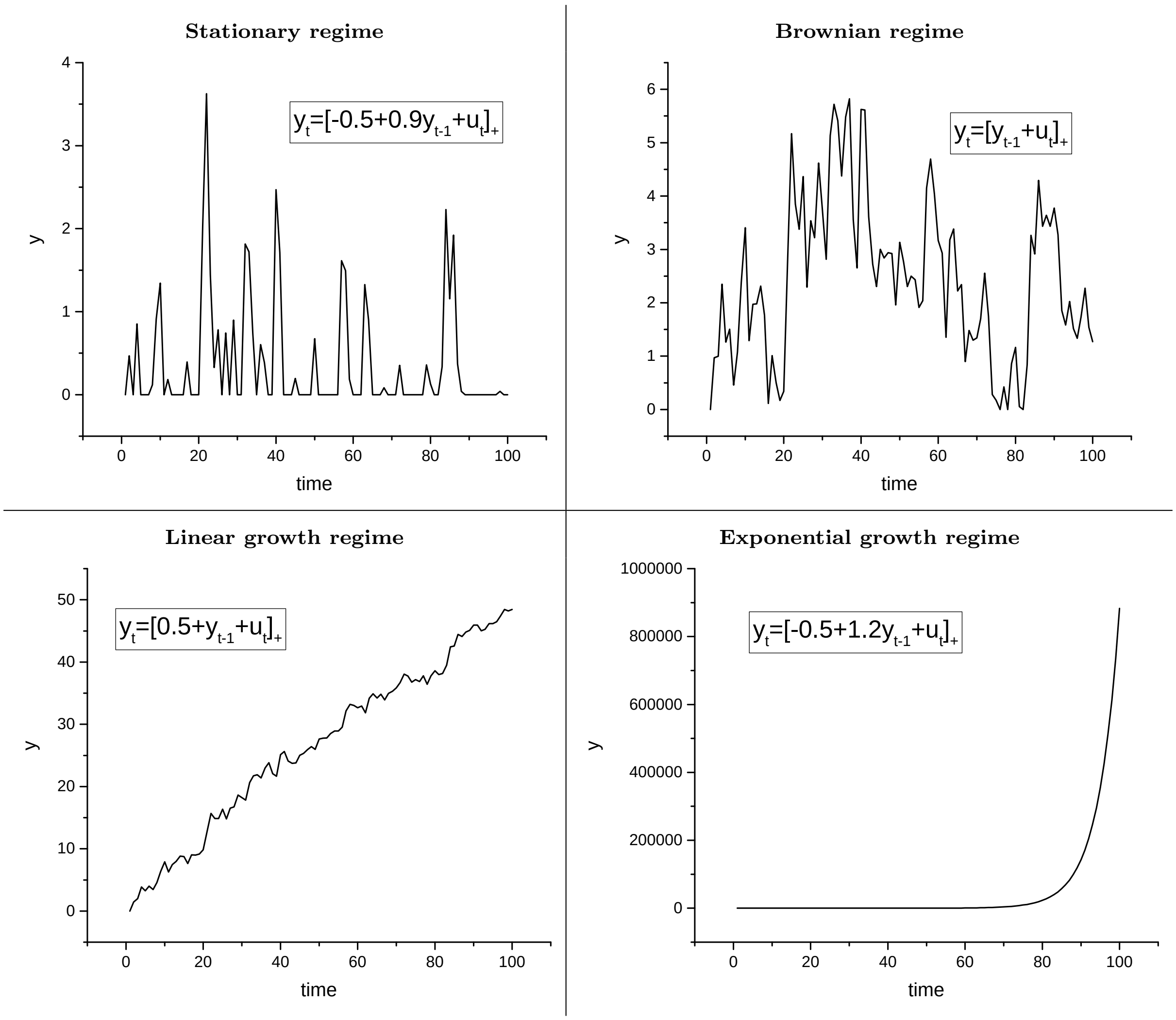

FIGURE 2. Illustration of Theorem 3.

Theorem 3 is proved in the Section A of the Appendix (Theorems A.4, A.5, A.6, and A.8). 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, when $\alpha = 0, \beta = 1$ we get a standard unit-root process while y_t is positive. This process always hits zero. However, it does not converge to a stationary distribution, and as Theorem 3 shows, the process $y_t = [y_{t-1} + u_t]_+$ has exploding mean. Thus, there is a discontinuity between the stationary and explosive regions. This is similar to what is observed in the classical linear autoregressive case.

The limiting distribution in the stationary case of Theorem 3 is complicated and can not be written explicitly as a function of α, β, and the distribution of u. Usually, it can only be obtained numerically. However, what can be calculated is the expected time the process y_t spends at zero once it hits zero. As shown in Lemma 4, it equals $\frac{1}{1-F_u(-\alpha)}$, where F_u is the cumulative distribution function of u. Thus, the smaller is α, the longer are the intervals of zeros.

Lemma 4. *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

$$
\begin{aligned}
&1 \cdot (1 - F_u(-\alpha)) + 2 \cdot F_u(-\alpha)(1 - F_u(-\alpha)) + 3 \cdot F_u^2(-\alpha)(1 - F_u(-\alpha)) + \dots \\
&= (1 - F_u(-\alpha)) \sum_{k=1}^{\infty} k F_u^{k-1}(-\alpha) = \frac{1}{1 - F_u(-\alpha)}.
\end{aligned}
$$

□

3.1. Identification.

In this subsection we show that the parameters of the model, α and β, are identified when $\beta > 0$. This is done in the spirit of Chamberlain (1986). However, in our case, unbounded support of the regressor, y_{t-1}, is satisfied automatically as long as the error has unbounded support.

We provide identification results for each of our normalization assumptions for non-negative β. First, suppose the mean of the errors is normalized.

Theorem 5 (Identification when $\mathbb{E}u_t = 0$.)**.** *Suppose Assumptions 1, 3, and 4 hold and $\beta \geq 0$. Then the coefficient β is always identified, and the coefficient α is identified only when $\beta > 0$.*

The intuition for the non-identifiability of α when $\beta = 0$ is that when $y_t = [\alpha + u_t]_+$, the error distribution below $-\alpha$ is not observed, so it can be defined arbitrarily. That is, fixing $\mathbb{E}u_t = 0$ is no longer meaningful.

If instead of normalizing the mean we normalize the median, we can expand the set of identified values. Yet, as shown in Theorem 6, $\alpha < 0$ is still unidentified when $\beta = 0$.

Theorem 6 (Identification when $\mathrm{med}(u_t) = 0$.). *Suppose Assumptions 1, 3, and 5 hold and β is non-negative. Then the coefficient β is always identified, and the coefficient α is identified only when $\beta > 0$ or when $\alpha \geq 0$. When $\beta = 0$ and $\alpha < 0$, any value $\alpha' < 0$ can rationalize the data.*

Here the intuition for non-identifiability is similar to the previous case. If $y_t = [\alpha + u_t]_+$, then the error distribution below $-\alpha$ is not observed, so it can be defined arbitrarily. Moreover, when $\alpha < 0$, $\mathbb{P}(u > -\alpha) < \mathbb{P}(u > 0) = 0.5$, and the distribution below $-\alpha$ can be adjusted so that the median can be any number less than $-\alpha$. That is, fixing $\mathrm{med}(u_t) = 0$ is not meaningful if more than half of u_t's distribution is unobserved.

Both Theorems 5 and 6 imply that when β is small the estimator of α will be poor.

Remark 3. The case $\beta < 0$ is technically more complicated and is related to singularity/non-singularity of some matrix. To be more specific, see conditions for consistency in Theorem 8.

4. Estimation

We now return back to the general model summarized in Eq. (2). The two main difficulties in our model are the interactions between the outcome variables and the non-separable errors. We can overcome the former by estimating the following model

$$y_t = [\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t]_+. \tag{5}$$

To be more specific, for each pair (i, j) we define

$$z_{ijt-1} = p_{ij}\left(\left\{y_{kls}\right\}_{\substack{\text{k,l=1,...,n} \\ \text{s=t-H,...,t-1}}}\right)$$

and ignore the fact that z_{ijt} are functions of lags of y_{klt}, $k, l = 1, \ldots, n$. That is, after we calculate the values of z_{ijt}, we are not going to use the fact that those values were obtained from $\{y_{klt}\}_{k,l=1,\ldots,n}$ and its lages. Instead we treat z_{ijt} as any other regressors. For each edge $(i \to j)$ we separately estimate Eq. (5) with $y_t = y_{ijt}$, $z_t = z_{ijt}$.

The natural question is whether we lose predictive power by treating each equation independently or not. Generally it is not clear and the answer should depend on the class of estimators we consider. Yet, there is a reason to believe that we do not lose a lot. Gourieroux and Monfort (1980) show that for the linear models like vector autoregressions if the errors are independent across components of y_t (in our case across edges) or the matrices of regressors span the same subspace for each component of y_t, then single-equation generalized least squares (GLS) is equivalent to overall GLS.

In the following subsections we present three different approaches to estimating the model (5) and discuss the strengths and weaknesses of the estimators. In all of them, unless otherwise noted, we assume that y_t is strongly mixing and converges to a stationary distribution, as in Theorems 2 and 3.

Remark 4. 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 $z^1_{t-1}, \dots, z^K_{t-1}$

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

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 8 and 10 is a $K+2$ by $K+2$ matrix composed of all second moments of the vector $(1, y_t, z^1_t, \dots, z^K_t)\mathbf{1}(\alpha + \beta y_t + \sum_{k=1}^{K} \gamma^k z^k_t \geq R)$.

4.1. $\mathbf{L_1}$ Estimation. For this and the following subsection we assume that errors have strictly positive density at zero, $f_u(0)$.

The least absolute deviations (LAD) estimator is the solution to the following minimization problem

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

The LAD estimation procedure for the case of censored regression was first proposed by Powell (1984). He considers a cross-section model $y_t = [x_t'\beta + u_t]_+$. We can not directly use his proofs, as they require u_t to be independent of all x_s, which does not hold for $s > t$ in an autoregressive model.

The minimization problem (6) is convex when $a > 0$, $b \geq 0$, $c \geq 0$, and $z \geq 0$. Thus, the numerical solution is a global maximum when the true $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$. It turns out, as Theorems 7 and 9 show, that the LAD estimators are consistent and asymptotically normal.

Moreover, $[a + by_{t-1} + cz_{t-1}]_+ \equiv a + by_{t-1} + cz_{t-1}$ when a is positive and b, c and z_{t-1} are non-negative (y is non-negative by assumption). Thus, in this case the positive part in the minimization problem never binds, and the additional complicated condition on the stationary distribution and true parameter values from Powell (1984) does not arise in our case.

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

Theorem 7. *When $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$, the peer effect function is non-negative (i.e., $z_t \geq 0$ for all t), u_t has a continuous density at 0, and the random variables $1, y, z$ are linearly independent, the LAD estimator is consistent:*

$$\begin{pmatrix} \hat{\alpha}_{LAD} \\ \hat{\beta}_{LAD} \\ \hat{\gamma}_{LAD} \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \text{ as } T \to \infty.$$

Remark 5. *The linear independence condition 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 can not be removed by taking linear combinations.*

If we do not impose positivity, we still can get consistency if the matrix M_R is nonsingular for some $R > 0$, where

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

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 $\Delta = 0$. When $\alpha = 0$ the latter condition indeed causes a problem, because $y = 0$ with positive probability. If z is also zero, then the indicator jumps depending on the sign of Δ. On the other hand if $\alpha > 0$, $\beta, \gamma \geq 0$ and $z > 0$ with probability 1, then this function is identical 1 and we return to the setting of Theorem 7.

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. When $\beta > 0$, the peer effect function $p_{ij}(\cdot)$ does not depend on y_{ijt}, and the random variable z is non-constant, 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 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 8. *Suppose that $(\alpha, \beta, \gamma) \in \Theta$, where Θ is some compact space in $\mathbb{R}^3$. When M_R is nonsingular for some $R > 0$ 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 $\Delta = 0$, the LAD estimator is consistent:*

$$\begin{pmatrix} \hat{\alpha}_{LAD} \\ \hat{\beta}_{LAD} \\ \hat{\gamma}_{LAD} \end{pmatrix} \xrightarrow[T \to \infty]{\mathbb{P}} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}.$$

Remark 6. Note that if $R_1 > R_2$ and M_{R_1} is nonsingular, then M_{R_2} is also nonsingular. Similarly, 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 $R > \bar{R}$.

Remark 7. The positivity 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 8, we only need $\min\left(\int\limits_{-\tau}^{0} (\tau + u) dF_u(u), \int\limits_{0}^{\tau} (\tau - u) dF_u(u)\right)$ to be positive for any $\tau > 0$, which is satisfied in this case.

Asymptotic normality in the positive case does not require any additional conditions, as is shown in Theorem 9. This is in line with consistency result for the positive case (Theorem 7).

Theorem 9. *When $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$, the peer effect function is non-negative (i.e., $z_t \geq 0$ for all t), and u_t has a continuous density at 0, and the random variables $1, y, z$ are linearly independent, the LAD estimator is asymptotically normal:*

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

Similar to the consistency result with general α, we get asymptotic normality if the random function

$$\Delta \mapsto \mathbf{1}(\alpha + \beta y + \gamma z + \Delta > 0)$$

is continuous with probability at the point corresponding to the true parameter values and $\Delta = 0$.

Theorem 10. *Suppose that $(\alpha, \beta, \gamma) \in \Theta$, where Θ is some compact space in $\mathbb{R}^3$. When 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 $\Delta = 0$, the LAD estimator is asymptotically normal:*

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

Remark 8. Note that we do not impose a condition on the existence of nonsingular M_R for some $R > 0$ in Theorem 10. This is because non-degeneracy of M_0 together with continuity of the indicator $\mathbf{1}(\alpha + \beta y + \gamma z + \Delta > 0)$ implies the existence of a (possibly small) $R > 0$, for which M_R is also nonsingular.

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 stationary distribution, but also for explosive and mean-explosive scenarios. This corresponds to cases when $\beta > 1$ or $\beta = 1$, $\alpha \geq 0$.

Theorem 11. *Suppose $\gamma = 0$ (no peer effects) and suppose that $(\alpha, \beta) \in \Theta$, where Θ is some compact space in $\mathbb{R}^2$. Then if u_t has a continuous density at 0, the LAD estimator is consistent for $\beta = 1$, $\alpha \geq 0$ and $\beta > 1$:*

$$\begin{pmatrix} \hat{\alpha}_{LAD} \\ \hat{\beta}_{LAD} \end{pmatrix} \xrightarrow[T\to\infty]{\mathbb{P}} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

We conjecture that consistency holds for any value of γ. Because the proof of Theorem 11 uses a separate argument depending on the asymptotic behavior of y_t, we can not 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 ($\beta = 1, \alpha > 0$), and Brownian regime ($\beta = 1, \alpha = 0$). The process y_t under different regimes is shown in Figure 2. Yet, it is unclear what the analogue of these regimes is when $\gamma \neq 0$.

4.3. Discussion: LAD vs OLS. Interestingly, although the model is non-linear we still can treat it as linear to 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.

Example 2. Suppose $y_t = [\alpha + u_t]_+$, $a \geq 0$, $\text{med}(u_t) = \mathbb{E}u_t = 0$. 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 $\hat{\alpha}_{LAD} = \text{med}(y_1, \ldots, y_T)$. 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 $\hat{\alpha}_{LAD} \xrightarrow[T\to\infty]{\mathbb{P}} \text{med}(y) = \alpha$.

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, $\hat{\alpha}_{OLS} = \frac{\sum_t y_t}{T}$. 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.

4.4. Truncated ordinary least squares. As Example 2 suggests, the ordinary least squares (OLS) does not produce a consistent estimator. We show below the same inconsistency for generic values of parameters such that y_t is strongly mixing and converges to a stationary distribution, as in Theorems 2 and 3.

Lemma 12. *Suppose that $\mathbb{E}u_t = 0$. Then OLS is inconsistent.*

Proof. Define $\theta = (\alpha, \beta, \gamma)'$. If X is the matrix with rows $(1,\, y_{t-1},\, z_{t-1})$ and $Y = (y_1, \dots, y_T)'$, then the OLS estimator is

$$
\begin{aligned}
\hat{\theta}_{OLS} &= (X'X)^{-1}X'Y \\
&= \theta + (X'X)^{-1}X'U - (X'X)^{-1} \begin{pmatrix} \sum\limits_{t:\, y_t=0} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) \\ \sum\limits_{t:\, y_t=0} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) y_{t-1} \\ \sum\limits_{t:\, y_t=0} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) z_{t-1} \end{pmatrix}
\end{aligned}
\tag{8}
$$

The term $(X'X)^{-1}X'U$ converges to zero as T goes to infinity by the law of large numbers, because $(1,\, y_{t-1},\, z_{t-1})$ is independent of u_t. However, the last term does not converge to zero:

$$
\begin{aligned}
&\frac{1}{T} \begin{pmatrix} \sum\limits_{t:\, y_t=0} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) \\ \sum\limits_{t:\, y_t=0} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) y_{t-1} \\ \sum\limits_{t:\, y_t=0} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) z_{t-1} \end{pmatrix} \\
&\xrightarrow[T\to\infty]{\mathbb{P}} \begin{pmatrix} \mathbb{E}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) \mathbf{1}(u_t < -\alpha - \beta y_{t-1} - \gamma z_{t-1}) \\ \mathbb{E}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) y_{t-1} \mathbf{1}(u_t < -\alpha - \beta y_{t-1} - \gamma z_{t-1}) \\ \mathbb{E}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t) z_{t-1} \mathbf{1}(u_t < -\alpha - \beta y_{t-1} - \gamma z_{t-1}) \end{pmatrix},
\end{aligned}
$$

where the expectations do not equal to zero. Thus, OLS is inconsistent. Moreover, each term in the expectation is negative when the indicator equals to 1. So OLS overestimates the coefficients as T goes to infinity. □

The advantage of the OLS procedure is the closed form for the estimator. We also recall that in the linear models the OLS estimator is more efficient than the LAD. These two properties motivate us to attempt to adjust the OLS procedure to restore consistency.

The idea of the modified procedure is similar to identification: when y_{t-1} (z_{t-1}) is large, while z_{t-1} (y_{t-1}) is small, we can treat the constant and the second regressor as part of an error. Thus, we are left effectively with the classical autoregression model and can use standard theory. Mathematically, to estimate β, we need to condition on $T_{1M} = \{t|y_t > 0,\, y_{t-1} > M,\, z_{t-1} < M/h(M)\}$ for some number $M > 0$ and function $h(\cdot)$ such that $h(M) \xrightarrow[M\to\infty]{} \infty$. When M is large, $-\alpha - \beta y_{t-1} - \gamma z_{t-1}$ is very negative, so the indicator $\mathbf{1}(u_t < -\alpha - \beta y_{t-1} - \gamma z_{t-1})$ almost always equals zero, and the last term in Eq. (8) disappears as $T \to \infty$. Similarly, we can condition on $T_{2M} = \{t|y_t > 0,\, y_{t-1} < M/h(M),\, z_{t-1} > M\}$ to recover γ. The next theorem, which is proved in the Appendix, summarizes the above heuristics.

Theorem 13. *Separate OLS estimates of β and γ based on T_{1M} and T_{2M} are consistent, respectively, as $(M,T)_{seq} \to \infty$:*

$$\frac{\sum\limits_{T_{1M}} y_{t-1}y_t}{\sum\limits_{T_{1M}} y_{t-1}^2} \xrightarrow[(M,T)_{seq}\to\infty]{\mathbb{P}} \beta, \qquad \frac{\sum\limits_{T_{2M}} z_{t-1}y_t}{\sum\limits_{T_{2M}} z_{t-1}^2} \xrightarrow[(M,T)_{seq}\to\infty]{\mathbb{P}} \gamma.$$

After β and γ are estimated, one can estimate α using

$$\frac{1}{|T_{1M}| + |T_{2M}|} \sum_{T_{1M}\cup T_{2M}} (y_t - \beta y_{t-1} - \gamma z_{t-1}) \xrightarrow[(M,T)_{seq}\to\infty]{\mathbb{P}} \alpha. \tag{9}$$

In practice, to estimate α we need to use a different, smaller threshold. That is, we first estimate β and γ based on some M_1 and then we plug the estimates into Eq. (9), evaluated at $M_2 < M_1$, to estimate α.

Let us note, that in practice we can use a simpler procedure. We denote it as $\mathrm{OLS_M}$. One can condition on $T_M := \{t|y_t > 0,\, y_{t-1} > M\}$ for some $M > 0$ and run OLS with three regressors. The problem here is that the limit behavior of the inverse of conditional matrix $\begin{pmatrix} 1 & \mathbb{E}(y|T_M) & \mathbb{E}(z|T_M) \\ \mathbb{E}(y|T_M) & \mathbb{E}(y^2|T_M) & \mathbb{E}(yz|T_M) \\ \mathbb{E}(z|T_M) & \mathbb{E}(yz|T_M) & \mathbb{E}(z^2|T_M) \end{pmatrix}^{-1}$ is unclear. It may crucially depend on the properties of the error distribution. As long as post-multiplication by the vector of cross product covariances $(0, Cov(y_{t-1}, u_t|T_M), Cov(z_{t-1}, u_t|T_M))'$ results in the zero vector in the limit, the sequential limit of the corresponding OLS estimate equals (α, β, γ). That is, the inverse matrix must not explode faster than the conditional covariance vector goes to zero. This is summarized in the next theorem.

Theorem 14. *The sequential limit* $(M,T)_{seq} \to \infty$ *of the OLS estimator based on* $t \in T_M$ *equals the true value* (α, β, γ) *when the product*

$$\begin{pmatrix} 1 & \mathbb{E}(y_{t-1}|T_M) & \mathbb{E}(z_{t-1}|T_M) \\ \mathbb{E}(y_{t-1}|T_M) & \mathbb{E}(y_{t-1}^2|T_M) & \mathbb{E}(y_{t-1}z_{t-1}|T_M) \\ \mathbb{E}(z_{t-1}|T_M) & \mathbb{E}(y_{t-1}z_{t-1}|T_M) & \mathbb{E}(z_{t-1}^2|T_M) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Cov(y_{t-1}, u_t|T_M) \\ Cov(z_{t-1}, u_t|T_M) \end{pmatrix}$$

converges to zero as $M \to \infty$.

In simulations, the product of the inverse conditional matrix of second moments and the conditional covariance vector goes to zero. Thus, in the empirical example studied below we use the above procedure to calculate adjusted OLS estimates. Moreover, as the next theorem suggests, when there are no peer effects ($\gamma \equiv 0$) and both u_t and y_t have exponential tails, the product goes to zero. When there is no γ, the product of the inverse conditional matrix of second moments and the conditional covariance vector reduces to

$$\begin{pmatrix} 1 & \mathbb{E}(y_{t-1}|T_M) \\ \mathbb{E}(y_{t-1}|T_M) & \mathbb{E}(y_{t-1}^2|T_M) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Cov(y_{t-1}, u_t|T_M) \end{pmatrix}$$
$$= \frac{1}{\mathbb{V}(y_{t-1}|T_M)} \begin{pmatrix} -\mathbb{E}(y_{t-1}|T_M) Cov(y_{t-1}, u_t|T_M) \\ Cov(y_{t-1}, u_t|T_M) \end{pmatrix}$$

Theorem 15. *Assume that the stationary distribution of* y_t *has density* $f_y(x)$ *for large positive* x *and that the noise* u_t *has density* $f_u(x)$ *for large negative* x. *Further, assume that there exist six positive constants* $c_1, c_2, c_3, d_1, d_2, d_3 > 0$, *such that for all large enough positive* x:

$$f_y(x) = \exp(-g_y(x)), \ \textit{where } c_1 x^{d_1} \le g_y'(x) \le c_2 x^{d_2} \tag{10}$$

and for all large enough negative x:

$$f_u(x) = \exp(-g_u(x)), \ \textit{where } g_u(x) \ge c_3 |x|^{d_3}. \tag{11}$$

Then the vector $\frac{1}{\mathbb{V}(y_{t-1}|T_M)} \begin{pmatrix} -\mathbb{E}(y_{t-1}|T_M) Cov(y_{t-1}, u_t|T_M) \\ Cov(y_{t-1}, u_t|T_M) \end{pmatrix}$ *goes to* 0 *as* $M \to \infty$, *i.e. the* OLS_M *estimate is consistent as* $M \to \infty$.

Remark 9. The conditions (10), (11) mean that both the noise and the stationary distribution have light tails. The condition (10) additionally requires that the tail probability of the stationary distribution of y_t does not decay too fast. These conditions are not intended to be optimal and can likely be considerably weakened. Instead, they are intended to illustrate the type of conditions where Theorem 14 holds.

The disadvantage of the adjusted OLS procedures is that we have to discard a lot of observations. Moreover, it is unclear how to choose M and $h(M)$. The tradeoff is that the larger is M, the more observations we have to discard, yet the closer to the consistent limit we are. Thus, we see that as we restore the consistency by increasing M, we lose the efficiency of the estimator.

4.5. Maximum likelihood estimator. Suppose that we know the density, f_u, of the error u_t. Then we can calculate the likelihood. It will consist of two types of terms. The first type corresponds to the cases when y_t is non-zero, the positive part is non-binding, so we can write $y_t = \alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t$ or $u_t = y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1}$. The second type corresponds to time periods with $y_t = 0$. If y_t is zero, then it is equivalent to $\alpha + \beta y_{t-1} + \gamma z_{t-1} + u_t$ being non-positive. That is, $y_t = 0$ is equivalent to $u_t \leq -\alpha - \beta y_{t-1} - \gamma z_{t-1}$. Thus, the likelihood and its logarithm are

$$
\begin{aligned}
\mathcal{L} &= \prod_{t:\, y_t>0} f_u(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1}) \times \prod_{t:\, y_t=0} F_u(-\alpha - \beta y_{t-1} - \gamma z_{t-1}), \\
\log \mathcal{L} &= \sum_{t:\, y_t>0} \log f_u(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1}) + \sum_{t:\, y_t=0} \log F_u(-\alpha - \beta y_{t-1} - \gamma z_{t-1}).
\end{aligned} \tag{12}
$$

Following common practice, we assume a normal distribution for u_t. It turns our, as the Theorem 16 shows, that when the true distribution is normal, the MLE produces consistent estimators. However, as the simulations suggest, and in agreement with the well-known results in the i.i.d. censored regression model, when the true distribution is far from normal, the estimates are poor. Moreover, numerical optimization is very sensitive to the choice of the initial point and the calculations for the MLE sometimes explode.

Theorem 16. *If $u_t \sim i.i.d. \mathcal{N}(0, \sigma^2)$, then MLE is consistent.*

The proof of Theorem 16, which is shown in Appendix, uses extremum estimation techniques. In a similar spirit it is possible to show $\sqrt{T}$- asymptotic normality of the MLE estimator under Gaussian errors.

5. Prediction

In this section we again treat equations for each edge separately. 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.

To measure predictive power, we can use a rolling window approach. That is, we choose some number $T' < T$, and estimate the model based on observations $t, \ldots, t + T' - 1$. For each $t = 1, \ldots, T - T'$ we can calculate the difference between the predicted value $\hat{y}_{t+T'}$ and the actual value $y_{t+T'}$. Thus, we get a measure of how well we can predict the data:

$$R_{abs} = \frac{1}{T - T'} \sum_{t=1}^{T-T'} |y_{t+T'} - \hat{y}_{t+T'}|. \tag{13}$$

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 optimal prediction in L_2 norm is $\int\limits_{-\alpha-\beta y_{t-1}-\gamma z_{t-1}}^{\infty} (\alpha + \beta y_{t-1} + \gamma z_{t-1} + u) f_u(u) du$. Thus, it crucially depends on the distribution of the error term, f_u. However, optimal prediction in L_1 is $[\alpha + \beta y_{t-1} + \gamma z_{t-1}]_+$, as shown below. That is, it does not depend on the distribution of the error, and is more convenient to work with.

Remark 10. To see that optimal 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

$$\begin{aligned} \int |y_{t+1} - \hat{y}_{t+1}| f(u) du &= \int |[\alpha + \beta y_t + \gamma z_t + u]_+ - \hat{y}_{t+1}| f(u) du \\ &= \int |\max(u, -\alpha - \beta y_t - \gamma z_t) - \Delta \hat{y}_{t+1}| f(u) du. \end{aligned} \tag{14}$$

Because $\arg\min_C \mathbb{E}|v - C| = \text{med}(v)$, and

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

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

minimizing Eq. (14), 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 $\hat{y}_{t+1} = [\alpha + \beta y_t + \gamma z_t]_+$.

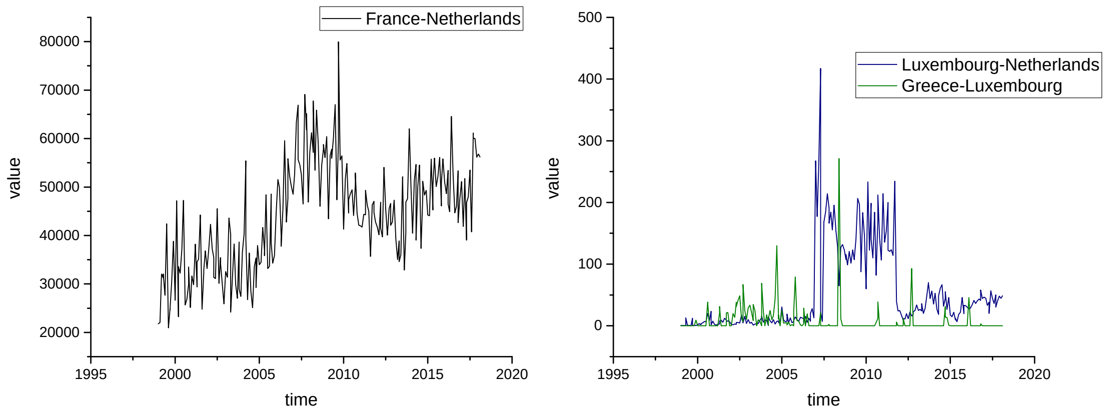

FIGURE 3. Exports of pharmaceutical products between EU countries. Value is in 1000 Euro.

6. Empirical Application

We apply our model to monthly exports of pharmaceutical products (hs2 code 30). The time period of observation is from January of 1999 until February of 2018, so that $T = 230$. We choose 12 European Union countries. Those are the countries which joined the EU first. Figure 3 shows exports 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.

To measure predictive power we use a rolling window of length 200. This gives 30 rolling windows and we report the average predictive absolute error over them. For testing in the next subsection we need more windows and with that aim in mind we also report results for the rolling window of size 115, which gives 115 windows. 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.

The benchmark prediction is "today equals tomorrow", i.e., $\hat{y}_{t+1} = y_t$. Another option is to ignore positivity and treat the model as linear, i.e., $y_t = \alpha + \beta y_{t-1} + u_t$. We compare these two approaches with the LAD estimator without peer effects. Results are shown in Table 1. We can see that for the larger rolling window model-based predictors outperform both alternatives. Moreover, the LAD estimate outperforms the MLE and adjusted OLS (with $M = 0$) predictors. However, for the smaller window size 115 all estimates lead to poor results. This suggests that the available T is at the border of applicability of our methods. Yet, in the next paragraph we will see that the addition of peer effects improves the results for small T as well. The fact that LAD performs significantly better than

	R_{abs}	
	window size = 200	window size = 115
LAD	$1.9865e+06$	$1.8977e+06$
MLE	$2.0036e+06$	$1.8847e+06$
OLS	$2.0037e+06$	$1.8848e+06$
OLS_0	$2.0038e+06$	$1.8848e+06$
"today"	$2.0316e+06$	$1.9096e+06$

Table 1. Prediction errors under different estimation techniques.

	R_{abs}	
	window size = 200	window size = 115
LAD	$1.8934e+06$	$1.8315e+06$
MLE	$1.9215e+06$	$1.8274e+06$
LAD w/o p.e.	$1.9865e+06$	$1.8977e+06$
MLE w/o p.e.	$2.0036e+06$	$1.8847e+06$
"today"	$2.0316e+06$	$1.9096e+06$

Table 2. Prediction errors with and without peer effects.

the MLE for the larger window size suggests that the error distribution may be far from normal.

We use the triangular peer effect function: $z_{ijt-1} = \sum\limits_k \frac{\sqrt{y_{ikt-1} y_{kjt-1}}}{n-2}$. Table 2 shows that adding peer effects reduces the prediction error both under LAD and MLE estimation approaches. This suggests the presence of peer effects in the data. The scatter plot of the LAD estimates $(\beta_{ij}, \gamma_{ij})$ 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, can expect that the use of additional relevant regressors may improve the prediction results reported in Table 2. It turns out, as Table 3 shows, that using four lags and one peer effect regressor or using five lags gives the best results ($z_{ijt-1} = \sum\limits_k \frac{\sqrt{y_{ikt-4} y_{kjt-4}}}{n-2}$). Also using peer effects evaluated at $t-4$ leads to better performance than using peer effects 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.

	R_{abs}	
	window size = 200	window size = 115
5 lags	$1.6731e+06$	$1.6650e+06$
4 lags+p.e. at $t-4$	$1.6733e+06$	$1.6437e+06$
4 lags+p.e. at $t-1$	$1.6797e+06$	$1.6511e+06$
1 lag+p.e. at $t-4$	$1.8569e+06$	$1.8172e+06$
1 lag+p.e. at $t-1$	$1.8934e+06$	$1.8315e+06$
1 lag	$1.9865e+06$	$1.8977e+06$
"today"	$2.0316e+06$	$1.9096e+06$

TABLE 3. Prediction errors with different lags and peer effects.

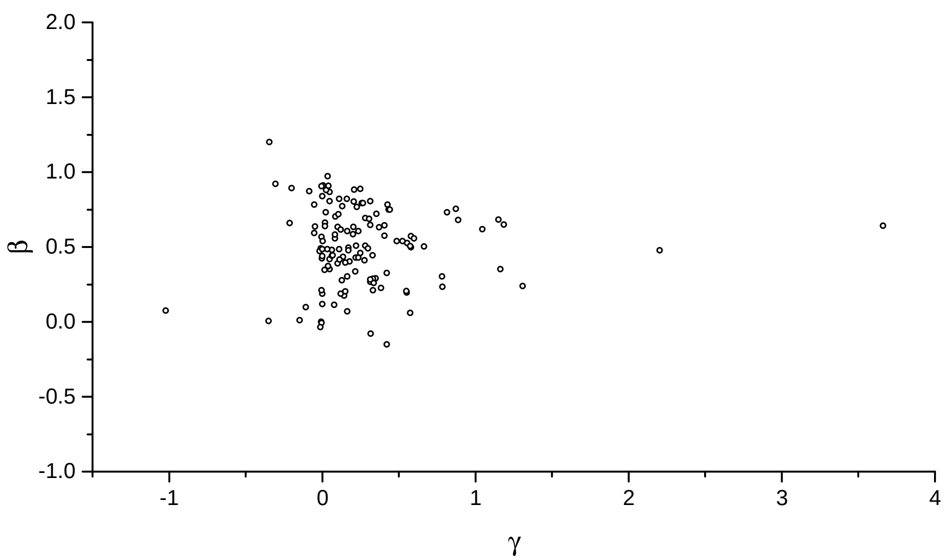

FIGURE 4. LAD estimates for 132 country pairs for the model

$$y_t = [\alpha + \beta y_{t-1} + u\gamma z_{t-1} + u_t]_+.$$

6.1. Tests. 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. We compare the model without peer effects against the benchmark prediction "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

	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 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. 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.

7. Conclusion

This paper presents a novel approach to modeling and estimating networks. Estimation does not require knowledge of the error distribution, thereby making the whole process more attractive to use. Instead of using the variation across individuals, one can use variation across time to identify parameters of the network. In this approach, we treat networks as multivariate time series. 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 period, 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.

It would be interesting in the future 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. This is currently under investigation by the author.

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.

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

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

Diebold, F.X. and R.S. Mariano, "Comparing predictive accuracy," *Journal of Business & economic statistics*, 1995, *13* (3), 253–63.

Feller, W., *An introduction to probability theory and its applications*, Vol. 2, John Wiley & Sons, 2008.

Gourieroux, C. and A. Monfort, "Sufficient linear structures: Econometric applications," *Econometrica*, 1980, *48* (5), 1083–1097.

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

Hayashi, F., *Econometrics*, Princeton University Press, 2000.

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

Jackson, M.O., *Social and economic networks*, Princeton University Press, 2010.

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.

Karatzas, I. and S. Shreve, *Brownian motion and stochastic calculus*, Springer Science & Business Media, 2012.

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

Lévy, P., *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars,Paris, 1948.

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.

Pin, P. and B.W. Rogers, "Cooperation, punishment and immigration," *Journal of Economic Theory*, 2015, *160*, 72–101.

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.

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.

Appendix A. Stationarity/Explosiveness.

This subsection presents proofs on stationary/explosive behavior of the process y_{ijt} as t goes to infinity.

Proof of Theorem 1.

Proof. To simplify notation, let us denote

$$z_{ijt-1} = p\left(\left\{y_{kls}\right\}_{\substack{(k,l)\neq(i,j)\\ s=t-H,\ldots,t-1}}\right).$$

Then

$$\begin{aligned} y_{ijt} &= [\alpha_{ij} + \beta_{ij}y_{ijt-1} + \gamma_{ij}z_{ijt-1} + u_{ijt}]_+ \\ &\leq [\max(0,\beta_{ij})y_{ijt-1} + |\gamma_{ij}z_{ijt-1}| + \max(0,\alpha_{ij}+u_{ijt})]_+ \\ &= \max(0,\beta_{ij})y_{ijt-1} + |\gamma_{ij}z_{ijt-1}| + \max(0,\alpha_{ij}+u_{ijt}) \\ &\leq (\max(0,\beta_{ij}) + |\gamma_{ij}|)\max_{k,l,s=t-H,\ldots,t-1} y_{kls} + |\mathcal{A}| + \max(0,\alpha_{ij}+u_{ijt}) \\ &\leq C \max_{k,l,s=t-H,\ldots,t-1} y_{kls} + |\mathcal{A}| + \max(0,\alpha_{ij}+u_{ijt}). \end{aligned} \tag{15}$$

Denote $v_t = |\mathcal{A}| + \max_{k,l,s=t-H,\ldots,t-1} \max(0,\alpha_{kl}+u_{kls})$. Then from Eq. (15) we get

$$\max_{i,j} y_{ijt} \leq C \max_{k,l,s=t-H,\ldots,t-1} y_{kls} + v_t. \tag{16}$$

We are going to show that $\max_{\substack{i,j\\ s=mH+1,\ldots,mH+H}} y_{ijs} \leq C \max_{\substack{i,j\\ s=(m-1)H+1,\ldots,mH}} y_{ijs} + w_m$, where $w_m = v_{mH+1} + \ldots + v_{mH+H}$.

By Eq. (16),

$$\max_{i,j} y_{i,j,mH+1} \leq C \max_{i,j;s=(m-1)H+1,\ldots,mH} y_{ijs} + v_{mH+1}.$$

Applying Eq. (16) twice (for $t = mH + 2$ and $t = mH + 1$) we get

$$\begin{aligned}\max_{i,j} y_{i,j,mH+2} &\leq C \max_{i,j;s=(m-1)H+2,\dots,mH+1} y_{ijs} + v_{mH+2} \\ &\leq C \max(C \max_{i,j;s=(m-1)H+1,\dots,mH} y_{ijs} + v_{mH+1}, \max_{i,j;s=(m-1)H+2,\dots,mH} y_{ijs}) + v_{mH+2} \\ &\leq C \max_{i,j;s=(m-1)H+1,\dots,mH} y_{ijs} + v_{mH+1} + v_{mH+2},\end{aligned}$$

because $C < 1$ and $v_t \geq 0$.

Similarly, applying Eq. (16) for $t = mH + 3$ and previous bounds on $\max_{i,j} y_{i,j,mH+1}$ and $\max_{i,j} y_{i,j,mH+2}$, we get

$$\max_{i,j} y_{i,j,mH+3} \leq C \max_{i,j;s=(m-1)H+1,\dots,mH} y_{ijs} + v_{mH+1} + v_{mH+2} + v_{mH+3}.$$

We can redo the same trick for $t = mH + 4, \dots, t = mH + H$, so that

$$\max_{i,j} y_{i,j,mH+r} \leq C \max_{i,j;s=(m-1)H+1,\dots,mH} y_{ijs} + v_{mH+1} + \dots + v_{mH+r}.$$

Thus,

$$\max_{\substack{i,j \\ s=mH+1,\dots,mH+H}} y_{ijs} \leq C \max_{\substack{i,j \\ s=(m-1)H+1,\dots,mH}} y_{ijs} + w_m,$$

where $w_m = v_{mH+1} + \dots + v_{mH+H} \geq 0$.

Iterative back-substitution leads to

$$\begin{aligned}\max_{\substack{i,j \\ s=mH+1,\dots,mH+H}} y_{ijs} &\leq C \max_{\substack{i,j \\ s=(m-1)H+1,\dots,mH}} y_{ijs} + w_m \\ &\leq w_m + Cw_{m-1} + C^2 \max_{\substack{i,j \\ s=(m-2)H+1,\dots,(m-1)H}} y_{ijs} \\ &\leq \sum_{s=0}^{m-1} C^s w_{m-s} + C^m \max_{\substack{i,j \\ s=1,\dots,H}} y_{ijs}.\end{aligned} \tag{17}$$

Note that $\mathbb{E}w = H\mathbb{E}v < C_1 < \infty$, as H and n, number of agents, are finite. Then from Eq. (17)

$$\mathbb{E}y_{ijt} \leq \mathbb{E} \max_{\substack{k,l \\ s=H\lfloor \frac{t-1}{H} \rfloor+1,\dots,H\lfloor \frac{t-1}{H} \rfloor+H}} y_{kls} \leq \frac{\mathbb{E}w}{1-C} + const < C_2 < \infty.$$

□

In the following set of lemmas and theorems we show sufficient condition for stationarity for the general model (Theorem 2) and the full characterization of stationary/explosive limiting behavior of y_t depending on α, β (Theorem 3) for the model without peer effects.

Lemma A.1. *If Assumptions 1, 2, and 6 are satisfied, $\mathbb{E}u_{ijt} < \infty$ for all i, j, t, and for all i, j $\max(0, \beta_{ij}) + |\gamma_{ij}| < C < 1$, then $\mathbb{E}$(time until graph is empty for H periods) is finite. That is, the expected time until $y_{ijt} = \dots = y_{ij,t+H-1} = 0$ for all i, j is finite.*

Proof. Denote by $\bar{u}_t = \{u_{ijt}\}_{i,j}$ the vector of all errors at time t. Fix some number $M > 0$ (it will be specified later) and define three independent random variables

$$\bar{u}_t^- = \{\{u_{ijt}\}_{i,j} | u_{ijt} < -M \; \forall i, j\},$$

$$\bar{u}_t^+ = \{\{u_{ijt}\}_{i,j} | u_{i'j't} \geq -M \text{ for some } i', j'\},$$

$$\xi_t = \begin{cases} 0, \text{with probability } \mathbb{P}(\forall i, j \;\; u_{ijt} < -M), \\ 1, \text{with probability } \mathbb{P}(\exists i, j \text{ s.t. } \; u_{ijt} \geq -M). \end{cases}$$

Then

$$\bar{u}_t \stackrel{d}{=} \xi_t \bar{u}_t^+ + (1 - \xi_t)\bar{u}_t^-. \tag{18}$$

Fix realizations of $(\bar{u}_t^+, \bar{u}_t^-, \xi_t)$ for $t = 1, \ldots, T$ and calculate the corresponding $\bar{u}_t$ from Eq. (18). Define

$$v_t = |\mathcal{A}| + \max_{\substack{i,j \\ s=t,\ldots,t-H+1}} \left[u_{ijs}^+ + \alpha_{ij}\right]_+ \geq 0.$$

Now construct a new time series

$$y'_t = C \max_{s=t-1,\ldots,t-H} y'_s + v_t,$$

$$y'_p = \max_{i,j} y_{ijp} \text{ for } p = 0, \ldots, H-1.$$

One can easily show by induction that $y'_t \geq y_{ijt}$ for all i, j, t.

By the same argument as in proof of Theorem 1, we can divide time periods into blocks of length H and get a bound

$$y'_{t+p} \leq C \max_{s=t-1,\ldots,t-H} y'_s + \sum_{s=0}^{H-1} v_{t+s} \;\text{ for all } p = 0, \ldots, H-1. \tag{19}$$

Now define another random process and error, $x_\tau = \max_{s=(\tau-1)H,\ldots,\tau H-1} y'_s$, $w_\tau = \sum_{s=0}^{H-1} v_{(\tau-1)H+s}$. Then by Eq. (19),

$$x_{\tau+1} \leq C x_\tau + w_{\tau+1}.$$

We need to find M such that for some $\varepsilon > 0$, $\mathbb{P}(\sharp\{\tau \in [1, \ldots, \lfloor T/H \rfloor] | x_\tau < M\} \geq \varepsilon T) \geq \frac{1}{T^2}$ and $\mathbb{P}(\forall i, j \;\; u_{ijt} < -M) > 0$. This is a condition on u_{ijt} which generally may fail to be true. For example, if u_{ijt} are almost surely larger than some positive constant. Let us show that such M exists under assumptions 2 and $\mathbb{E}u_{ijt} < C_4 < \infty$. Assumption 2 implies that for all M large enough $\mathbb{P}(\forall i, j \;\; u_{ijt} < -M) > 0$.

Let us show that if $\mathbb{E}u_{ijt} < C_4 < \infty$, then $\mathbb{E}w_\tau < \tilde{C}_4 < \infty$, and constant $\tilde{C}_4$ does not depend on M. Note that the fourth moment of u^+_{ijt} is bounded as

$$\mathbb{E}|u^+_{ijt}|^4 \le \mathbb{E}|u^4_{ijt}| \frac{1}{\mathbb{P}(\exists i, j \text{ s.t. } u_{ijt} \ge -M)}.$$

Further, as $|w_\tau| = \sum_{s=0}^{H-1} v_{(\tau-1)H+s}$, it is less than the sum of absolute values of several instances of u^+_{ijt} and constants. Thus, the fourth moment of the sum can be bounded by a combination of the individual fourth moments, which are bounded. As $M \to \infty$, $\mathbb{P}(\exists i, j \text{ s.t. } u_{ijt} \ge -M) \to 1$, so that setting a lower bound for M to be such that $\mathbb{P}(\exists i, j \text{ s.t. } u_{ijt} \ge -M') = 0.5$, we get a bound which does not depend on $M > M'$ and then $\mathbb{E}|u^+_{ijt}|^4 \le 2\mathbb{E}|u^4_{ijt}|$.

Define one more process, $\tilde{x}_\tau$, by

$$\tilde{x}_{\tau+1} = C\tilde{x}_\tau + w_{\tau+1}, \ \tilde{x}_1 = x_1.$$

It can be shown by induction, that for all τ, $\tilde{x}_\tau \ge x_\tau$. Thus, it is enough to show that $\mathbb{P}(\sharp\{\tau \in [1, \dots, T/H] | \tilde{x}_\tau < M\} \ge \varepsilon T) \ge \frac{1}{T^2}$. Let us show that $\exists Q$ such that $\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_{\lfloor T/H \rfloor} < Q\lfloor T/H \rfloor$ with probability greater that $1 - \frac{const}{T^2}$.

$$\begin{aligned}\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_{\lfloor T/H \rfloor} &= \tilde{x}_1 + \sum_{\tau=2}^{\lfloor T/H \rfloor} (w_\tau + Cw_{\tau-1} + \ldots + C^{\tau-2}w_2 + C^{\tau-1}\tilde{x}_1) \\ &\le \frac{1}{1-C} \sum_{\tau=1}^{\lfloor T/H \rfloor} w_\tau + \frac{\tilde{x}_1}{1-C}.\end{aligned}$$

The expectation of the right hand side of the last expression is $\frac{1}{1-C}\left(\lfloor T/H \rfloor \mathbb{E}w_\tau + \mathbb{E}x_1\right)$

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{1-C}\sum_{\tau=1}^{\lfloor T/H \rfloor}(w_\tau - \mathbb{E}w_\tau)\right| > \lfloor T/H \rfloor\right) &\le \frac{\mathbb{E}\left|\sum_{\tau=1}^{\lfloor T/H \rfloor}(w_\tau - \mathbb{E}w_\tau)\right|^4}{(1-C)^4\lfloor T/H \rfloor^4} \\ &\le \frac{const \cdot T^2}{T^4} \le \frac{const}{T^2},\end{aligned}$$

where we used the fact that $w_\tau - \mathbb{E}w_\tau$ are i.i.d. with zero mean and with bounded fourth and second moments. Thus,

$$\mathbb{P}\left(\frac{1}{1-C}\sum_{\tau=1}^{\lfloor T/H \rfloor} w_\tau > Q\lfloor T/H \rfloor\right) \le \mathbb{P}\left(\left|\frac{1}{1-C}\sum_{\tau=1}^{\lfloor T/H \rfloor}(w_\tau - \mathbb{E}w_\tau)\right| > \lfloor T/H \rfloor\right) \le \frac{const}{T^2},$$

where $Q = 1 + \frac{\mathbb{E}w_\tau}{1-C}$. Thus,

$$\mathbb{P}\left(\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_{\lfloor T/H \rfloor} < Q\lfloor T/H \rfloor\right) \geq \mathbb{P}\left(\frac{1}{1-C}\sum_{\tau=1}^{\lfloor T/H \rfloor} w_\tau + \frac{\tilde{x}_1}{1-C} < Q\lfloor T/H \rfloor\right) \tag{20}$$
$$\geq 1 - \frac{const}{T^2}.$$

Finally, if $\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_{\lfloor T/H \rfloor} < Q\lfloor T/H \rfloor$, then $x_1 + \ldots + x_{\lfloor T/H \rfloor} < Q\lfloor T/H \rfloor$. The latter implies that $\sharp\{\tau | x_\tau > 2Q\} < 0.5\lfloor T/H \rfloor$ and $\sharp\{\tau | x_\tau \leq 2Q\} > 0.5\lfloor T/H \rfloor$. That is, we have shown that $\exists M$ (any number larger than $2Q$ and M') such that $\sharp\{\tau | x_\tau \leq M\} > \varepsilon T$ has probability greater than $1 - \frac{const}{T^2}$. For each such τ we flip a coin to determine ξ_t. If it zero, then the whole process y_{ijt} jumps to zero. Thus, with probability of at most $(\mathbb{P}(\exists i, j \text{ s.t. } u_{ijt} \geq -M))^{\varepsilon T}$ the process does not jump to zero. Thus,

$$\mathbb{E}(\text{length until } H \text{ zero periods}) = \sum_{T=1}^{\infty} \mathbb{P}(length \geq T)$$
$$\leq \sum_{T=1}^{\infty}\left(\frac{const}{T^2} + (\mathbb{P}(\exists i, j \text{ s.t. } u_{ijt} \geq -M))^{\varepsilon T}\right) < \infty.$$

□

Corollary A.2. *If Assumptions 1 and 2 are satisfied for the model without γ and if $\beta < 1$, then $\mathbb{E}$(length until zero) is finite.*

Proof. If $\beta < 1$ and there is no γ, $\max(0, \beta) + |\gamma| = \max(0, \beta) < 1$. Thus, Theorem A.1 applies. □

Lemma A.3. *If Assumptions 1 and 2 are satisfied $\alpha < 0$, $\beta = 1$ and if $\mathbb{E}u_t^4 < \infty$, then $\mathbb{E}$(length until zero) is finite.*

Proof. We can write the expected length until $y_t = 0$ as

$$\mathbb{E}(\text{length until zero}) = \sum_{T=1}^{\infty} \mathbb{P}(length \geq T). \tag{21}$$

Define $S_t = y_0 + t\alpha + u_1 + \ldots + u_t$ for all $t \in \mathbb{N}$. If length until zero is greater than T, then $S_1 > 0, \ldots, S_{T-1} > 0$. (Otherwise the process S_t becomes negative, so that non-negative process y_t becomes zero before T). Thus, $\mathbb{P}(length \geq T) \leq \mathbb{P}(S_1 > 0, \ldots, S_{T-1} > 0)$.

Note that

$$\mathbb{P}(S_1 > 0, \dots, S_{T-1} > 0) = \mathbb{P}\left(y_0 + \alpha + u_1 > 0, \dots, y_0 + (T-1)\alpha + \sum_{t=1}^{T-1} u_t > 0\right)$$
$$\leq \mathbb{P}\left(y_0 + (T-1)\alpha + \sum_{t=1}^{T-1} u_t > 0\right) = \mathbb{P}\left(\sum_{t=1}^{T-1} u_t > -y_0 - (T-1)\alpha\right).$$

Because $\alpha < 0$, there exists T' such that $\forall T > T'$ $-y_0 - (T-1)\alpha > 0$. Let us look at any $T > T'$.

(22)

$$\mathbb{P}\left(\sum_{t=1}^{T-1} u_t > -y_0 - (T-1)\alpha\right) \leq \mathbb{P}\left(\left|\sum_{t=1}^{T-1} u_t\right| > -y_0 - (T-1)\alpha\right) \leq \frac{\mathbb{E}\left|\sum_{t=1}^{T-1} u_t\right|^4}{(y_0 + (T-1)\alpha)^4}$$
$$= \frac{(T-1)\mathbb{E}u^4 + 3(T-1)(T-2)\mathbb{E}u^2}{(y_0 + (T-1)\alpha)^4} \leq \frac{const}{T^2},$$

where we used Markov inequality to bound probability by expectation.

Plugging Eq. (22) into Eq. (21), we get

$$\mathbb{E}(\text{length until zero}) \leq \sum_{T=1}^{\infty} \frac{const}{T^2} < const_1 < \infty. \qquad \square$$

The next theorem proves Theorem 2 and the stationary cases of Theorem 3.

Theorem A.4. *Suppose that Assumptions 1, 2, and 6 are satisfied and* $\mathbb{E}u_{ijt}^4 < \infty$ *for all* i, j, t. *For the model without* γ, *if* $\beta < 1$ *or* $\beta = 1, \alpha < 0$, *then* y_t *is strongly mixing and converges to a stationary distribution. For the model with* γ, *if for some* C $\max(0, \beta_{ij}) + |\gamma_{ij}| < C < 1$ *for all* i, j, *then* y_{ijt} *is strongly mixing and converges to a stationary distribution*[2].

Proof. Let us first show convergence to a stationary distribution. The proof follows the lines of section $XI.8$ in Feller (2008). Let us show the proof for the model without peer effects.

From Lemma A.3 and Corollary A.2, we know that $\mathbb{E}$(length until zero) is finite. Let us denote the expected time until zero by μ. Thus, with probability one the process reaches zero. The continuation of a process after it reaches zero is a probabilistic replica of the whole process started from the previous zero.

[2]Formally this means that the finite-dimensional distributions of the process $\{y_{t+\tau}\}_{\tau\in\mathbb{Z}}$, converge to those of a stationary in τ process as $t \to \infty$.

For any Borel set Δ denote by $P_\Delta(t)$ the probability that $y_{t+s} \in \Delta$ given that s is the (finite) time before the process first hits zero. The process y_t is by definition strongly Markov, so that such probability does not depend on s.

We are going to show that for any Borel set Δ there exists $\lim\limits_{t\to\infty} P_\Delta(t) = P_\Delta$ such that $P_\Delta \geq 0$, $P_{\mathbb{R}_+} = 1$, and P_Δ is countably additive. This would imply that the one-point distribution of y_t converges to a limit at $t \to \infty$; by the Markov property the latter further implies the desired convergence of all finite-dimensional distributions of $\{y_{t+\tau}\}_{\tau\in\mathbb{Z}}$.

Define by S_1 the first time of hitting zero, by S_2 the second time of hitting zero, etc. Also define $q_\Delta(t) = \mathbb{P}(S_1 > t, y_t \in \Delta)$. Then

$$q_\Delta(t) + q_{\mathbb{R}_+\setminus\Delta}(t) = 1 - F(t),$$

where F is a distribution of time between two consequent moments of hitting zero ($S_{n+1} - S_n$).

Because the probability that $y_{t+s} \in \Delta$ given $S_1 = s$ does not depend on s, we can write

$$P_\Delta(t) = \mathbb{P}(S_1 > t,\, y_t \in \Delta) + \mathbb{P}(S_1 \leq t,\, y_t \in \Delta) = q_\Delta(t) + \int_0^t P_\Delta(t-y)F(dy).$$

The function $q_\Delta(t)$ is directly integrable since it is dominated by the monotone integrable function $1 - F$. Therefore by the renewal theorem

$$\lim_{t\to\infty} P_\Delta(t) = \frac{1}{\mu}\sum_{t=0}^{\infty} q_\Delta(t) \geq 0.$$

By definition, $q_{\mathbb{R}_+}(t) = 1 - F(t)$ so that $\lim\limits_{t\to\infty} P_{\mathbb{R}_+}(t) = \frac{1}{\mu}\sum\limits_{t=0}^{\infty}(1-F(t)) = \frac{\mu}{\mu} = 1$. Similarly, for any Δ, $q_\Delta(t) \leq 1 - F(t)$, and, thus, $\lim\limits_{t\to\infty} P_\Delta(t) \leq 1$. Finally, we need to check that P_Δ is countably additive. This follows from the fact that for a countable number of pairwise disjoint sets Δ_i,

$$q_{\cup_i\Delta_i}(t) = \mathbb{P}(S_1 > t, y_t \in \cup_i\Delta_i) = \sum_i q_{\Delta_i}(t).$$

Thus, $\lim\limits_{t\to\infty} P_{\cup\Delta_i}(t) = \frac{1}{\mu}\sum\limits_{t=0}^{\infty} q_{\cup\Delta_i}(t) = \frac{1}{\mu}\sum\limits_{t=0}^{\infty}\sum\limits_i q_{\Delta_i}(t) \geq 0 = \sum\limits_i \lim\limits_{t\to\infty} P_{\Delta_i}(t)$.

The proof for the model with peer effects is the same, with the only difference that now y_t is a vector $\{y_{ijt}\}_{i,j}$. Such process is strongly Markov in extended space, where the element is a vector $\{y_{ij,t-1}, \ldots, y_{ij,t-H}\}_{i,j}$. (That is, we divide the time scale into blocks of length H.) Moreover, by Lemma A.1 $\mathbb{E}$(length until zero) is finite.

Let us now show strong mixing. All bounds on $\mathbb{E}$(length until zero) were uniform with respect to the choice of starting point y_0 (i.e. if $y_0 \in [0, M]$, then there exists a constant such that $\mathbb{E}$(length until zero) $< const$). (See Eq. (20) and Eq. (22).) Thus, because in the proof above $\lim\limits_{t\to\infty} P_{\cup\Delta_i}(t)$ is uniform over the choice of y_0 in a bounded set

$$\lim_{t\to\infty} |\mathbb{P}(y_s \in \Delta_1, y_{t+s} \in \Delta_2) - \mathbb{P}(y_s \in \Delta_1)\mathbb{P}(y_s \in \Delta_1, y_{t+s} \in \Delta_2)| = 0. \quad \square$$

Theorem A.5. *If $\beta = 1$, $\alpha > 0$ or $\beta > 1$, then y_t is divergent ($y_t \xrightarrow[t\to\infty]{a.s.} \infty$).*

Proof. We need to show that $\forall M \geq 0$ $\mathbb{P}(\lim\limits_{t\to\infty} y_t < M) = 0$.

Assume first that $\alpha > 0$ and note, that because taking positive part of a random variable can only increase it, we get

$$y_t = [\alpha + \beta y_{t-1} + u_t]_+ \geq \alpha + \beta y_{t-1} + u_t \geq \alpha + y_{t-1} + u_t.$$

Thus, if $\alpha > 0$, then $y_t \geq \alpha t + y_0 + \sum\limits_{s=1}^{t} u_s$, and

$$\text{If } y_t < M \text{ then } \alpha t + y_0 + \sum_{s=1}^{t} u_s < M \text{ and } \frac{1}{t}\sum_{s=1}^{t} u_s < \frac{M - y_0}{t} - \alpha.$$

By the strong law of large numbers, $\frac{1}{t}\sum\limits_{s=1}^{t} u_s \xrightarrow[t\to\infty]{a.s.} \mathbb{E}u_t = 0$. Therefore, $\mathbb{P}\left(\lim\limits_{t\to\infty} \frac{1}{t}\sum\limits_{s=1}^{t} u_s = 0\right) = 1$ and $\forall \varepsilon > 0$ $\mathbb{P}\left(\lim\limits_{t} \frac{1}{t}\sum\limits_{s=1}^{t} u_s < -\varepsilon\right) = 0$.

Fix $\varepsilon = \frac{2\alpha}{3}$ and T' such that $\frac{M-y_0}{T'} < \frac{\alpha}{3}$. Then

$$\mathbb{P}\left(\lim_{t\to\infty} y_t < M\right) \leq \mathbb{P}\left(\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} u_s < \frac{M - y_0}{t} - \alpha\right)$$
$$\leq \mathbb{P}\left(\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} u_s < -2\alpha/3\right) = 0.$$

So that $\mathbb{P}\left(\lim\limits_{t\to\infty} y_t < M\right) \xrightarrow[t\to\infty]{} 0$.

Now suppose that $\alpha \leq 0$ and $\beta > 1$. We first will show convergence in probability. We are going to show that

$$\forall M \quad \mathbb{P}(y_t \in [M, M+1]) \xrightarrow[t\to\infty]{} 0.$$

Then $\mathbb{P}(|y_t| < M) \to 0$. As $y_t \geq 0$, we only need to consider intervals in $\mathbb{R}_+$.

Consider events $S_t(M) = \begin{cases} y_t \in [M, M+1] \\ y_{t+s} > M+1 \ \forall s \geq 1. \end{cases}$ Such events are disjoint for $t \neq t'$ ($S_t \cap S_{t'} = \emptyset$ when $t \neq t'$). Thus,

$$\sum_{t=1}^{\infty} \mathbb{P}(S_t) \leq 1. \tag{23}$$

Choose $A > 1$ such that

$$\left(\mathbb{E}u^4 + \left(\mathbb{E}u^2\right)^2\right) \sum_{s=1}^{\infty} \frac{3s^2}{(As-1)^4} < \frac{1}{2}. \tag{24}$$

Also choose $\varepsilon > 0$ such that $\beta - \varepsilon > 1$ and M' such that $\varepsilon M' > |\alpha| + A$. Thus, conditional on being in $S_t(M')$, $\forall s \geq 1$

$$\begin{aligned} y_{t+s} &= [\alpha + \beta y_{t-1} + u_t]_+ \geq \alpha + \beta y_{t+s-1} + u_{t+s} \geq A + (\beta - \varepsilon) y_{t+s-1} + u_{t+s} \\ &\geq A + y_{t+s-1} + u_{t+s} \geq As + y_t + \sum_{r=1}^{s} u_{t+r}. \end{aligned} \tag{25}$$

If $y_t \geq M$, then from Eq. (25), if $A + u_{t+1} > 1$, we get $y_{t+1} > M + 1$. If also $2A + u_{t+1} + u_{t+2} > 1$, then similarly $y_{t+2} > M + 1$. Thus, if $Ap + \sum\limits_{r=1}^{p} u_{t+r} > 1$ for all $p = 1, \ldots, s$ we get $y_{t+s} > M + 1$. Let us calculate the probability that $Ap + \sum\limits_{r=1}^{p} u_{t+r} > 1$ for all $p \geq 1$.

$$\begin{aligned} &\mathbb{P}\left(As + \sum_{r=1}^{s} u_{t+r} \leq 1\right) = \mathbb{P}\left(\sum_{r=1}^{s} u_{t+r} \leq 1 - As\right) \leq \mathbb{P}\left(\left|\sum_{r=1}^{s} u_{t+r}\right| \leq As - 1\right) \\ &\leq \frac{\mathbb{E}\left(\sum\limits_{r=1}^{s} u_{t+r}\right)^4}{(As-1)^4} = \frac{s\mathbb{E}u^4 + 3s(s-1)\left(\mathbb{E}u^2\right)^2}{(As-1)^4} \leq \frac{3s^2\left(\mathbb{E}u^4 + \left(\mathbb{E}u^2\right)^2\right)}{(As-1)^4}, \end{aligned}$$

where we used the Markov inequality for $\left|\sum\limits_{r=1}^{s} u_{t+r}\right|^4$ to bound probability by expectation.

Therefore,

$$\begin{aligned} \mathbb{P}\left(Ap + \sum_{r=1}^{p} u_{t+r} > 1 \;\; \forall p \geq 1\right) &\geq 1 - \sum_{s=1}^{\infty} \mathbb{P}\left(As + \sum_{r=1}^{s} u_{t+r} \leq 1\right) \\ &\geq 1 - \sum_{s=1}^{\infty} \frac{3s^2\left(\mathbb{E}u^4 + \left(\mathbb{E}u^2\right)^2\right)}{(As-1)^4} \geq 1 - \frac{1}{2} = \frac{1}{2}. \end{aligned} \tag{26}$$

Thus, for $M \geq M'$

$$\begin{aligned} &\mathbb{P}(S_t(M)) \geq \mathbb{P}\left(y_t \in [M, M+1],\; Ap + \sum_{r=1}^{p} u_{t+r} > 1 \;\; \forall p \geq 1\right) \\ &\quad = \mathbb{P}\left(y_t \in [M, M+1]\right) \mathbb{P}\left(Ap + \sum_{r=1}^{p} u_{t+r} > 1 \;\; \forall p \geq 1\right) \geq 0.5\mathbb{P}\left(y_t \in [M, M+1]\right). \end{aligned} \tag{27}$$

Plugging Eq. (27) into Eq. (23), we get $\sum_{t=1}^{\infty} \mathbb{P}\left(y_t \in [M, M+1]\right) \leq 2$. Thus, the series converges, so it must be that $\mathbb{P}\left(y_t \in [M, M+1]\right) \xrightarrow[t\to\infty]{} 0$.

We have shown that for $M \geq M'$, $\mathbb{P}\left(y_t \in [M, M+1]\right) \xrightarrow[t\to\infty]{} 0$. Let us show that this also holds for $M < M'$.

We know that for all $M \geq M'$, $\lim_t \mathbb{P}\left(y_t \in [M, M+1]\right) = 0$. Thus, also $\lim_t \mathbb{P}\left(y_t \in [M', \beta M'+1]\right) = 0$. Suppose by contradiction that

$$\lim_t \mathbb{P}\left(y_t \in [\beta^{-1}M', M']\right) > 0.$$

But then with positive probability $u_{t+1} \in [-\alpha, -\alpha+1]$ and with positive probability $y_{t+1} = [\alpha + \beta y_t + u_{t+1}]_+ \in [M', \beta M'+1]$, so that we get a contradiction. Thus, $\lim_t \mathbb{P}\left(y_t \in [\beta^{-1}M', M']\right) = 0$. We can repeat the argument with $\beta^{-1}M'$ instead of M', then with $\beta^{-2}M'$ instead of $\beta^{-1}M'$ and so on. Thus, we get that $\mathbb{P}\left(y_t \in (0, M]\right) \xrightarrow[t\to\infty]{} 0$ for all $M > 0$ (as $\beta^{-k} \to 0$). If $\lim_t \mathbb{P}\left(y_t = 0\right) > 0$, then by a similar argument we must have that $\lim_t \mathbb{P}\left(y_t \in [1, 2]\right) > 0$, which is a contradiction. (Take $u_{t+1} \in [-\alpha+1, -\alpha+2]$.)

Therefore, $\mathbb{P}\left(y_t \in [M, M+1]\right) \xrightarrow[t\to\infty]{} 0$ for all M and $y_t \xrightarrow{\mathbb{P}} \infty$.

Now let us show that $y_t \xrightarrow{a.s.} \infty$. Note that we can choose $A(k)$ in Eq. (24) such that

$$\left(\mathbb{E}u^4 + \left(\mathbb{E}u^2\right)^2\right) \sum_{s=1}^{\infty} \frac{3s^2}{(A(k)s-1)^4} < \frac{1}{k}$$

and $M(k)$ such that $\varepsilon M(k) > |\alpha| + A(k)$. Then if $y_t > M(k)$, for Eq. (26) with A replaced by $A(k)$, we get that $y_{t+s} > M(k)$ for all $s \geq 1$ with probability at least $1 - \frac{1}{k}$.

Because $y_t \xrightarrow{\mathbb{P}} \infty$, for any M $\mathbb{P}\left(y_t > M\right) \xrightarrow[t\to\infty]{} 1$. Thus, for any M and $\delta > 0$ there exists T such that $\mathbb{P}\left(y_T > M\right) > 1 - \delta$. Therefore, if $M > M(k)$, $\lim_t y_t > M$ with probability of at least $(1-\delta)(1-1/k)$. Because δ is arbitrary, we must have $\mathbb{P}(\lim_t y_t > M) \geq 1 - 1/k$ for any $M > M(k)$. Because k can be chosen arbitrary, we must have $\mathbb{P}(\lim_t y_t > M) = 1$ for any M. (Note that if $M' > M''$, then $\mathbb{P}(\lim_t y_t > M'') \geq \mathbb{P}(\lim_t y_t > M')$). Thus, $y_t \xrightarrow{a.s.} \infty$. $\square$

Theorem A.6. *If $\beta = 1$, $\alpha = 0$, then y_t is mean-divergent ($\mathbb{E}y_t \xrightarrow[t\to\infty]{} \infty$).*

Proof. Because u_t has full support, with positive probability $u_t < -y_{t-1}$. Thus,

$$\mathbb{E}y_t = \mathbb{E}[y_{t-1} + u_t]_+ > \mathbb{E}(y_{t-1} + u_t) = \mathbb{E}y_{t-1},$$

and $\mathbb{E}y_t$ is a strictly increasing sequence of t. Therefore, either $\mathbb{E}y_t \to \infty$ or $\mathbb{E}y_t \to const$. Suppose that the latter is true. Then by Markov's inequality for any $C > 0$,

$$\mathbb{P}(y_t \geq C) \leq \frac{\mathbb{E}y_t}{C} \leq \frac{\lim_t \mathbb{E}y_t}{C}.$$

Let us choose C such that $\frac{\lim_t \mathbb{E}y_t}{C} \leq \frac{1}{2}$. Thus, for any t, $\mathbb{P}(y_t \geq C) \leq \frac{1}{2}$ and $\mathbb{P}(y_t < C) \geq \frac{1}{2}$.

$$\mathbb{E}y_t = \mathbb{E}[y_{t-1} + u_t]_+ (\mathbf{1}\{y_{t-1} \geq C\} + \mathbf{1}\{y_{t-1} < C\}), \tag{28}$$

$$\mathbb{E}[y_{t-1} + u_t]_+ \mathbf{1}\{y_{t-1} \geq C\} \geq \mathbb{E}(y_{t-1} + u_t)\mathbf{1}\{y_{t-1} \geq C\} = \mathbb{E}y_{t-1}\mathbf{1}\{y_{t-1} \geq C\}, \tag{29}$$

$$\begin{aligned} \mathbb{E}[y_{t-1} + u_t]_+ \mathbf{1}\{y_{t-1} < C\} &= \mathbb{E}[y_{t-1} + \max(-C, u_t)]_+ \mathbf{1}\{y_{t-1} < C\} \\ &\geq \mathbb{E}(y_{t-1} + \max(-C, u_t))\mathbf{1}\{y_{t-1} < C\} \\ &= \mathbb{E}y_{t-1}\mathbf{1}\{y_{t-1} < C\} + \mathbb{P}(y_{t-1} < C)\mathbb{E}\max(-C, u_t). \end{aligned} \tag{30}$$

Because $\mathbb{P}(y_t < C) \geq \frac{1}{2}$ and $\mathbb{E}\max(-C, u_t) > \mathbb{E}u_t = 0$, combining Eq. (28), (29), and (30), we get

$$\begin{aligned} \mathbb{E}y_t &\geq \mathbb{E}y_{t-1}\mathbf{1}\{y_{t-1} \geq C\} + \mathbb{E}y_{t-1}\mathbf{1}\{y_{t-1} < C\} + \mathbb{P}(y_{t-1} < C)\mathbb{E}\max(-C, u_t) \\ &\geq \mathbb{E}y_{t-1} + const,\ const > 0. \end{aligned}$$

Thus, $\mathbb{E}y_t \geq const \cdot t + const_1$, $const > 0$, and $\mathbb{E}y_t \to \infty$. $\square$

Lemma A.7. *If $\beta = 1$, $\alpha = 0$, then we can equivalently rewrite the evolution of y_t as follows*

$$y_t = [y_{t-1} + u_t]_+ = y_0 + \sum_{s=1}^{t} u_s + \sup_{r=0,\dots,t} \left[-y_0 - \sum_{s=1}^{r} u_s \right]_+ .$$

Proof. Define $z_t = y_0 + \sum_{s=1}^{t} u_s + \sup_{r=0,\dots,t} \left[-y_0 - \sum_{s=1}^{r} u_s \right]_+$, $z_0 = y_0$. Note that by definition z_t is always non-negative, as when $y_0 + \sum_{s=1}^{t} u_s$ becomes negative, we are adding its absolute value or even a larger positive number ($\sup_{r=0,\dots,t} \left[-y_0 - \sum_{s=1}^{r} u_s \right]_+$). Let us show that $z_t = y_t$ for all t. Let us proceed by induction.

By definition $z_0 = y_0 \geq 0$. Let us look at $t = 1$. If $y_0 + u_1 \geq 0$, then $\sup_{r=0,1} \left[-y_0 - \sum_{s=1}^{r} u_s \right]_+ = 0$ and $z_1 = y_0 + u_1 = y_1$. If $y_0 + u_1 < 0$, then $y_1 = [y_0 + u_1]_+ = 0$ and $\sup_{r=0,1} \left[-y_0 - \sum_{s=1}^{r} u_s \right]_+ = -y_0 - u_1 > 0$. Thus, $z_t = y_0 + u_1 + (-y_0 - u_1) = 0 = y_1$.

Suppose that $z_t = y_t$ for all $t \le t'$. Let us prove that $z_{t'+1} = y_{t'+1}$. First, suppose that $\exists p \in \{0,\dots,t'\}$ such that $\left[-y_0 - \sum_{s=1}^{p} u_s\right]_+ \ge \left[-y_0 - \sum_{s=1}^{t'+1} u_s\right]_+$. Thus, either $t+1$ is not an argmaximum or it is not a unique argmaximum over $\{0,\dots,t'+1\}$. In that case,

$$\sup_{r=0,\dots,t'} \left[-y_0 - \sum_{s=1}^{r} u_s\right]_+ = \sup_{r=0,\dots,t'+1} \left[-y_0 - \sum_{s=1}^{r} u_s\right]_+$$

and

$$\begin{aligned} z_{t'+1} &= y_0 + \sum_{s=1}^{t'+1} u_s + \sup_{r=0,\dots,t'+1} \left[-y_0 - \sum_{s=1}^{r} u_s\right]_+ \\ &= y_0 + \sum_{s=1}^{t'} u_s + u_{t'+1} + \sup_{r=0,\dots,t'} \left[-y_0 - \sum_{s=1}^{r} u_s\right]_+ = z_{t'} + u_{t'+1} = y_{t'} + u_{t'+1} \ge 0, \end{aligned}$$

where we used the induction hypothesis and the observation that $z_t \ge 0$ for all t

Thus,

$$y_{t'+1} = [y_{t'} + u_{t'+1}]_+ = y_{t'} + u_{t'+1} = z_{t'+1}.$$

Now suppose that $\forall p \in \{0,\dots,t'\}$, $\left[-y_0 - \sum_{s=1}^{p} u_s\right]_+ < \left[-y_0 - \sum_{s=1}^{t'+1} u_s\right]_+$. Thus, $\left[-y_0 - \sum_{s=1}^{t'+1} u_s\right]_+ > 0$ and

$$\begin{aligned} z_{t'+1} &= y_0 + \sum_{s=1}^{t'+1} u_s + \sup_{r=0,\dots,t'+1} \left[-y_0 - \sum_{s=1}^{r} u_s\right]_+ \\ &= y_0 + \sum_{s=1}^{t'+1} u_s + \left(-y_0 - \sum_{s=1}^{t'+1} u_s\right)_+ = 0. \end{aligned}$$

In turn,

$$\begin{aligned} y_{t'+1} = [y_{t'} + u_{t'+1}]_+ &= \left[y_0 + \sum_{s=1}^{t'} u_s + \sup_{r=0,\dots,t'} \left[-y_0 - \sum_{s=1}^{r} u_s\right]_+ u_{t'+1}\right]_+ \\ &= \left[y_0 + \sum_{s=1}^{t'+1} u_s + \sup_{r=0,\dots,t'} \left[-y_0 - \sum_{s=1}^{r} u_s\right]_+\right]_+ = 0 = z_{t'+1}, \end{aligned}$$

as $y_0 + \sum_{s=1}^{t'+1} u_s < 0$ and $\left[-y_0 - \sum_{s=1}^{p} u_s\right]_+ < -y_0 - \sum_{s=1}^{t'+1} u_s$ for all $p \in \{0,\dots,t'\}$.

Therefore, $z_{t'+1} = y_{t'+1}$. By induction we get that $y_t = z_t$ for all t. $\square$

Theorem A.8. *If $\beta = 1$, $\alpha = 0$, then for all $r \in (0,1]$, $\frac{1}{\sqrt{T}} y_{\lfloor rT \rfloor} \xrightarrow[T\to\infty]{d} \sigma|W(r)|$, where $\sigma = \mathbb{E}u^2$ and W is a standard Brownian motion.*

Proof. By Lemma A.7, y_t can be alternatively written as

$$y_t = y_0 + \sum_{s=1}^{t} u_s + \sup_{p=0,\dots,t} \left[-y_0 - \sum_{s=1}^{p} u_s \right]_+ .$$

This is Skorokhod transformation for $y_0 + \sum_{s=1}^{t} u_s$, which is a continuous transformation. Thus,

$$\frac{1}{\sqrt{T}} y_{\lfloor rT \rfloor} = \frac{1}{\sqrt{T}} y_0 + \frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor rT \rfloor} u_s + \sup_{p=0,\dots,\lfloor rT \rfloor} \left[-\frac{y_0}{\sqrt{T}} - \frac{1}{\sqrt{T}} \sum_{s=1}^{p} u_s \right]_+$$

By functional central limit theorem $\frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor rT \rfloor} u_s \xrightarrow[T\to\infty]{d} \sigma W(r)$. So that using the continuity of Skorokhod transformation,

$$\frac{1}{\sqrt{T}} y_{\lfloor rT \rfloor} \xrightarrow[T\to\infty]{d} \sigma W(r) + \sup_{p\in[0,r]} \left[\sigma W(p) \right]_+ \stackrel{d}{=} \sigma|W(r)|,$$

where the last equation, which gives equivalence in distribution, was first proved in Lévy (1948). (See, for example, Section 3.6.C in Karatzas and Shreve (2012).) ☐

Appendix B. Identification.

This subsection presents proofs of the identification theorems.

Proof of Theorem 5.

Proof. Define

$$\begin{aligned} q(y_{t-1}) :&= \mathbb{E}\left(y_t | y_t > 0,\, y_{t-1}\right) = \mathbb{E}\left(\alpha + \beta y_{t-1} + u_t | u_t > -\alpha - \beta y_{t-1},\, y_{t-1}\right) \\ &= \alpha + \beta y_{t-1} + \mathbb{E}\left(u_t | u_t > -\alpha - \beta y_{t-1}\right). \end{aligned}$$

For fixed value of y_{t-1}, $\mathbb{E}\left(u_t | u_t > -\alpha - \beta y_{t-1}\right) > \mathbb{E}u_t = 0$. However, in the limit when y_{t-1} goes to infinity, the expectation converges to zero if $\beta > 0$, as $-\alpha - \beta y \xrightarrow[y\to\infty]{} -\infty$. Thus,

$$\lim_{y_{t-1}\to\infty} \frac{q(y_{t-1})}{y_{t-1}} = \beta + \lim_{y_{t-1}\to\infty} \frac{\alpha + \mathbb{E}\left(u_t | u_t > -\alpha - \beta y_{t-1}\right)}{y_{t-1}} = \beta,$$

and β is identified. After β is identified, α can be identified from

$$\alpha = \lim_{y_{t-1}\to\infty} \left[q(y_{t-1}) - y_{t-1} \lim_{y\to\infty} \frac{q(y)}{y} \right].$$

When $\beta = 0$ it is still identified, as $\lim_{y_{t-1}\to\infty} \frac{\mathbb{E}(u_t | u_t > -\alpha - \beta y_{t-1})}{y_{t-1}} = \lim_{y_{t-1}\to\infty} \frac{\mathbb{E}(u_t | u_t > -\alpha)}{y_{t-1}} = 0$. However, α is not identified.

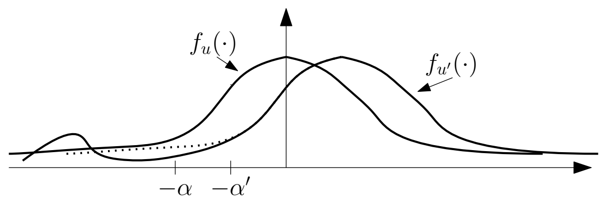

FIGURE 5. Densities of u and u'.

Suppose that $y_t = [\alpha + u_t]_+$. Fix some constant $C \neq 0$ and define $\alpha' = \alpha - C$, and a random variable u' such that

- If $x \geq -\alpha' = -\alpha + C$, then $f_{u'}(x) = f_u(x - C)$;
- If $x < -\alpha'$, then choose $f_{u'}(x)$ such that $\mathbb{E}u' = \mathbb{E}u = 0$.

The densities of u and u' are shown in Figure 5. For $x > -\alpha'$, $u' = u + C$, so that $y_t = [\alpha + u_t]_+ = [\alpha' + u'_t]_+$, and α is not identified. The reason is that error distribution below $-\alpha$ is not observed, so we can set it to arbitrary. That is, fixing $\mathbb{E}u = 0$ is no longer meaningful.

□

Proof of Theorem 6.

Proof. First, let us show that β is always identified. For any $M > 0$, conditioning on $y_{t-1} > M$ the ratio $\frac{y_t}{y_{t-1}}$ is well defined. Let us show that $\lim\limits_{M\to\infty}\lim\limits_{T\to\infty} \operatorname{med}\left(\frac{y_t}{y_{t-1}}|_{y_{t-1}>M}\right) = \beta$.

As $T \to \infty$, the median of $\frac{y_t}{y_{t-1}}|_{y_{t-1}>M}$ converges to the true median of a corresponding stationary process, which has a mass point at zero and then a continuous cdf on $\mathbb{R}_{++}$. Let us look at that median.

$$\frac{y_t}{y_{t-1}}|_{y_{t-1}>M} = \begin{cases} 0, & y_t = 0; \\ \beta + \frac{\alpha + u_t}{y_{t-1}}, & y_t > 0. \end{cases} \tag{31}$$

As $\operatorname{med}(u_t) = 0$, in half of the cases u_t is negative and in half of the cases it is positive. First, suppose $u_t > 0$. Then $\frac{y_t}{y_{t-1}} = \frac{[\alpha+\beta y_{t-1}+u_t]_+}{y_{t-1}} \geq \frac{[\alpha+\beta y_{t-1}]_+}{y_{t-1}} = \left[\beta + \frac{\alpha}{y_{t-1}}\right]_+ \geq \beta - \frac{|\alpha|}{M}$. Now suppose that $u_t < 0$. Then $\frac{y_t}{y_{t-1}} = \frac{[\alpha+\beta y_{t-1}+u_t]_+}{y_{t-1}} \leq \frac{[\alpha+\beta y_{t-1}]_+}{y_{t-1}} = \left[\beta + \frac{\alpha}{y_{t-1}}\right]_+ \leq \beta + \frac{|\alpha|}{M}$. Thus, the median must lie on the interval $[\beta - \frac{|\alpha|}{M}, \beta + \frac{|\alpha|}{M}]$. As M goes to infinity, the interval shrinks to one point, β. This gives us the desired result, $\lim\limits_{M\to\infty}\lim\limits_{T\to\infty} \operatorname{med}\left(\frac{y_t}{y_{t-1}}|_{y_{t-1}>M}\right) = \beta$, which implies that β is identified.

Second, let us show that if $\beta > 0$, then α is identified. We already know that β per se is identified, so it can be used to track down α. If $\beta > 0$, then here exists $M > 0$ such that $\alpha + \beta M > 0$. Note that if $\alpha \geq 0$, then any positive M will work, and if

$M' > M$, then $\beta M' > \beta M > -\alpha$. We will show that $\lim_{T\to\infty} \text{med}\left(y_t - by_{t-1}|_{y_{t-1}>M}\right) = \alpha$. As before, the median converges to the true median of a corresponding stationary process. For that process, suppose $u_t > 0$. Then, as $y_{t-1} > M$, $y_t = [\alpha + \beta y_{t-1} + u_t]_+ > 0$, so that $y_t - \beta y_{t-1} = \alpha + u_t > \alpha$. Now suppose that $u_t < 0$. Then $y_t = [\alpha + \beta y_{t-1} + u_t]_+ \leq [\alpha + \beta y_{t-1}]_+ = \alpha + \beta y_{t-1}$, where we used the facts that $y_{t-1} > M$ and $\alpha + \beta M > 0$. Thus, $y_t - \beta y_{t-1} \leq \alpha + \beta y_{t-1} - \beta y_{t-1} = \alpha$. To sum up, in half of the cases our random variable is above α and in half below. Thus, $\alpha = \lim_{T\to\infty} \text{med}\left(y_t - \beta y_{t-1}|_{y_{t-1}>M}\right)$. We do not know the value of M, as it depends on α, however, taking the limit as M goes to infinity gets rid of any such dependence and α is identified.

Thirdly, suppose that $\beta = 0$ and let us show that if α is non-negative, then it is identified. We proceed similar to the previous arguments and show that $\alpha = \lim_{T\to\infty} \text{med} y_t$. Suppose $u_t > 0$. Then, as $\alpha \geq 0$, $y_t = [\alpha + u_t]_+ = \alpha + u_t > a$. Now suppose that $u_t < 0$. Then $y_t = [\alpha + u_t]_+ \leq [\alpha]_+ = \alpha$. Thus, $\alpha = \lim_{T\to\infty} \text{med} y_t$. Note that when $\alpha < 0$, the median of y_t will be zero, because in the more than half of the cases y_t will be negative. To separate those cases from $\alpha = 0$ it is enough to look at the average length the process spends at zero. Once y_t hits zero, it stays at zero with probability $F(-a)$. Thus, the average length of a period of zeroes is

$$
\begin{aligned}
&1 \cdot (1 - F(-\alpha)) + 2 \cdot F(-\alpha)(1 - F(-\alpha)) + 3 \cdot F^2(-\alpha)(1 - F(-\alpha)) + \dots \\
&= (1 - F(-\alpha)) \sum_{k=1}^{\infty} k F^{k-1}(-\alpha) = \frac{1}{1 - F(-\alpha)}.
\end{aligned}
$$

As $\text{med}(u_t) = 0$, we get $F(0) = 0.5$ and when $\alpha = 0$, the average length of a period of zeros is $\frac{1}{1-0.5} = 2$. In contrast, when $\alpha < 0$, $F(-\alpha) > F(0) = 0.5$, so that once the process hits zero, it stays at zero for a longer time: $\frac{1}{1-F(-\alpha)} > \frac{1}{1-F(0)} = 2$.

Finally, suppose that $y_t = [\alpha + u_t]_+$, $\alpha < 0$. We will show that multiple values of α are consistent with the data. Fix some constant $C > 0$ and define $\alpha' = \alpha - C$, and random variable u' such that

- If $x \geq -\alpha' = -\alpha + C$, then $f_{u'}(x) = f_u(x - C)$;
- If $x < -\alpha'$, then choose $f_{u'}(x)$ such that $\text{med}(u') = \text{med}(u) = 0$.

The densities of u and u' are shown in Figure 6. For $x > -\alpha'$, $u' = u + C$, so that $y_t = [\alpha + u_t]_+ = [\alpha' + u'_t]_+$, and α is not identified. The reason is that the error distribution below $-\alpha$ is not observed, so we can set it to be arbitrary. Moreover, as $\mathbb{P}(u > -\alpha) < \mathbb{P}(u > 0) = 0.5$ we can adjust the distribution below $-\alpha$ so that the median can be any number below $-\alpha$. That is, fixing $\text{med}(u) = 0$ is no longer meaningful if more than half of u's distribution is unobserved. $\square$

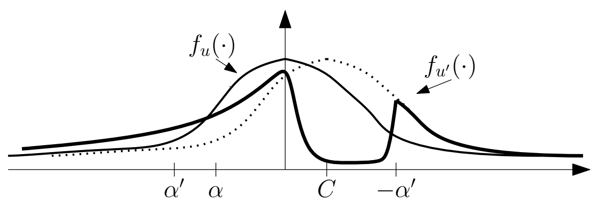

FIGURE 6. Densities of u and u'.

Appendix C. Properties of the estimators.

All properties of estimators are proved in this section. We use the following lemma to justify cases, where sums are approximated by expectations.

Lemma C.1. *If y_t converges to a stationary process in Theorem 2 and Theorem 3 and there exists $\mathbb{E}|u|^{4+\varepsilon}$ for some $\varepsilon > 0$, then as $T \to \infty$*

$$\frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{\mathbb{P}} \mathbb{E}y,\ \frac{1}{T}\sum_{t=1}^{T} y_t^2 \xrightarrow{\mathbb{P}} \mathbb{E}y^2,\ \frac{1}{T}\sum_{t=1}^{T} y_t z_t \xrightarrow{\mathbb{P}} \mathbb{E}yz,\ \frac{1}{T}\sum_{t=1}^{T} z_t \xrightarrow{\mathbb{P}} \mathbb{E}z,\ \frac{1}{T}\sum_{t=1}^{T} z_t^2 \xrightarrow{\mathbb{P}} \mathbb{E}z^2,$$

where expectations are taken with respect to the stationary distribution.

Proof. First note that if the error u_t has a moment of order k, then y_t also has a moment of order k.

Let us show that $Cov(y_s, y_{s+t}) \xrightarrow[t\to\infty]{} 0$.

$$\begin{aligned} Cov(y_s, y_{s+t}) &= \mathbb{E}y_s y_{t+s} - \mathbb{E}y_s \mathbb{E}y_{t+s} = \mathbb{E}y_s y_{t+s}\mathbf{1}(y_s < M, y_{t+s} < M) - \mathbb{E}y_s\mathbb{E}y_{t+s} \\ &+ \mathbb{E}y_s y_{t+s}\mathbf{1}(y_s \geq M, y_{t+s} < M) + \mathbb{E}y_s y_{t+s}\mathbf{1}(y_s < M, y_{t+s} \geq M) \\ &+ \mathbb{E}y_s y_{t+s}\mathbf{1}(y_s \geq M, y_{t+s} \geq M) \end{aligned}$$

Because y_t is mixing and converges in distribution to a random variable, its limit is independent of y_s,

$$\mathbb{E}y_s y_{t+s}\mathbf{1}(y_s < M, y_{t+s} < M) \to \mathbb{E}y_s\mathbf{1}(y_s < M)\mathbb{E}y_{s+t}\mathbf{1}(y_{s+t} < M).$$

As M goes to infinity, $\mathbb{E}y_s\mathbf{1}(y_s < M) \to \mathbb{E}y_s$ for all s. Thus, by choice of M large enough we can make $\lim_{t\to\infty}(\mathbb{E}y_s y_{t+s}\mathbf{1}(y_s < M, y_{t+s} < M) - \mathbb{E}y_s\mathbb{E}y_{t+s})$ arbitrarily close to zero. Moreover,

$$\begin{aligned} &\mathbb{E}y_s y_{t+s}\mathbf{1}(y_s \geq M, y_{t+s} < M) \leq \mathbb{E}\frac{y_s^{1+\varepsilon/2}}{M^{\varepsilon/2}} y_{t+s}\mathbf{1}(y_s \geq M, y_{t+s} < M) \\ &\leq \frac{1}{M^{\varepsilon/2}}\mathbb{E}(y_s^{2+\varepsilon} + y_{t+s}^2)\mathbf{1}(y_s \geq M, y_{t+s} < M) \leq \frac{1}{M^{\varepsilon/2}}\mathbb{E}(y_s^{2+\varepsilon} + y_{t+s}^2) \xrightarrow[M\to\infty]{} 0. \end{aligned}$$

Similarly,

$$\mathbb{E}y_s y_{t+s}\mathbf{1}(y_s < M, y_{t+s} \geq M) \leq \frac{1}{M^{\varepsilon/2}}\mathbb{E}(y_s^2 + y_{t+s}^{2+\varepsilon}) \xrightarrow[M\to\infty]{} 0.$$

Finally,

$$\begin{aligned}&\mathbb{E}y_s y_{t+s}\mathbf{1}(y_s \geq M, y_{t+s} \geq M) \leq \mathbb{E}(y_s^2 + y_{t+s}^2)\mathbf{1}(y_s \geq M, y_{t+s} \geq M)\\ &\leq \frac{1}{M^{\varepsilon}}\mathbb{E}(y_s^{2+\varepsilon} + y_{t+s}^{2+\varepsilon})\mathbf{1}(y_s \geq M, y_{t+s} \geq M) \leq \frac{1}{M^{\varepsilon}}\mathbb{E}(y_s^{2+\varepsilon} + y_{t+s}^{2+\varepsilon}) \xrightarrow[M\to\infty]{} 0.\end{aligned}$$

Thus, by choice of M large enough, $\mathbb{E}y_s y_{t+s}\mathbf{1}(y_s \geq M, y_{t+s} < M) + \mathbb{E}y_s y_{t+s}\mathbf{1}(y_s < M, y_{t+s} \geq M) + \mathbb{E}y_s y_{t+s}\mathbf{1}(y_s \geq M, y_{t+s} \geq M)$ can be made arbitrarily close to zero. That is, $Cov(y_s, y_{s+t}) \to 0$ as $t \to \infty$. This means that the variance of $\frac{1}{T}\sum_{t=1}^{T} y_t$ goes to zero as $T \to \infty$, so that law of large numbers holds for y_t.

To see that the variance of $\frac{1}{T}\sum_{t=1}^{T} y_t$ goes to zero as $T \to \infty$, note that

$$\mathbb{V}\sum_{t=1}^{T} y_t = \mathbb{E}\left(\sum_{t=1}^{T} y_t - T\mathbb{E}y\right)^2 = \sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(y_t - \mathbb{E}y)(y_s - \mathbb{E}y) \leq \sum_{s=1}^{T}\left(\sum_{t=0}^{T} Cov(y_s, y_{s+t})\right)$$

Because $Cov(y_s, y_{s+t})$ goes to zero as $t \to \infty$, $\sum_{t=0}^{T} Cov(y_s, y_{s+t}) = o(T)$ (otherwise terms would not disappear as $t \to \infty$). Thus, $\sum_{s=1}^{T}\left(\sum_{t=0}^{T} Cov(y_s, y_{s+t})\right) = o(T^2)$ (there are T terms each of order $o(T)$). That is, $\mathbb{V}\sum_{t=1}^{T}\frac{1}{T}y_t = \frac{o(T^2)}{T^2} = o(1)$ and the law of large numbers holds for y_t.

The same logic applies to the four other limits. Corresponding moments of z_t are finite, because z_t is bounded by the maximum of finite number of variables y_{ijs}, $s = t, \ldots, t-H+1$ (see Assumption 6). □

Proof of Theorem 7.

Proof. Denote the LAD estimate as $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$. The estimate is the solution to minimization problem (6). The corresponding first order conditions are

$$\begin{aligned}&\sum_{t=1}^{T}\operatorname{sgn}(y_t - a - by_{t-1} - cz_{t-1}) = 0,\\ &\sum_{t=1}^{T} y_{t-1}\operatorname{sgn}(y_t - a - by_{t-1} - cz_{t-1}) = 0,\\ &\sum_{t=1}^{T} z_{t-1}\operatorname{sgn}(y_t - a - by_{t-1} - cz_{t-1}) = 0.\end{aligned} \tag{32}$$

Define $\Delta\alpha = \hat{\alpha} - \alpha$, $\Delta\beta = \hat{\beta} - \beta$, $\Delta\gamma = \hat{\gamma} - \gamma$, $v_{t+1} = \max(u_{t+1}, -\alpha - \beta y_t - \gamma z_t)$. Then, noting that $y_t = \alpha + \beta y_{t-1} + \gamma z_{t-1} + v_t$, we can rewrite Eq. (32) as

$$\begin{aligned}
&\sum_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) = 0,\\
&\sum_t y_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) = 0,\\
&\sum_t z_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) = 0.
\end{aligned} \tag{33}$$

Dividing each line of Eq. (33) by T and applying the law of large numbers, we get

$$\begin{aligned}
&\mathbb{E}\operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) = 0,\\
&\mathbb{E} y_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) = 0,\\
&\mathbb{E} z_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) = 0.
\end{aligned} \tag{34}$$

Let us linearize the expectations in Eq. (34).

$$\begin{aligned}
&\mathbb{E}\operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t)\\
&= \int_y \int_z \int_v \operatorname{sgn}(v - \Delta\alpha - \Delta\beta y - \Delta\gamma z)\, f_{y,z,v}(y,z,v) dv dz dy\\
&= \int_y \int_z \int\limits_{\Delta\alpha + \Delta\beta y + \Delta\gamma z}^{\infty} f_{y,z,v}(y,z,v) dv dz dy - \int_y \int_z \int\limits_{-\infty}^{\Delta\alpha + \Delta\beta y + \Delta\gamma z} f_{y,z,v}(y,z,v) dv dz dy\\
&= \mathbb{E}_{y,z}\left(1 - F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z | y,z)\right) - \mathbb{E}_{y,z} F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z | y,z)\\
&= 1 - 2\mathbb{E}_{y,z} F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z | y,z).
\end{aligned} \tag{35}$$

Taylor expanding around $\Delta\alpha + \Delta\beta y + \Delta\gamma z = 0$, we can rewrite Eq. (35) as

$$1 - 2\mathbb{E}_{y,z}\left(F_{v|y,z}(0|y,z) + f_{v|y,z}(0|y,z)(\Delta\alpha + \Delta\beta y + \Delta\gamma z | y,z)\right).$$

As $v_{t+1} = \max(u_{t+1}, -\alpha - \beta y_t - \gamma z_t)$ and $-\alpha - \beta y_t - \gamma z_t < 0$ for $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$, the density of v_{t+1} has a mass point at $-\alpha - \beta y_t - \gamma z_t$ and then coincides with the density of u_{t+1}. Thus, $F_{v|y,z}(0|y,z) = F_u(0) = 0.5$ and $f_{v|y,z}(0|y,z) = f_u(0)$. To sum up, the first order condition with respect to α becomes

$$-2 f_u(0)\mathbb{E}(\Delta\alpha + \Delta\beta y + \Delta\gamma z) = 0. \tag{36}$$

Let us now analyze the second line of Eq. (34) in a similar fashion.

$$\begin{aligned}
&\mathbb{E}y_t \mathrm{sgn}\left(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t\right) \\
&= \int_y \int_z \int_v y \mathrm{sgn}\left(v - \Delta\alpha - \Delta\beta y - \Delta\gamma z\right) f_{y,z,v}(y,z,v) dv dz dy \\
&= \mathbb{E}_{y,z} y \left(1 - F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z|y,z)\right) - \mathbb{E}_{y,z} y F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z|y,z) \\
&= \mathbb{E}y - 2\mathbb{E}_{y,z} y F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z|y,z) \\
&\approx \mathbb{E}y - 2\mathbb{E}_{y,z} y \left(F_{v|y,z}(0|y,z) + f_{v|y,z}(0|y,z)(\Delta\alpha + \Delta\beta y + \Delta\gamma z|y,z)\right) \\
&= -2f_u(0)\mathbb{E}y(\Delta\alpha + \Delta\beta y + \Delta\gamma z).
\end{aligned} \tag{37}$$

Thus, first order condition with respect to β becomes

$$-2f_u(0)\mathbb{E}y(\Delta\alpha + \Delta\beta y + \Delta\gamma z) = 0. \tag{38}$$

Similarly, the first order condition with respect to γ becomes

$$-2f_u(0)\mathbb{E}z(\Delta\alpha + \Delta\beta y + \Delta\gamma z) = 0. \tag{39}$$

Combining Eq. (36), (38), and (39), we get

$$\begin{cases} \Delta\alpha + \Delta\beta\mathbb{E}y + \Delta\gamma\mathbb{E}z = 0, \\ \Delta\alpha\mathbb{E}y + \Delta\beta\mathbb{E}y^2 + \Delta\gamma\mathbb{E}yz = 0, \\ \Delta\alpha\mathbb{E}z + \Delta\beta\mathbb{E}yz + \Delta\gamma\mathbb{E}z^2 = 0. \end{cases}$$

Equivalently in the matrix form

$$\begin{pmatrix} 1 & \mathbb{E}y & \mathbb{E}z \\ \mathbb{E}y & \mathbb{E}yz & \mathbb{E}yz \\ \mathbb{E}z & \mathbb{E}yz & \mathbb{E}z^2 \end{pmatrix} \begin{pmatrix} \Delta\alpha \\ \Delta\beta \\ \Delta\gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Matrix $\begin{pmatrix} 1 & \mathbb{E}y & \mathbb{E}z \\ \mathbb{E}y & \mathbb{E}yz & \mathbb{E}yz \\ \mathbb{E}z & \mathbb{E}yz & \mathbb{E}z^2 \end{pmatrix}$ is nonsingular. If it were singular, it would have an eigenvector $(\lambda_1, \lambda_2, \lambda_2)$ corresponding to the zero eigenvalue. Then the random variable $\lambda_1 + \lambda_2 y_t + \lambda_3 z_t$ must have zero second moment. That is, $\lambda_1 + \lambda_2 y_t + \lambda_3 z_t$ must be zero, which contradicts conditions of the theorem.

Thus, the only solution to the above system is $\Delta\alpha = 0$, $\Delta\beta = 0$, $\Delta\gamma = 0$. That is, the LAD estimator is consistent and converges to the true value of the parameters as $T \to \infty$. □

Proof of Theorem 8.

Proof. This proof follows the lines of Powell (1984).

Define $\theta = (\alpha, \beta, \gamma)'$, $x_t = (1,\ y_{t-1},\ z_{t-1})'$, $\mathcal{F}_t = \{y_t, z_t, y_{t-1}, z_{t-1}, \ldots\}$, and

$$S_T(\theta) = \frac{1}{T}\sum_t |y_t - [\alpha + \beta y_{t-1} + \gamma z_{t-1}]_+| = \frac{1}{T}\sum_t |[x_t'\theta_0 + u_t]_+ - [x_t'\theta]_+|\,,$$

where θ_0 corresponds to the true value of θ.

We want to show that $S_T(\theta) - S_T(\theta_0)$ is uniformly bounded away from zero for large T and $||\theta - \theta_0|| > \varepsilon$ for any $\varepsilon > 0$. Then $\hat{\theta}_{LAD} \underset{\mathbb{P}}{\to} \theta_0$.

Let us write

$$\begin{aligned}
Q_T(\theta) &:= S_T(\theta) - S_T(\theta_0) \\
&= \frac{1}{T}\sum_t \Bigg[|[x_t'\theta_0 + u_t]_+ - [x_t'\theta]_+| - |[x_t'\theta_0 + u_t]_+ - [x_t'\theta_0]_+| \\
&- \mathbb{E}\left(|[x_t'\theta_0 + u_t]_+ - [x_t'\theta]_+| - |[x_t'\theta_0 + u_t]_+ - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right)\Bigg] \\
&+ \frac{1}{T}\sum_t \mathbb{E}\left(|[x_t'\theta_0 + u_t]_+ - [x_t'\theta]_+| - |[x_t'\theta_0 + u_t]_+ - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right)
\end{aligned} \tag{40}$$

Let us analyze the first summation in Eq. (40). Define

$$s_t(\theta, x_t) := |[x_t'\theta_0 + u_t]_+ - [x_t'\theta]_+| - |[x_t'\theta_0 + u_t]_+ - [x_t'\theta_0]_+|$$

and note that $s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})$ is a martingale difference sequence. Thus,

$$\mathbb{E}\left(\sum_t \left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right)\right)^2 = \sum_t \mathbb{E}\left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right)^2.$$

$$\begin{aligned}
\mathbb{E}\left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right)^2 = \mathbb{E}\Bigg(&|[x_t'\theta_0 + u_t]_+ - [x_t'\theta]_+| - |[x_t'\theta_0 + u_t]_+ - [x_t'\theta_0]_+| \\
- \int \left(|[x_t'\theta_0 + u]_+ - [x_t'\theta]_+| - |[x_t'\theta_0 + u]_+ - [x_t'\theta_0]_+|\right) f(u)du\Bigg)^2,
\end{aligned}$$

which is a function of stationary distribution of y_t and z_t and parameter θ. Because $\theta \in \Theta$, which is compact, and second moments of y and z exist, $\mathbb{E}\left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right)^2$ is bounded by some constant C (which depends on first two moments of y and z). Therefore, by Markov inequality, for any $a > 0$

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_t \left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right)\right| > a\right) \leq \frac{\mathbb{E}\left|\sum_t \left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right)\right|^2}{T^2a^2}$$
$$= \frac{T\mathbb{E}\left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right)^2}{T^2a^2} \leq \frac{C}{Ta^2} \xrightarrow[T\to\infty]{} 0.$$

Thus,

$$\frac{1}{T}\sum_t \left(s_t(\theta, x_t) - \mathbb{E}(s_t(\theta, x_t)|\mathcal{F}_{t-1})\right) \xrightarrow[T\to\infty]{\mathbb{P}} 0$$

uniformly over θ.

Let us now analyze the second summation in Eq. (40). We are going to show that $\mathbb{E}\left(|y_t - [x_t'\theta]_+| - |y_t - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right)$ is always non-negative.

Because $y_t = [x_t'\theta_0 + u_t]_+$,

$$\begin{aligned}\mathbb{E}(|y_t - [x_t'\theta]_+||\mathcal{F}_{t-1}) = \mathbf{1}(x_t'\theta < 0) \int_{-x_t'\theta_0}^{\infty} (x_t'\theta_0 + u) f_u(u)du \\ + \mathbf{1}(x_t'\theta \geq 0)\left(x_t'\theta F_u(-x_t'\theta_0) + \int_{-x_t'\theta_0}^{\infty} |u - x_t'(\theta - \theta_0)| f_u(u)du\right)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(|y_t - [x_t'\theta_0]_+||\mathcal{F}_{t-1}) = \mathbf{1}(x_t'\theta_0 < 0) \int_{-x_t'\theta_0}^{\infty} (x_t'\theta_0 + u) f_u(u)du \\ + \mathbf{1}(x_t'\theta_0 \geq 0)\left(x_t'\theta_0 F_u(-x_t'\theta_0) + \int_{-x_t'\theta_0}^{\infty} |u| f_u(u)du\right).\end{aligned}$$

Therefore, omitting the derivations, we get

$$
\begin{aligned}
&\mathbb{E}\left(|y_t - [x_t'\theta]_+| - |y_t - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right) \\
&= 2\mathbf{1}(x_t'\theta_0 \geq 0, x_t'\theta < 0) \int_{-x_t'\theta_0}^{0} (x_t'\theta_0 + u) f_u(u) du \\
&+ 2\mathbf{1}(x_t'\theta_0 < 0, x_t'\theta \geq 0) \left(\int_{-x_t'\theta_0}^{x_t'(\theta-\theta_0)} (x_t'(\theta - \theta_0) - u) f_u(u) du + \int_{0}^{-x_t'\theta_0} x_t'\theta f_u(u) du \right) \\
&+ 2\mathbf{1}(x_t'\theta_0 \geq 0, x_t'\theta \geq 0) \int_{0}^{x_t'(\theta-\theta_0)} (x_t'(\theta - \theta_0) - u) f_u(u) du \geq 0,
\end{aligned}
\tag{41}
$$

as every function under integral is non-negative over the domain of integration.

Because all terms in the Eq. (41) are non-negative,

$$
\begin{aligned}
\mathbb{E}\left(|y_t - [x_t'\theta]_+| - |y_t - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right) &\geq 2\mathbf{1}(x_t'\theta_0 \geq 0, x_t'\theta < 0) \int_{-x_t'\theta_0}^{0} (x_t'\theta_0 + u) f_u(u) du \\
+ 2\mathbf{1}(x_t'\theta_0 \geq 0, x_t'\theta \geq 0) \int_{0}^{x_t'(\theta-\theta_0)} (x_t'(\theta - \theta_0) - u) f_u(u) du.
\end{aligned}
$$

Moreover, for $R \geq 0$ such that M_R is nonsingular and any $\tau \in (0, R]$,

$$
\begin{aligned}
&\mathbb{E}\left(|y_t - [x_t'\theta]_+| - |y_t - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right) \\
&\geq 2\mathbf{1}(x_t'\theta_0 \geq R, x_t'\theta < 0)\mathbf{1}(|x_t'(\theta - \theta_0)| \geq \tau) \int_{-\tau}^{0} (\tau + u) f_u(u) du \\
&+ 2\mathbf{1}(x_t'\theta_0 \geq R, x_t'\theta \geq 0)\mathbf{1}(|x_t'(\theta - \theta_0)| \geq \tau) \int_{0}^{\tau} (\tau - u) f_u(u) du \\
&\geq 2 \min\left(\int_{-\tau}^{0} (\tau + u) f_u(u) du, \int_{0}^{\tau} (\tau - u) f_u(u) du \right) \mathbf{1}(x_t'\theta_0 \geq R)\mathbf{1}(|x_t'(\theta - \theta_0)| \geq \tau).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\frac{1}{T}\sum_t \mathbb{E}\left(|y_t - [x_t'\theta]_+| - |y_t - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right) \\
&\geq \frac{2}{T} \min\left(\int_{-\tau}^{0} (\tau + u) f_u(u) du, \int_{0}^{\tau} (\tau - u) f_u(u) du \right) \sum_t \mathbf{1}(x_t'\theta_0 \geq R)\mathbf{1}(|x_t'(\theta - \theta_0)| \geq \tau).
\end{aligned}
$$

As T goes to infinity, $\frac{1}{T}\sum_t \mathbf{1}(x_t'\theta_0 \geq R)\mathbf{1}(|x_t'(\theta-\theta_0)| \geq \tau)$ converges to

$$\begin{aligned}&\mathbb{E}\mathbf{1}(x_t'\theta_0 \geq R,\, |x_t'(\theta-\theta_0)| \geq \tau) = \mathbb{P}(x_t'\theta_0 \geq R,\, |x_t'(\theta-\theta_0)| \geq \tau)\\ &= \mathbb{P}(x_t'\theta_0 \geq R)\mathbb{P}(|x_t'(\theta-\theta_0)| \geq \tau|\, x_t'\theta_0 \geq R).\end{aligned}$$

Because M_R is nonsingular, $\mathbb{P}(x_t'\theta_0 \geq R) > 0$ (otherwise indicator in M_R will always be zero, so that the matrix M_R will be identically zero).

We are going to prove that

$$\mathbb{P}(|x_t'(\theta-\theta_0)| \geq \tau_0|\, x_t'\theta_0 \geq R) \geq C_1 > 0,$$

where $\tau_0 = const \cdot ||\theta-\theta_0||^2$ and $||\cdot||$ denotes L_2 norm. This will imply that the sum of conditional expectations (Eq. (41)) is bounded from zero uniformly in $||\theta-\theta_0||$. Thus, initial summation (Eq. (40)) is also uniformly bounded from zero, so that as T goes to infinity, for any $\theta \neq \theta_0$ with probability tending to one we have

$$\frac{1}{T}\sum_t \mathbb{E}\left(|y_t - [x_t'\theta]_+| - |y_t - [x_t'\theta_0]_+|\,|\mathcal{F}_{t-1}\right) \geq const(||\theta-\theta_0||) > 0$$

and

$$\lim_{T\to\infty} \mathbb{P}(Q_T(\theta) \geq const(||\theta-\theta_0||)) = 1,$$

where constant is increasing in $||\theta-\theta_0||$.

In contrast, for any T, $Q_T(\theta_0) = 0$. Thus, we must have that LAD estimate $\hat{\theta}_T$ converges to θ_0 as $T\to\infty$, as $Q_T(\theta)$ is bounded from zero for $\theta\neq\theta_0$ and T large enough.

Let us show that $\mathbb{P}(|x_t'(\theta-\theta_0)| \geq \tau_0|\, x_t'\theta_0 \geq R) \geq C_1 > 0$, where $\tau_0 = const\cdot||\theta-\theta_0||^2$. Define by $\lambda_{\min}$ the minimal eigenvalue of matrix $\mathbb{E}(x_t x_t'|x_t'\theta_0 > R)$. It is non-zero, because the matrix is nonsingular. Then

$$\mathbb{E}\left(|x_t'(\theta-\theta_0)|^2\,|x_t'\theta_0 \geq R\right) \geq ||\theta-\theta_0||^2\lambda_{\min},$$

as such conditional expectation corresponds to the value of the quadratic form $\mathbb{E}(x_t x_t'|x_t'\theta_0 > R)$ on vector $\theta-\theta_0$.

Choose $\varepsilon > 0$ such that $\varepsilon < ||\theta-\theta_0||^2\lambda_{\min}$ and note that for any $A > 0$

$$\begin{aligned}&\mathbb{E}\left(|x_t'(\theta-\theta_0)|^2\,|x_t'\theta_0 \geq R\right) = \mathbb{E}\left(|x_t'(\theta-\theta_0)|^2\mathbf{1}(|x_t'(\theta-\theta_0)|^2) < A\,|x_t'\theta_0 \geq R\right)\\ &+\mathbb{E}\left(|x_t'(\theta-\theta_0)|^2\mathbf{1}(|x_t'(\theta-\theta_0)|^2 \geq A)\,|x_t'\theta_0 \geq R\right)\\ &\leq \mathbb{E}\left(|x_t'(\theta-\theta_0)|^2\mathbf{1}(|x_t'(\theta-\theta_0)|^2 < A)\,|x_t'\theta_0 \geq R\right)\\ &+\mathbb{E}\left(\frac{|x_t'(\theta-\theta_0)|^4}{A}\mathbf{1}(|x_t'(\theta-\theta_0)|^2 \geq A)\,|x_t'\theta_0 \geq R\right)\\ &\leq \mathbb{E}\left(|x_t'(\theta-\theta_0)|^2\mathbf{1}(|x_t'(\theta-\theta_0)|^2 < A)\,|x_t'\theta_0 \geq R\right) + \frac{\mathbb{E}\left(|x_t'(\theta-\theta_0)|^4\,|x_t'\theta_0 \geq R\right)}{A},\end{aligned}$$

where the fourth moment exists, because u_t has fourth moment. Choose $A(\varepsilon)$ such that $A(\varepsilon) > \frac{||\theta-\theta_0||^4 \mathbb{E}\left(||x_t||^4 \,|x_t'\theta_0 \geq R\right)}{\varepsilon}$. Then

$$\frac{\mathbb{E}\left(|x_t'(\theta-\theta_0)|^4 \,|x_t'\theta_0 \geq R\right)}{A} < \frac{\varepsilon \mathbb{E}\left(|x_t'(\theta-\theta_0)|^4 \,|x_t'\theta_0 \geq R\right)}{||\theta-\theta_0||^4 \mathbb{E}\left(||x_t||^4 \,|x_t'\theta_0 \geq R\right)} \leq \varepsilon,$$

as $\mathbb{E}\left(|x_t'(\theta-\theta_0)|^4 \,|x_t'\theta_0 \geq R\right) \leq ||\theta-\theta_0||^4 \mathbb{E}\left(||x_t||^4 \,|x_t'\theta_0 \geq R\right)$. Thus,

$$\begin{aligned}\mathbb{E}\left(|x_t'(\theta-\theta_0)|^2 \mathbf{1}(|x_t'(\theta-\theta_0)|^2 < A(\varepsilon)) \,|x_t'\theta_0 \geq R\right) &\geq \mathbb{E}\left(|x_t'(\theta-\theta_0)|^2 \,|x_t'\theta_0 \geq R\right) - \varepsilon \\ &\geq ||\theta-\theta_0||^2 \lambda_{\min} - \varepsilon > 0.\end{aligned}$$

Finally,

$$\begin{aligned}&\mathbb{E}\left(|x_t'(\theta-\theta_0)|^2 \mathbf{1}(|x_t'(\theta-\theta_0)|^2 < A(\varepsilon)) \,|x_t'\theta_0 \geq R\right) \\ &= \mathbb{E}\left(|x_t'(\theta-\theta_0)|^2 \mathbf{1}(|x_t'(\theta-\theta_0)|^2 < A(\varepsilon)) \mathbf{1}(|x_t'(\theta-\theta_0)| \geq \tau) \,|x_t'\theta_0 \geq R\right) \\ &+ \mathbb{E}\left(|x_t'(\theta-\theta_0)|^2 \mathbf{1}(|x_t'(\theta-\theta_0)|^2 < A(\varepsilon)) \mathbf{1}(|x_t'(\theta-\theta_0)| < \tau) \,|x_t'\theta_0 \geq R\right) \\ &\leq A(\varepsilon) \mathbb{P}(|x_t'(\theta-\theta_0)| \geq \tau |\, x_t'\theta_0 \geq R) + \tau,\end{aligned}$$

so that for $\tau \in (0, ||\theta-\theta_0||^2 \lambda_{\min} - \varepsilon)$,

$$\mathbb{P}(|x_t'(\theta-\theta_0)| \geq \tau |\, x_t'\theta_0 \geq R) \geq \frac{||\theta-\theta_0||^2 \lambda_{\min} - \varepsilon - \tau}{A(\varepsilon)} = C_1 > 0.$$

Fixing $\varepsilon = \frac{1}{2}||\theta-\theta_0||^2 \lambda_{\min}$, $A(\varepsilon) = 4\frac{||\theta-\theta_0||^2 \mathbb{E}\left(||x_t||^4 \,|x_t'\theta_0 \geq R\right)}{\lambda_{\min}}$, $\tau = \frac{1}{4}||\theta-\theta_0||^2 \lambda_{\min}$. Then

$$\mathbb{P}(|x_t'(\theta-\theta_0)| \geq \tau |\, x_t'\theta_0 \geq R) \geq \frac{\lambda_{\min}^2}{16 \mathbb{E}\left(||x_t||^4 | x_t'\theta_0 \geq R\right)} = C_1 > 0.$$

If $||\theta'-\theta_0|| > ||\theta-\theta_0||$, then

$$\begin{aligned}&\mathbb{P}(|x_t'(\theta'-\theta_0)| \geq \frac{1}{4}||\theta-\theta_0||^2 \lambda_{\min} |\, x_t'\theta_0 \geq R) \\ &\geq \mathbb{P}(|x_t'(\theta'-\theta_0)| \geq \frac{1}{4}||\theta'-\theta_0||^2 \lambda_{\min} |\, x_t'\theta_0 \geq R) \geq \frac{\lambda_{\min}^2}{16 \mathbb{E}\left(||x_t||^4 | x_t'\theta_0 \geq R\right)}.\end{aligned}$$

□

Proof of Theorem 9.

Proof. Denote the LAD estimate as $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$. We are going to use results from the proof of Theorem 7. The difference is that now we will apply the martingale central limit theorem instead of the law of large numbers. Define the filtration $\mathcal{F}_t = \{y_t, z_t, y_{t-1}, z_{t-1}, \dots\}$. Then

$$\xi_t := \begin{pmatrix} \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) - \mathbb{E}\left(\operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) \,|\mathcal{F}_t\right) \\ y_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) - \mathbb{E}\left(y_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) \,|\mathcal{F}_t\right) \\ z_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) - \mathbb{E}\left(z_t \operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t - \Delta\gamma z_t) \,|\mathcal{F}_t\right) \end{pmatrix}$$

is a martingale difference with respect to filtration $\mathcal{F}_t$.

Let us calculate the corresponding asymptotic covariance matrix.

First, observe that

$$\mathbb{E}\left(\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)=1-2\mathbb{P}_v\left(v_{t+1}<\Delta\alpha+\Delta\beta y_t+\Delta\gamma z_t\right).$$

When $\Delta\alpha+\Delta\beta y_t+\Delta\gamma z_t\approx 0$, the probability that v_{t+1} is smaller than $\Delta\alpha+\Delta\beta y_t+\Delta\gamma z_t$ equals to $1/2$, as $\operatorname{med}(v)=0$. Thus, $\mathbb{E}\left(\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\approx 0$ for $\Delta\alpha+\Delta\beta y_t+\Delta\gamma z_t\approx 0$.

Second, note that $\operatorname{sgn}^2\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)\equiv 1$, so that

- $\mathbb{E}\left(y_t\operatorname{sgn}^2\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\equiv y_t$,
- $\mathbb{E}\left(z_t\operatorname{sgn}^2\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\equiv z_t$,
- $\mathbb{E}\left(y_t z_t\operatorname{sgn}^2\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\equiv y_t z_t$,
- $\mathbb{V}\left(\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)=1-0=1$,
- $\mathbb{V}\left(y_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)=y_t^2(1-0)=y_t^2$,
- $\mathbb{V}\left(z_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)=z_t^2(1-0)=z_t^2$.

Under stationarity $\frac{1}{T}\sum_t y_t\to\mathbb{E}y$, $\frac{1}{T}\sum_t z_t\to\mathbb{E}z$, $\frac{1}{T}\sum_t y_t^2\to\mathbb{E}y^2$, $\frac{1}{T}\sum_t y_t z_t\to\mathbb{E}yz$, $\frac{1}{T}\sum_t z_t^2\to\mathbb{E}z^2$. We want to apply the martingale CLT. Thus, we need to check the Lindeberg condition. Because

$$\sum_{t=1}^{T}\mathbb{V}\left(\frac{1}{\sqrt{T}}\xi_t|\mathcal{F}_t\right)=\begin{pmatrix}1 & \frac{1}{T}\sum_t y_t & \frac{1}{T}\sum_t z_t\\ \frac{1}{T}\sum_t y_t & \frac{1}{T}\sum_t y_t^2 & \frac{1}{T}\sum_t y_t z_t\\ \frac{1}{T}\sum_t z_t & \frac{1}{T}\sum_t y_t z_t & \frac{1}{T}\sum_t z_t^2\end{pmatrix},$$

the variance matrix is $O(1)$. So we need to check that for any numbers λ_1, λ_2, λ_3 and $\varepsilon>0$

$$\sum_{t=1}^{T}\mathbb{E}\left(\frac{1}{T}(\lambda_1\xi_{1t}+\lambda_2\xi_{2t}+\lambda_3\xi_{3t})^2\mathbf{1}\left(\frac{1}{\sqrt{T}}|\lambda_1\xi_{1t}+\lambda_2\xi_{2t}+\lambda_3\xi_{3t}|>\varepsilon\right)\right)\xrightarrow[T\to\infty]{}0. \tag{42}$$

Because y_t and z_t are stationary, Eq. (42) reduces to

$$\mathbb{E}(\lambda_1\xi_{1t}+\lambda_2\xi_{2t}+\lambda_3\xi_{3t})^2\mathbf{1}\left(\frac{1}{\sqrt{T}}|\lambda_1\xi_{1t}+\lambda_2\xi_{2t}+\lambda_3\xi_{3t}|>\varepsilon\right)\xrightarrow[T\to\infty]{}0,$$

which is true as $(1,y_t,z_t)$ and, thus, $\xi_t=(\xi_{1t},\xi_{2t},\xi_{3t})'$ and any linear combination of ξ_t's coordinates, have finite second moments.

So by the martingale CLT,

$$\frac{1}{\sqrt{T}}\begin{pmatrix} \sum_t \left[\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)-\mathbb{E}\left(\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\right] \\ \sum_t \left[y_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)-\mathbb{E}\left(y_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\right] \\ \sum_t \left[z_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)-\mathbb{E}\left(z_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\right] \end{pmatrix} \xrightarrow[T\to\infty]{d} \mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1 & \mathbb{E}y & \mathbb{E}z\\ \mathbb{E}y & \mathbb{E}y^2 & \mathbb{E}yz \\ \mathbb{E}z & \mathbb{E}yz & \mathbb{E}z^2\end{pmatrix}\right). \tag{43}$$

By Eq. (43), first order conditions (32) can be rewritten as

$$\begin{aligned}&\mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1 & \mathbb{E}y & \mathbb{E}z\\ \mathbb{E}y & \mathbb{E}y^2 & \mathbb{E}yz \\ \mathbb{E}z & \mathbb{E}yz & \mathbb{E}z^2\end{pmatrix}\right)\\ &+\frac{1}{\sqrt{T}}\begin{pmatrix}\sum_t \mathbb{E}\left(\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\\ \sum_t \mathbb{E}\left(y_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\\ \sum_t \mathbb{E}\left(z_t\operatorname{sgn}\left(v_{t+1}-\Delta\alpha-\Delta\beta y_t-\Delta\gamma z_t\right)|\mathcal{F}_t\right)\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}.\end{aligned} \tag{44}$$

Using approximations (36), (38), and (39) for the second term in Eq. (44) and taking the limit as $T\to\infty$ to approximate sums with expectations, we get

$$\begin{aligned}&\mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1 & \mathbb{E}y & \mathbb{E}z\\ \mathbb{E}y & \mathbb{E}y^2 & \mathbb{E}yz \\ \mathbb{E}z & \mathbb{E}yz & \mathbb{E}z^2\end{pmatrix}\right)\\ &-2f_u(0)\sqrt{T}\begin{pmatrix}\Delta\alpha+\Delta\beta\mathbb{E}y+\Delta\gamma\mathbb{E}z\\ \Delta\alpha\mathbb{E}y+\Delta\beta\mathbb{E}y^2+\Delta\gamma\mathbb{E}yz\\ \Delta\alpha\mathbb{E}z+\Delta\beta\mathbb{E}yz+\Delta\gamma\mathbb{E}z^2\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}.\end{aligned}$$

Thus,

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

□

Proof of Theorem 10.

Proof. Let us first linearize expectations of the first order conditions in spirit of the proof of Theorem 7.

$$\sum_{t=1}^{T} \mathrm{sgn}\left(y_t - a - by_{t-1} - cz_{t-1}\right) \mathbf{1}(a + by_{t-1} + cz_{t-1} > 0) = 0,$$
$$\sum_{t=1}^{T} y_{t-1}\mathrm{sgn}\left(y_t - a - by_{t-1} - cz_{t-1}\right) \mathbf{1}(a + by_{t-1} + cz_{t-1} > 0) = 0,$$
$$\sum_{t=1}^{T} z_{t-1}\mathrm{sgn}\left(y_t - a - by_{t-1} - cz_{t-1}\right) \mathbf{1}(a + by_{t-1} + cz_{t-1} > 0) = 0.$$

Thus, we will be left with expectations of the form

$$\begin{aligned} &\mathbb{E}_{y,z}(1 - 2F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z|y, z)) \\ &\cdot \mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + \Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1} > 0) = 0, \\ &\mathbb{E}_{y,z} y(1 - 2F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z|y, z)) \\ &\cdot \mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + \Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1} > 0) = 0, \\ &\mathbb{E}_{y,z} z(1 - 2F_{v|y,z}(\Delta\alpha + \Delta\beta y + \Delta\gamma z|y, z)) \\ &\cdot \mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + \Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1} > 0) = 0. \end{aligned} \tag{45}$$

Instead of Taylor expansion, we are going to use the direct formula: $f(x + \Delta x)g(x + \Delta x) = f(x)g(x) + f(x)(g(x + \Delta x) - g(x)) + g(x + \Delta x)(f(x + \Delta x) - f(x))$. Thus, the first expectation in Eq. (45) can be rewritten as

$$\begin{aligned} &\mathbb{E}_{y,z}(1 - 2F_{v|y,z}(0|y, z))\mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} > 0) \\ &- 2\mathbb{E}_{y,z} f_{0|y,z}(0|y, z)\mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + \Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1} > 0)(\Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1}) \\ &+ \mathbb{E}_{y,z}(1 - 2F_{v|y,z}0|y, z))\Big(\mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + \Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1} > 0) \\ &- \mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} > 0)\Big) \\ &\approx -2\mathbb{E}_{y,z} f_u(0)\mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} > 0)(\Delta\alpha + \Delta\beta y + \Delta\gamma z), \end{aligned} \tag{46}$$

where we used the fact that $F_{v|y,z}(0|y, z) = F_u(0) = 0.5$ and $f_{v|y,z}(0|y, z) = f_u(0)$ when $\alpha + \beta y_{t-1} + \gamma z_{t-1} > 0$, and continuity of $\mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} + \Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1} > 0)$ with respect to $\Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1} = 0$.

Similarly, the other two expectations in Eq. (45) can be rewritten as

$$-2\mathbb{E}_{y,z} f_u(0) y_{t-1}\mathbf{1}(\alpha + \beta y_{t-1} + \gamma z_{t-1} > 0)(\Delta\alpha + \Delta\beta y_{t-1} + \Delta\gamma z_{t-1})$$

and

$$-2\mathbb{E}_{y,z}f_u(0)z_{t-1}\mathbf{1}(\alpha+\beta y_{t-1}+\gamma z_{t-1}>0)(\Delta\alpha+\Delta\beta y_{t-1}+\Delta\gamma z_{t-1}).$$

The rest of the proof follows the lines of Theorem 9: we add and subtract conditional expectations from each term in the first order conditions. We then apply the martingale CLT for the difference and use the linearizations above (Eq. (46) and two others) for the expectations. The only difference is that everything will be multiplied by an indicator $\mathbf{1}(\alpha+\beta y+\gamma z+\Delta\alpha+\Delta\beta y+\Delta\Gamma z>0)$. Thus, we get

$$\sqrt{T}\begin{pmatrix}\Delta\alpha\\ \Delta\beta\\ \Delta\gamma\end{pmatrix}=\sqrt{T}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\\ \hat{\gamma}-\gamma\end{pmatrix}$$

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

□

Lemma C.2. *Consider the process $y_{t+1}=[\alpha+\beta y_t+u_t]_+$, $\beta>1$. For any $\beta'\in(1,\beta)$, almost surely exists T such that $y_{t+1}>\beta' y_t$ for all $t>T$.*

Proof. Fix $\varepsilon>0$. Denote $v_t=|\alpha|+|u_t|$. Write $y_{t+1}\ge\beta y_t-v_t$. Iterating, we get

$$y_{t+k}\ge\beta^k\left(y_t-\sum_{i=1}^{k}v_{t+i-1}\beta^{-i}\right)\tag{47}$$

Further, note $\sum_{i=1}^{\infty}v_{t+i-1}\beta^{-i}$ is a positive finite random variable, whose distribution does not depend on the choice of t. Choose large $M>2$ such that this random variable is less than M with probability greater than $1-\varepsilon$. By Theorem A.5 we already know that almost surely $y_t\xrightarrow[t\to\infty]{}\infty$. Thus, we can choose T such that $y_T>2M$ with probability greater than $1-\varepsilon$. Then with probability greater than $1-2\varepsilon$, we have by (47):

$$y_{T+k}\ge\beta^k(y_T/2),\qquad\text{for all } k=1,2,\ldots.\tag{48}$$

Let us call the event where (48) holds $\mathcal{A}_T$. We thus know that $\mathbb{P}(\mathcal{A}_T)\ge 1-2\varepsilon$ for large enough T.

Next, consider the events

$$\mathcal{B}_k=\{|v_{T+k}|>\beta^k\}.$$

Note that $\sum_k\mathbb{P}(\mathcal{B}_k)<\infty$, since v_t is a random variable, whose distribution does not depend on t and whose expectation exists. Therefore, there exists K such that for the event $\mathcal{C}_K=\{v_{T+k}\le\beta^k \text{ for all } k>K\}$, $\mathbb{P}(\mathcal{C}_K)\ge 1-\sum_{k=K}^{\infty}\mathbb{P}(\mathcal{B}_k)>1-\varepsilon$.

Now consider the event $\mathcal{D} = \mathcal{A}_T \cap \mathcal{C}_K$. We have $\mathbb{P}(\mathcal{D}) \geq 1 - (\mathbb{P}(\neg\mathcal{A}_T) + \mathbb{P}(\neg\mathcal{C}_K)) > 1 - 3\varepsilon$. On the other hand, on this event, for each $t > T + K$, we have

$$y_{t+1} \geq \beta y_t - v_t = \beta' y_t + (\beta - \beta') y_t \left(1 - \frac{v_t}{y_t}\right) \tag{49}$$

Since $y_t \geq \frac{y_T}{2}\beta^{t-T} > M\beta^{t-T} > 2\beta^{t-T}$ and $v_t \leq \beta^{t-T}$, the last term in (49) is positive and we conclude that $y_{t+1} \geq \beta' y_t$, as desired.

Since $\varepsilon > 0$ was arbitrary, we conclude that with probability 1 for all large enough t, $y_{t+1} \geq \beta' y_t$, as desired. □

Proof of Theorem 11.

Proof. Denote the LAD estimate as $(\hat{\alpha}, \hat{\beta})$. The estimate is the solution to minimization problem

$$\min_{a,b} \sum_t |y_{t+1} - [a + by_t]_+| \,. \tag{50}$$

The corresponding first order conditions are

$$\sum_t \operatorname{sgn}(y_{t+1} - a - by_t)\, \mathbf{1}(a + by_t > 0) = 0, \tag{51}$$

$$\sum_t y_t \operatorname{sgn}(y_{t+1} - a - by_t)\, \mathbf{1}(a + by_t > 0) = 0. \tag{52}$$

Note that formally the summations may not equal to zero, as they involve discrete increments. Thus, we need to find point, where the summations switch signs from minus to plus.

The proof differs depending on the behavior of y_t ($\beta > 1$ or $\beta = 1$, $\alpha > 0$ or $\beta = 1$, $\alpha = 0$).

Let us analyze Eq. (51). Define $\xi_t = \operatorname{sgn}(y_{t+1} - a - by_t)\, \mathbf{1}(a + by_t > 0)$. Then Eq. (51) can be rewritten as

$$\sum_t \mathbb{E}(\xi_t | y_t) + \sum_t (\xi_t - \mathbb{E}(\xi_t | y_t)) \,. \tag{53}$$

The second term, $\sum_t (\xi_t - \mathbb{E}(\xi_t|y_t))$ is of order $O(\sqrt{T})$. This follows from the fact that

$$\mathbb{E} \sum_t (\xi_t - \mathbb{E}(\xi_t|y_t)) = 0$$

and

$$\mathbb{E}\left(\sum_t (\xi_t - \mathbb{E}(\xi_t|y_t))\right)^2 = \sum_t \mathbb{E}\left(\xi_t - \mathbb{E}(\xi_t|y_t)\right)^2 < const \cdot T,$$

as each $\xi_t \in \{-1, 0, 1\}$.

Define $\Delta\alpha = \hat{\alpha} - \alpha$, $\Delta\beta = \hat{\beta} - \beta$, $v_{t+1} = \max(u_{t+1}, -\alpha - \beta y_t)$. Then

$$\begin{aligned}\mathbb{E}(\xi_t|y_t) &= \mathbb{E}\left(\operatorname{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t)\mathbf{1}(\Delta\alpha + \Delta\beta y_t + \alpha + \beta y_t > 0)|y_t\right) \\ &= (1 - 2F_v(\Delta\alpha + \Delta\beta y_t))\mathbf{1}(\Delta\alpha + \Delta\beta y_t + \alpha + \beta y_t > 0).\end{aligned}$$

We want to linearize $\mathbb{E}(\xi_t|y_t)$. When $\beta > 1$ or $\alpha > 0, \beta = 1$ we know from Theorem A.5, that $y_t \xrightarrow{a.s.} \infty$. Thus, for T' large enough, starting from $t > T'$ we get $\alpha + \beta y_t \gg 0$, so that the indicator does not bind.

- Case I: $\beta > 1$.

Assume that $\Delta\beta y_t \approx 0$, so that we can linearize $\mathbb{E}(\xi_t|y_t)$ around $\Delta\alpha + \Delta\beta y_t = 0$. Because y_t goes to infinity, we can not assume $\Delta\beta y_t \approx 0$ for all t. However, we can assume that $\Delta\beta y_t \approx 0$ for $t < T - \sqrt{T}$. At the end of the proof we will find solution to first order conditions, which indeed satisfies this assumption, and has $\Delta\beta y_T = O(1)$ (so that $\Delta\beta y_{T-\sqrt{T}} = o(1)$) and $\Delta\alpha = o(1)$.

Thus, for $t \in (T', T - \sqrt{T})$,

$$\begin{aligned}(1 - 2F_v(\Delta\alpha + \Delta\beta y_t))\mathbf{1}(\Delta\alpha + \Delta\beta y_t + \alpha + \beta y_t > 0) &= (1 - 2F_v(\Delta\alpha + \Delta\beta y_t)) \\ &\approx -2f_u(0)(\Delta\alpha + \Delta\beta y_t),\end{aligned}$$

so that

$$\sum_t \mathbb{E}(\xi_t|y_t) \approx -2Tf_u(0)\Delta\alpha - 2f_u(0)\Delta\beta \sum_t y_t + O(\sqrt{T}),$$

where we again used the fact that $\xi_t \in \{-, 1, 0, 1\}$ to bound terms with $t \notin (T', T - \sqrt{T})$.

Therefore, Eq. (51) can be rewritten as

$$-2Tf_u(0)\Delta\alpha - 2f_u(0)\Delta\beta \sum_{t=T'}^{T-\sqrt{T}} y_t + O(\sqrt{T}) = 0. \tag{54}$$

Let us now analyze the second first order condition, Eq. (52).

Suppose that $\beta > 1$. Then terms corresponding to $t \approx T$ dominate the summation, as they have the largest y_{t+1}'s, and the indicator is no longer binding. By Lemma C.2 we know that almost surely there exists T_0 such that $y_{t+1} > \beta' y_t$ for some $\beta' \in (1, \beta)$ and any $t \geq T_0$. Thus, $y_{T-k} \leq \beta'^{-k} y_T$. Simultaneously we assume $y_{T_0} > 1$ to avoid pathologies.

Choose an additional integer τ to be fixed later and split the sum in Eq. (52) into three:

$$\sum_{t=1}^{T_0} y_t \mathbf{1}(\hat\alpha + \hat\beta y_t > 0)\mathrm{sgn}\big(u_t - \Delta\alpha - \Delta\beta\big) + \sum_{t=T_0+1}^{T-\tau} y_t \mathrm{sgn}\big(u_t - \Delta\alpha - \Delta\beta y_t\big) + \sum_{t=T-\tau+1}^{T} y_t \mathrm{sgn}\big(u_t - \Delta\alpha - \Delta\beta\big) \tag{55}$$

The first sum is bounded by a (random) number M as $T \to \infty$. Now choose (random, independent from T) τ such that $(\beta')^\tau > \frac{2}{1-1/\beta'}$ and $(\beta')^\tau > 2M$. We claim that in this case whenever all the signs in the third sum are positive, (55) is positive, and whenever all the signs in the third sum are negative, (55) is negative. Indeed, the first sum is bounded by M, and the very last term with $t = T$ is at least twice larger due to our choice of τ and the bound $y_{T_0} > 1$ that we started from. Since y_t grows faster than geometric series with denominator β' for $t > T_0$, the second sum can be bounded from above by $y_{T-\tau}$ multiplied by geometric series with denominator $1/\beta'$. Hence, by our choice of τ and inequality $y_T > (\beta')^\tau y_{T-\tau}$, the last term with $t = T$ is again at least twice as large as the sum.

The conclusion is that the value of $\Delta\beta$ lies between the maximum value of b which makes all $\mathrm{sgn}\big(u_t - \Delta\alpha - b y_t\big)$ positive for $t = T - \tau + 1, \dots, T$, and the minimum value of b which makes all $\mathrm{sgn}\big(u_t - \Delta\alpha - b y_t\big)$ negative for $t = T - \tau + 1, \dots, T$. Thus, looking at the points, where $\mathrm{sgn}\big(u_t - \Delta\alpha - b y_t\big)$ changes, we obtain

$$\min_{t=T-\tau-1,\dots,T} \frac{u_t - \Delta\alpha}{y_t} \le \Delta\beta \le \max_{t=T-\tau-1,\dots,T} \frac{u_t - \Delta\alpha}{y_t}. \tag{56}$$

Expressing $\Delta\beta$ via $\Delta\alpha$ using (54) and plugging into (56), we get

$$\frac{1}{T}\left(\sum_{t=1}^{T-\sqrt{T}} y_t\right) \min_{t=T-\tau-1,\dots,T} \frac{u_t - \Delta\alpha}{y_t} \le O\left(T^{-\frac{1}{2}}\right) - \Delta\alpha \le \frac{1}{T}\left(\sum_{t=1}^{T-\sqrt{T}} y_t\right) \max_{t=T-\tau-1,\dots,T} \frac{u_t - \Delta\alpha}{y_t}. \tag{57}$$

The last inequality implies that $\Delta\alpha \to 0$ almost surely as $T \to \infty$, since $\frac{\sum_{t=T'}^{T-\sqrt{T}} y_t}{y_{t'}}$ stays bounded as $T \to \infty$ for all $t' = T - \tau + 1, \dots, T$ (the numerator is at most $y_{T-\sqrt{T}}$ multiplied by a geometric series with denominator $1/\beta'$ and the denominator is at least $y_{T-\sqrt{T}} \beta'^{\sqrt{T}-\tau}$, so that their ratio is bounded by $1/(1 - \beta'^{-1})$). In fact, we see that the speed of decay is $\frac{1}{T}$. Since $\Delta\alpha \to 0$, it stays bounded, and therefore, (56) implies that $\Delta\beta \to 0$ as fast as $1/y_T$, i.e. exponentially fast.

Finally, the solution to the first order conditions is a global minimum, not a local one, because y_t grows exponentially fast in t ($y_{t+1} > \beta' y_t$). Thus, for all large t, $\max(a + by_t, 0) = a + by_t$. This implies that most of the terms in (50) are convex functions of a and b (if the parameters are taken in a compact set), and therefore, the solution to the minimization problem can be found (up to a small error) as a point satisfying the first order conditions.

- Case II: $\beta = 1$, $\alpha > 0$.

In this case we do not need to consider separately observations with $t > T - \sqrt{T}$, and we will find solution to first order conditions, which has $\Delta\beta y_T = o(1)$ and $\Delta\alpha = o(1)$. Thus, for $t > T'$,

$$\begin{aligned}(1 - 2F_v(\Delta\alpha + \Delta\beta y_t))\mathbf{1}(\Delta\alpha + \Delta\beta y_t + \alpha + \beta y_t > 0) &= (1 - 2F_v(\Delta\alpha + \Delta\beta y_t))\\ &\approx -2f_u(0)(\Delta\alpha + \Delta\beta y_t),\end{aligned}$$

and Eq. (51) can be rewritten as

$$-2Tf_u(0)\Delta\alpha - 2f_u(0)\Delta\beta \sum_{t=T'}^{T} y_t + O(\sqrt{T}) = 0. \tag{58}$$

Let us now analyze the second first order condition, Eq. (52). Define

$$\eta_t = y_t \mathrm{sgn}(y_{t+1} - (\hat{\alpha} + \hat{\beta} y_t)).$$

Then for $t > T'$ the sum in Eq. (52) can be rewritten as

$$\sum_t \mathbb{E}(\eta_t | y_t) + \sum_t (\eta_t - \mathbb{E}(\eta_t|y_t)),$$

where

$$\mathbb{E} \sum_t (\eta_t - \mathbb{E}(\eta_t|y_t)) = 0$$

and

$$\begin{aligned}\mathbb{E}\left(\sum_t (\eta_t - \mathbb{E}(\eta_t|y_t))\right)^2 &= \mathbb{E}\sum_t (\eta_t - \mathbb{E}(\eta_t|y_t))^2 = \mathbb{E}\sum_t \left(\eta_t^2 - (\mathbb{E}(\eta_t|y_t))^2\right)\\ &= \mathbb{E}\sum_t y_t^2 \left(1 - \left(\mathbb{E}(\mathrm{sgn}(y_{t+1} - (\hat{\alpha} + \hat{\beta} y_t))|y_t)\right)^2\right) \le \mathbb{E}\sum_t y_t^2,\end{aligned}$$

because $\mathrm{sgn} \in [-1, 1]$.

Random variable y_t grows linearly in t (i.e. $y_T = O(T)$). To see this, note that $y_t \geq \alpha + y_{t-1} + u_t \geq \alpha t + y_0 + \sum_{s=1}^{t} u_s$ and $y_t \leq \alpha + y_{t-1} + |u_t| \leq \alpha t + y_0 + \sum_{s=1}^{t} |u_s|$, so that

$$\alpha \xleftarrow[t\to\infty]{} \alpha + y_0/t + \frac{1}{t}\sum_{s=1}^{t} u_s \leq y_t/t \leq \alpha + y_0/t + \frac{1}{t}\sum_{s=1}^{t} |u_s| \xrightarrow[t\to\infty]{} \alpha + \mathbb{E}|u_t|.$$

Thus, $\mathbb{E}\sum_t y_t^2$ is of order $\sum t^2 = \frac{T(T+1)(2T+1)}{6}$, so that $\mathbb{E}\sum_t y_t^2 = O(T^3)$ and $\sum_t(\eta_t - \mathbb{E}(\eta_t|y_t)) = O(T^{3/2})$.

Define $v_{t+1} = \max(u_{t+1}, -\alpha - \beta y_t)$. We are left with analyzing

$$\begin{aligned}\sum_t \mathbb{E}(\eta_t|y_t) &= \sum_t y_t \mathbb{E}(\text{sgn}(y_{t+1} - (\hat{\alpha} + \hat{\beta} y_t))|y_t) = \sum_t y_t \mathbb{E}(\text{sgn}(v_{t+1} - \Delta\alpha - \Delta\beta y_t)|y_t) \\ &= \sum_t y_t(1 - 2F_{v_{t+1}|y_t}(-\Delta\alpha - \Delta\beta y_t)).\end{aligned}$$

Taylor expanding $1 - 2F_{v_{t+1}|y_t}(-\Delta\alpha - \Delta\beta y_t)$ around $\Delta\alpha + \Delta\beta y_t = 0$ and using the fact that $F_{v|y}(0) = F_u(0) = 0.5$, $f_{v|y}(0) = f_u(u)$ as for $t > T'$ $\alpha + \beta y_t > 0$, we get

$$\sum_t \mathbb{E}(\eta_t|y_t) = -2f_u(0)\sum_t y_t(\Delta\alpha + \Delta\beta y_t + o(\Delta\alpha + \Delta\beta y_t)),$$

so that Eq. (52) can be rewritten as

$$\Delta\alpha \sum_{t=T'}^{T} y_t + \Delta\beta \sum_{t=T'}^{T} y_t^2 = O(T^{3/2}). \tag{59}$$

Solving Eq. (58) and (59), and using the fact that y_t grows linearly in T so that $\sum_t y_t = O(\sum_t t) = O(T^2)$, we get

$$\Delta\alpha = O(T^{-0.5}),\ \Delta\beta = O(T^{-1.5}).$$

Thus, $\Delta\alpha \xrightarrow[T\to\infty]{} 0$, $\Delta\beta \xrightarrow[T\to\infty]{} 0$, and the LAD estimator is consistent. Similarly to the Case I, the solution to the first order conditions is a global minimum, not a local one, because y_t grows linearly in t. Thus, for all large t, $\max(a + by_t, 0) = a + by_t$. This implies that most of the terms in (50) are convex functions of a and b (if the parameters are taken in a compact set), and therefore, the solution to the minimization problem can be found (up to a small error) as a point satisfying the first order conditions.

- Case III: $\beta = 1$, $\alpha = 0$.

Let us calculate the order of terms, where indicator binds. By Theorem A.8, $\frac{1}{\sqrt{T}} y_{\lfloor Ts \rfloor} \to \sigma|W(s)|$ for any $s \in (0, 1]$, where $\sigma = \mathbb{E}u_t^2$ and W is a standard Brownian motion. Thus, the indicator can be rewritten as

$$\mathbf{1}\left(b\frac{y_{\lfloor Ts\rfloor}}{\sqrt{T}} > -\frac{a}{\sqrt{T}}\right).$$

As T goes to infinity, $\frac{a}{\sqrt{T}}$ goes to zero and $b\frac{y_{\lfloor Ts\rfloor}}{\sqrt{T}}$ goes to $b\sigma|W(s)|$. The (random) time Brownian motion spends inside interval $[-\varepsilon,\varepsilon]$ goes to zero almost surely as $\varepsilon\to 0$ (see Section 3.6 in Karatzas and Shreve (2012)). Thus, for $b>0$ the time $b\sigma|W(s)|$ is smaller than $-\frac{a}{\sqrt{T}}$ goes to zero as $T\to\infty$, i.e. it is $o(1)$. Therefore, $\sharp\{t:\ b\frac{y_{\lfloor Ts\rfloor}}{\sqrt{T}}\le -\frac{a}{\sqrt{T}}\}=o(T)$.

Note also that $b\le 0$ can not solve the minimization problem (50). When $b<0$, $b\sigma|W(s)|\le 0$ and the indicator starts to bind all the time, so that we get $\sum\limits_t |y_{t+1}| = O(T^{3/2})$, as $\frac{1}{T\sqrt{T}}\sum\limits_t|y_{t+1}|\to\sigma\int\limits_0^1|W(s)|ds$. Yet, $\sum\limits_t|y_{t+1}-y_t|=\sum\limits_t|u_t|=O(T)$. When $b=0$, we again get $\sum\limits_t|y_{t+1}-[a]_+|=O(T^{3/2})$ as $\frac{1}{T\sqrt{T}}\sum\limits_t|y_{t+1}-[a]_+|=\frac{1}{T}\sum\limits_t|y_{t+1}/\sqrt{T}-[a]_+/\sqrt{T}|\to\sigma\int\limits_0^1|W(s)|ds$.

Thus, we can linearize Eq. (51) to get

$$-2Tf_u(0)\Delta\alpha - 2f_u(0)\Delta\beta\sum_t y_t + o(T) + O(\sqrt{T}) = 0. \tag{60}$$

To analyze the second first order condition, Eq. (52), we proceed as in Case II. The only difference is that now we use Theorem A.8 to calculate the order of different sums involving y_t. That is, $\frac{1}{T^2}\mathbb{E}\sum\limits_t y_t^2 \xrightarrow[T\to\infty]{} \mathbb{E}\sigma^2\int_0^1 W^2(s)ds$ implying $\mathbb{E}\sum\limits_t y_t^2 = O(T^2)$ and $\frac{1}{T^{3/2}}\mathbb{E}\sum\limits_t y_t \xrightarrow[T\to\infty]{} \mathbb{E}\sigma\int_0^1|W(s)|ds$ implying $\mathbb{E}\sum\limits_t y_t = O(T^{3/2})$. Thus, the second first order condition, Eq. (52) can be rewritten as

$$\Delta\alpha O(T^{3/2}) + \Delta\beta O(T^2) + o(T^{3/2}) = O(T). \tag{61}$$

Solving Eq. (60) and (61), we get

$$\Delta\alpha = o(1),\ \Delta\beta = o(T^{-0.5}).$$

Thus, $\Delta\alpha \xrightarrow[T\to\infty]{} 0$, $\Delta\beta \xrightarrow[T\to\infty]{} 0$, and the LAD estimator is consistent. The solution to the first order conditions is a global minimum, not a local one because the indicator in the maximization problem binds in $o(T)$ terms. This implies that the dominant terms in (50) are convex functions of a and b (if the parameters are taken in a compact set), and therefore, the solution to the minimization problem can be found (up to a small error) as a point satisfying the first order conditions. □

Proof of Theorem 13.

Proof. Define $\hat{\beta} = \frac{\sum\limits_{T_{1M}} y_{t-1}y_t}{\sum\limits_{T_{1M}} y_{t-1}^2}$ and $\hat{\gamma} = \frac{\sum\limits_{T_{1M}} z_{t-1}y_t}{\sum\limits_{T_{1M}} z_{t-1}^2}$. Let us show that $\hat{\beta} \xrightarrow[(M,T)_{seq}\to\infty]{\mathbb{P}} \beta$. The proof for $\hat{\gamma}$ is the same.

$$\begin{aligned}\hat{\beta} = \frac{\sum\limits_{T_{1M}} y_{t-1}y_t}{\sum\limits_{T_{1M}} y_{t-1}^2} &= \beta + \frac{\sum\limits_{T_{1M}} (\alpha + \gamma z_{t-1} + u_t)y_{t-1}}{\sum\limits_{T_{1M}} y_{t-1}^2} \\ &\xrightarrow[T\to\infty]{\mathbb{P}} \beta + \frac{\alpha\mathbb{E}(y_{t-1}|T_{1M}) + \gamma\mathbb{E}(y_{t-1}z_{t-1}|T_{1M}) + \mathbb{E}(u_t y_{t-1}|T_{1M})}{\mathbb{E}(y_{t-1}^2|T_{1M})}.\end{aligned}$$

First note that u_t and y_{t-1} are independent and $\mathbb{E}(u_t|T_{1M}) \xrightarrow[M\to\infty]{} 0$. Then $\frac{\mathbb{E}(y_{t-1}|T_{1M})}{\mathbb{E}(y_{t-1}^2|T_{1M})} \leq \frac{\mathbb{E}(y_{t-1}|T_{1M})}{\mathbb{E}(My_{t-1}|T_{1M})} = \frac{1}{M} \xrightarrow[M\to\infty]{} 0$. Finally, by Cauchy–Schwarz inequality, $\mathbb{E}(y_{t-1}z_{t-1}|T_{1M}) \leq \sqrt{\mathbb{E}(y_{t-1}^2|T_{1M})\mathbb{E}(z_{t-1}^2|T_{1M})}$ so that $\frac{\mathbb{E}(y_{t-1}z_{t-1}|T_{1M})}{\mathbb{E}(y_{t-1}^2|T_{1M})} \leq \sqrt{\frac{\mathbb{E}(z_{t-1}^2|T_{1M})}{\mathbb{E}(y_{t-1}^2|T_{1M})}} \leq \sqrt{\frac{M^2/h^2(M)}{M^2}} = \frac{1}{h(M)} \xrightarrow[M\to\infty]{} 0$.

Thus, $\hat{\beta} \xrightarrow[(M,T)_{seq}\to\infty]{\mathbb{P}} \beta$.

Finally, notice that both under T_{1M} and T_{2M},

$$\alpha + u_t = y_t - \beta y_{t-1} - \gamma z_{t-1},$$

so that

$$\begin{aligned}\frac{1}{|T_{1M}| + |T_{2M}|} \sum_{T_{1M}\cup T_{2M}} (y_t - \beta y_{t-1} - \gamma z_{t-1}) &= \alpha + \frac{1}{|T_{1M}| + |T_{2M}|} \sum_{T_{1M}\cup T_{2M}} u_t \\ &\xrightarrow[T\to\infty]{\mathbb{P}} \alpha + \mathbb{E}(u|T_{1M} \cup T_{2M})\end{aligned}$$

and

$$\alpha + \mathbb{E}(u|T_{1M} \cup T_{2M}) \xrightarrow[M\to\infty]{} \alpha + \mathbb{E}u = \alpha.$$

□

Proof of Theorem 14.

Proof. Conditional on T_M, the OLS estimate equals to

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix}+\begin{pmatrix}\sum\limits_{T_M}1 & \sum\limits_{T_M}y_{t-1} & \sum\limits_{T_M}z_{t-1}\\ \sum\limits_{T_M}y_{t-1} & \sum\limits_{T_M}y_{t-1}^2 & \sum\limits_{T_M}y_{t-1}z_{t-1}\\ \sum\limits_{T_M}z_{t-1} & \sum\limits_{T_M}y_{t-1}z_{t-1} & \sum\limits_{T_M}z_{t-1}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum\limits_{T_M}u_t\\ \sum\limits_{T_M}y_{t-1}u_t\\ \sum\limits_{T_M}z_{t-1}u_t\end{pmatrix}$$

$$\xrightarrow[T\to\infty]{\mathbb{P}}\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix}+\begin{pmatrix}1 & \mathbb{E}(y_{t-1}|T_M) & \mathbb{E}(z_{t-1}|T_M)\\ \mathbb{E}(y_{t-1}|T_M) & \mathbb{E}(y_{t-1}^2|T_M) & \mathbb{E}(y_{t-1}z_{t-1}|T_M)\\ \mathbb{E}(z_{t-1}|T_M) & \mathbb{E}(y_{t-1}z_{t-1}|T_M) & \mathbb{E}(z_{t-1}^2|T_M)\end{pmatrix}^{-1}\begin{pmatrix}\mathbb{E}u_t\\ \mathbb{E}(y_{t-1}u_t|T_M)\\ \mathbb{E}(z_{t-1}u_t|T_M)\end{pmatrix}.$$

Let us rewrite the second term. The goal is to show that it converges to zero as $M\to\infty$.

$$\begin{pmatrix}1 & \mathbb{E}(y_{t-1}|T_M) & \mathbb{E}(z_{t-1}|T_M)\\ \mathbb{E}(y_{t-1}|T_M) & \mathbb{E}(y_{t-1}^2|T_M) & \mathbb{E}(y_{t-1}z_{t-1}|T_M)\\ \mathbb{E}(z_{t-1}|T_M) & \mathbb{E}(y_{t-1}z_{t-1}|T_M) & \mathbb{E}(z_{t-1}^2|T_M)\end{pmatrix}^{-1}\begin{pmatrix}\mathbb{E}u_t\\ \mathbb{E}(y_{t-1}u_t|T_M)\\ \mathbb{E}(z_{t-1}u_t|T_M)\end{pmatrix}$$

$$=\begin{pmatrix}1 & \mathbb{E}(y|T_M) & \mathbb{E}(z|T_M)\\ \mathbb{E}(y|T_M) & \mathbb{E}(y^2|T_M) & \mathbb{E}(yz|T_M)\\ \mathbb{E}(z|T_M) & \mathbb{E}(yz|T_M) & \mathbb{E}(z^2|T_M)\end{pmatrix}^{-1}$$

$$\cdot\left(\mathbb{E}(u|T_M)\begin{pmatrix}1\\ \mathbb{E}(y|T_M)\\ \mathbb{E}(z|T_M)\end{pmatrix}+\begin{pmatrix}0\\ \mathbb{E}(yu|T_M)-\mathbb{E}(y|T_M)\mathbb{E}(u|T_M)\\ \mathbb{E}(zu|T_M)-\mathbb{E}(z|T_M)\mathbb{E}(u|T_M)\end{pmatrix}\right)$$

$$=\mathbb{E}(u|T_M)\begin{pmatrix}1\\ 0\\ 0\end{pmatrix}+\begin{pmatrix}1 & \mathbb{E}(y|T_M) & \mathbb{E}(z|T_M)\\ \mathbb{E}(y|T_M) & \mathbb{E}(y^2|T_M) & \mathbb{E}(yz|T_M)\\ \mathbb{E}(z|T_M) & \mathbb{E}(yz|T_M) & \mathbb{E}(z^2|T_M)\end{pmatrix}^{-1}\begin{pmatrix}0\\ Cov(y,u|T_M)\\ Cov(z,u|T_M)\end{pmatrix}.$$

Because $\mathbb{E}(u|T_M)\to\mathbb{E}u=0$ as $M\to\infty$, the first term converges to zero. By assumption, the second term also converges to zero. (Note that as $M\to\infty$ the correlation between u_t and y_{t-1} drops to zero, so that $Cov(y_{t-1},u_t|T_M)\to 0$. Similarly, $Cov(z_{t-1},u_t|T_M)\to 0$. However, the behavior of the inverse matrix per se is unclear.)

Thus, sequential limit of the OLS estimate based on T_M equals the true values of the parameters. $\square$

Proof of Theorem 15.

Proof. Because $g'_y(x) \geq c_1 x^{d_1}$, for any $x \geq M$ $g'_y(x) \geq c_1 M^{d_1}$ and

$$\begin{aligned}\mathbb{P}(y_{t-1} > M) &= \int_M^\infty f_y(x)dx = \int_M^\infty e^{-g_y(x)}dx = e^{-g_y(M)} \int_M^\infty e^{-\int_M^x g'_y(w)dw} dx \\ &\leq e^{-g_y(M)} \int_M^\infty e^{-c_1 M^{d_1}(x-M)}dx \leq e^{-g_y(M)} \frac{1}{c_1 M^{d_1}} \leq e^{-g_y(M)},\end{aligned} \tag{62}$$

where the last inequality holds for M large enough.

We want to show that conditional variance of y_{t-1} is polynomial in M^{-1}. To do this, let us show that if a variance of a random variable X is bounded from below by $C > 0$ on some interval $[a, b]$, then $\mathbb{V}X \geq \frac{C}{8}(b-a)^3$:

$$\mathbb{V}X = \int_{\mathbb{R}} (x - \mathbb{E}X)^2 f_X(x)dx \geq C \int_a^b (x - \mathbb{E}X)^2 dx \geq C \int_0^{\frac{b-a}{2}} x^2 dx = \frac{C}{24}(b-a)^3, \tag{63}$$

where the last inequality holds because if $a + \frac{b-a}{2} \geq \mathbb{E}X$ then $(x - \mathbb{E}X)^2 \geq \left(x - a - \frac{b-a}{2}\right)^2$ for $x \in [a + (b-a)/2, b]$ and if $a + \frac{b-a}{2} \leq \mathbb{E}X$ then $(x - \mathbb{E}X)^2 \geq (x-a)^2$ for $x \in [a, a + (b-a)/2]$.

Consider the interval $\Delta = [M, M + (c_2 M^{d_2})^{-1}]$. Because $f_y(x) = e^{-g_y(x)}$ and $g'_y > 0$, the density of y is decreasing for $x \in \Delta$ for M large enough. Thus, for $x \in \Delta$,

$$\begin{aligned} f_y(x) &\geq f_y(M + (c_2 M^{d_2})^{-1}) = \exp(-g_y(M + (c_2 M^{d_2})^{-1})) \\ &\geq \exp(-g_y(M) - g'_y(M + (c_2 M^{d_2})^{-1})(c_2 M^{d_2})^{-1}) \\ &\geq \exp(-g_y(M)) \exp(-c_2(M + (c_2 M^{d_2})^{-1})^{d_2}(c_2 M^{d_2})^{-1}) \\ &= \exp(-g_y(M)) \exp(-(1 + (c_2 M^{d_2+1})^{-1})^{d_2}) \geq \exp(-g_y(M)) e^{-2^{d_2}}. \end{aligned} \tag{64}$$

Therefore, combining Eq. (62) and Eq. (64), for $x \in \Delta$,

$$f_{y_{t-1}|T_M}(x) = f_y(x)/\mathbb{P}(y_{t-1} > M) \geq e^{-g_y(M)} e^{-2^{d_2}} / e^{-g_y(M)} = e^{-2^{d_2}}.$$

Using the bound from Eq. (63), we get

$$\mathbb{V}(y_{t-1}|T_M) = \int \left(x - \mathbb{E}(y_{t-1}|T_M)\right)^2 f_y(x)dx \geq \frac{1}{24 e^{2^{d_2}}}(c_2 M^{d_2})^{-3}. \tag{65}$$

Let us show that the conditional expectation of y_{t-1} does not grow faster than linearly in M.

(66)
$$\begin{aligned}\mathbb{E}(y_{t-1}|T_M) &= \int\limits_M^\infty x\frac{f_y(x)}{\mathbb{P}(y_{t-1}>M)}dx = \frac{\int\limits_M^\infty xe^{-g_y(x)}dx}{\int\limits_M^\infty e^{-g_y(x)}dx} \leq \frac{2\int\limits_M^{2M} xe^{-g_y(x)}dx}{\int\limits_M^\infty e^{-g_y(x)}dx}\\ &\leq 4M\frac{\int\limits_M^{2M} e^{-g_y(x)}dx}{\int\limits_M^\infty e^{-g_y(x)}dx} \leq 4M,\end{aligned}$$

where the first inequality comes from the fact that $xe^{-g_y(x)}$ is decreasing exponentially, so that for M large enough $\int\limits_M^{2M} xe^{-g_y(x)}dx > \int\limits_{2M}^\infty xe^{-g_y(x)}dx$.

We are left with analyzing conditional covariance between y_{t-1} and u_t.

(67)
$$\begin{aligned}&Cov(y_{t-1},u_t|T_M) = \mathbb{E}\left(y_{t-1}\mathbb{E}\left(u_t - \mathbb{E}(u_t|T_M)|y_{t-1}\right)|T_M\right)\\ &= \int\limits_M^\infty x\int\limits_{-\alpha-\beta x}^\infty v\frac{f_u(v)}{\mathbb{P}(u_t>-\alpha-\beta x)}dv\frac{f_y(x)}{\mathbb{P}(y_{t-1}>M)}dx - \mathbb{E}(y_{t-1}|T_M)\mathbb{E}(u_t|T_M).\end{aligned}$$

First, note that, for $x \geq M$,

(68)
$$\begin{aligned}\mathbb{P}(u_t>-\alpha-\beta x) \geq \mathbb{P}(u_t>-\alpha-\beta M) &= 1-\int\limits_{-\infty}^{-\alpha-\beta M} f_u(v)dv = 1-\int\limits_{-\infty}^{-\alpha-\beta M} e^{-g_u(v)}dv\\ &\geq 1-\int\limits_{-\infty}^{-\alpha-\beta M} e^{-c_3|v|^{d_3}}dv \xrightarrow[M\to\infty]{} 1,\end{aligned}$$

so that $\mathbb{P}(u_t>-\alpha-\beta x) \geq 0.5$ for M large enough.

Second, because $\mathbb{E}u = 0$,

(69)
$$\begin{aligned}\int\limits_{-\alpha-\beta x}^\infty vf_u(v)dv &= -\int\limits_{-\infty}^{-\alpha-\beta x} vf_u(v)dv = \int\limits_{-\infty}^{-\alpha-\beta x}(-v)e^{-g_u(v)}dv \leq \int\limits_{-\infty}^{-\alpha-\beta x}(-v)e^{-c_3|v|^{d_3}}dv\\ &\leq 2(\alpha+\beta x)e^{-c_3(\alpha+\beta x)^{d_3}},\end{aligned}$$

where the last inequality holds for M large enough as the integrand decreases exponentially.

Third, using Eq. (69),

$$\begin{aligned}
\mathbb{E}(u_t|T_M) &= \int_M^\infty \int_{-\alpha-\beta x}^\infty u \frac{f_u(v)}{\mathbb{P}(u_t > -\alpha - \beta x)} dv \frac{f_y(x)}{\mathbb{P}(y_{t-1} > M)} dx \\
&\leq \int_M^\infty 2(\alpha + \beta x) e^{-c_3(\alpha+\beta x)^{d_3}} \frac{f_y(x)}{\mathbb{P}(u_t > -\alpha - \beta x)\mathbb{P}(y_{t-1} > M)} dx \\
&\leq \int_M^\infty 4(\alpha + \beta x) e^{-c_3(\alpha+\beta x)^{d_3}} \frac{f_y(x)}{\mathbb{P}(y_{t-1} > M)} dx \\
&= 4\mathbb{E}\left((\alpha + \beta y_{t-1}) e^{-c_3(\alpha+\beta y_{t-1})^{d_3}} |T_M\right) \leq 4(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}},
\end{aligned} \tag{70}$$

because the function under expectation is decreasing in y for M large enough.

Plugging Eq. (66), (69), and (70) into Eq. (67), we get

$$\begin{aligned}
&|Cov(y_{t-1}, u_t|T_M)| \\
&\leq 4 \int_M^\infty x(\alpha + \beta x) e^{-c_3(\alpha+\beta x)^{d_3}} \frac{f_y(x) dx}{\mathbb{P}(y_{t-1} > M)} + 16M(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}} \\
&\leq 4M(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}} + 16M(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}}.
\end{aligned} \tag{71}$$

Combining Eq. (65) and (71), we get

$$\begin{aligned}
&\frac{|Cov(y_{t-1}, u_t|T_M)|}{\mathbb{V}(y_{t-1}|T_M)} \\
&\leq 24e^{2^{d_2}} \frac{4M(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}} + 16M(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}}}{(c_2 M^{d_2})^{-3}} \xrightarrow[M\to\infty]{} 0.
\end{aligned}$$

Combining Eq. (65), (66) and (71), we get

$$\begin{aligned}
&\frac{\mathbb{E}(y_{t-1}|T_M)|Cov(y_{t-1}, u_t|T_M)|}{\mathbb{V}(y_{t-1}|T_M)} \\
&\leq 96e^{2^{d_2}} M \frac{4M(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}} + 16M(\alpha + \beta M) e^{-c_3(\alpha+\beta M)^{d_3}}}{(c_2 M^{d_2})^{-3}} \xrightarrow[M\to\infty]{} 0.
\end{aligned}$$

So that $\frac{1}{\mathbb{V}(y_{t-1}|T_M)} \begin{pmatrix} -\mathbb{E}(y_{t-1}|T_M) Cov(y_{t-1}, u_t|T_M) \\ Cov(y_{t-1}, u_t|T_M) \end{pmatrix} \xrightarrow[M\to\infty]{} 0.$ □

Proof of Theorem 16.

Proof. Define $\theta = (\alpha, \beta, \gamma, \sigma)$ and assume that θ_0 is the true value of θ. MLE estimate $\hat{\theta}$ maximizes sample log-likelihood, Q_n. The sample and population log-likelihoods are

$$\begin{aligned} Q_n(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\Big[\log f_y(y_t|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t > 0) + \log F_y(0|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t = 0)\Big] \\ &= \frac{1}{T}\sum_{t=1}^{T}\Big[\log f_u(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1}|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t > 0) \\ &+ \log F_u(-\alpha - \beta y_{t-1} - \gamma z_{t-1}|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t = 0)\Big] \end{aligned} \tag{72}$$

and

$$\begin{aligned} Q(\theta) &= \mathbb{E}\Big[\log f_y(y_t|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t > 0) + \log F_y(0|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t = 0)\Big] \\ &= \mathbb{E}\Big[\log f_u(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1}|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t > 0) \\ &+ \log F_u(-\alpha - \beta y_{t-1} - \gamma z_{t-1}|y_{t-1}, z_{t-1}, \theta)\mathbf{1}(y_t = 0)\Big], \end{aligned} \tag{73}$$

where expectation is taken with respect to y_t, y_{t-1}, z_{t-1}.

Let us first show that θ_0 uniquely minimizes Q.

$$\begin{aligned} Q(\theta) - Q(\theta_0) &= \mathbb{E}(\log f_y(y_t|y_{t-1}, z_{t-1}, \theta) - \log f_y(y_t|y_{t-1}, z_{t-1}, \theta_0))\mathbf{1}(y_t > 0) \\ &+ \mathbb{E}(\log F_y(0|y_{t-1}, z_{t-1}, \theta) - \log F_y(0|y_{t-1}, z_{t-1}, \theta_0))\mathbf{1}(y_t = 0). \end{aligned} \tag{74}$$

First note, that

$$\begin{aligned} &\mathbb{E}\mathbf{1}(y_t = 0)\log F_y(0|y_{t-1}, z_{t-1}, \theta) = \mathbb{E}\log F_y(0|y_{t-1}, z_{t-1}, \theta)\left(\mathbb{E}\left(\mathbf{1}(y_t = 0)|y_{t-1}, z_{t-1}\right)\right) \\ &= \mathbb{E}\log F_y(0|y_{t-1}, z_{t-1}, \theta)\mathbb{P}_y(0|y_{t-1}, z_{t-1}) = \mathbb{E}\log F_y(0|y_{t-1}, z_{t-1}, \theta)F_y(0|y_{t-1}, z_{t-1}, \theta_0). \end{aligned}$$

Then

$$\begin{aligned} &\mathbb{E}(\log F_y(0|y_{t-1}, z_{t-1}, \theta) - \log F_y(0|y_{t-1}, z_{t-1}, \theta_0))\mathbf{1}(y_t = 0) \\ &= \mathbb{E}\log\left(\frac{F_y(0|y_{t-1}, z_{t-1}, \theta)}{F_y(0|y_{t-1}, z_{t-1}, \theta_0)}\right)F_y(0|y_{t-1}, z_{t-1}) \\ &\leq \mathbb{E}\left(\frac{F_y(0|y_{t-1}, z_{t-1}, \theta)}{F_y(0|y_{t-1}, z_{t-1}, \theta_0)} - 1\right)F_y(0|y_{t-1}, z_{t-1}) \\ &= \mathbb{E}\left(F_y(0|y_{t-1}, z_{t-1}, \theta) - F_y(0|y_{t-1}, z_{t-1}, \theta_0)\right). \end{aligned} \tag{75}$$

Similarly,

$$
\begin{aligned}
&\mathbb{E}(\log f_y(y_t|y_{t-1}, z_{t-1}, \theta) - \log f_y(y_t|y_{t-1}, z_{t-1}, \theta_0))\mathbf{1}(y_t > 0) \\
&= \mathbb{E} \log \left(\frac{f_y(y_t|y_{t-1}, z_{t-1}, \theta)}{f_y(y_t|y_{t-1}, z_{t-1}, \theta_0)} \right) \mathbf{1}(y_t > 0) \le \mathbb{E} \left(\frac{f_y(y_t|y_{t-1}, z_{t-1}, \theta)}{f_y(y_t|y_{t-1}, z_{t-1}, \theta_0)} - 1 \right) \mathbf{1}(y_t > 0) \\
&= \mathbb{E} \left(\mathbb{E} \left(\left(\frac{f_y(y_t|y_{t-1}, z_{t-1}, \theta)}{f_y(y_t|y_{t-1}, z_{t-1}, \theta_0)} - 1 \right) \mathbf{1}(y_t > 0) | y_{t-1}, z_{t-1} \right) \right) \\
&= \mathbb{E} \int \left(f_y(y_t|y_{t-1}, z_{t-1}, \theta) - f_y(y_t|y_{t-1}, z_{t-1}, \theta_0) \right) \mathbf{1}(y_t > 0) dy_t \\
&= (1 - \mathbb{E}F_y(0|y_{t-1}, z_{t-1}, \theta)) - (1 - \mathbb{E}F_y(0|y_{t-1}, z_{t-1}, \theta_0)) \\
&= \mathbb{E} \left(F_y(0|y_{t-1}, z_{t-1}, \theta_0) - F_y(0|y_{t-1}, z_{t-1}, \theta) \right)
\end{aligned} \tag{76}
$$

Plugging Eq. (75) and (76) into Eq. (74), we get

$$
\begin{aligned}
Q(\theta) - Q(\theta_0) \le \mathbb{E} &\left(F_y(0|y_{t-1}, z_{t-1}, \theta) - F_y(0|y_{t-1}, z_{t-1}, \theta_0) \right) \\
&+ \mathbb{E} \left(F_y(0|y_{t-1}, z_{t-1}, \theta_0) - F_y(0|y_{t-1}, z_{t-1}, \theta) \right) = 0.
\end{aligned}
$$

Thus, θ_0 minimizes Q. Moreover, equality holds only when $\mathbb{P}\left(f_y(y_t|y_{t-1}, z_{t-1}, \theta) = f_y(y_t|y_{t-1}, z_{t-1}, \theta_0) \right) = 1$, which can not happen for gaussian errors with density $f_y(y_t|y_{t-1}, z_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1})^2 \right)$.

To apply the theorem for extremum estimators, we need to reduce the domain of θ to a compact space. That is, we need to show that when some of the parameters go to infinity, Q_n goes to minus infinity and, thus, such values can not be solutions to $\max Q_n$. Here we are going to use the fact that $f_u(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{x^2/2\sigma^2}$. Let us plug the density into Eq. (72):

$$
\begin{aligned}
Q_n = &\frac{1}{T} \sum_{t=1}^{T} \left(-0.5 \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1})^2 \right) \mathbf{1}(y_t > 0) \\
&+ \frac{1}{T} \sum_{t=1}^{T} \left(-0.5 \log(2\pi\sigma^2) + \log \int_{-\infty}^{-\alpha - \beta y_{t-1} - \gamma z_{t-1}} e^{-u^2/2\sigma^2} du \right) \mathbf{1}(y_t = 0)
\end{aligned} \tag{77}
$$

If σ goes to infinity, then $-0.5 \log(2\pi\sigma^2) \to -\infty$, while other terms remain non-positive: $\int_{-\infty}^{A} e^{-u^2/2\sigma^2} du \le \sqrt{2\pi\sigma}$. Thus, $Q_n \to -\infty$ when $\sigma \to \infty$ independently of the values of other parameters, and we can restrict σ to a bounded interval.

After we know that σ is bounded, we can guarantee that the second summation is bounded by zero from above for any values of α, β, γ: $\left(-0.5 \log(2\pi\sigma^2) + \log \int_{-\infty}^{-\alpha - \beta y_{t-1} - \gamma z_{t-1}} e^{-u^2/2\sigma^2} du \right) \le -0.5 \log(2\pi\sigma^2) + 0.5 \log(2\pi\sigma^2) = 0$. When $|\alpha|$ goes to infinity or $|\beta| \to \infty$ or $|\gamma| \to \infty$, we have $(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1})^2 \to \infty$.

Note that as y and z are random with correlation below one, parameters can not compensate each other. Thus, $Q_n \to -\infty$ as $|\alpha| \to \infty$ or $\beta \to \infty$ or $|\gamma| \to \infty$. Therefore, those parameters can also be restricted to bounded intervals. Thus, we are left with compact set.

Plugging the density of u into Eq. (73), we get

$$\begin{aligned} Q(\theta) = \mathbb{E}\left(-0.5\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1})^2\right)\mathbf{1}(y_t > 0) \\ + \mathbb{E}\left(-0.5\log(2\pi\sigma^2) + \log \int_{-\infty}^{-\alpha-\beta y_{t-1}-\gamma z_{t-1}} e^{-u^2/2\sigma^2} du\right)\mathbf{1}(y_t = 0). \end{aligned} \tag{78}$$

Function under expectation in Eq. (78),

$$\begin{aligned} g(y_t, y_{t-1}, z_{t-1}, \theta) := \left(-0.5\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_t - \alpha - \beta y_{t-1} - \gamma z_{t-1})^2\right)\mathbf{1}(y_t > 0) \\ + \left(-0.5\log(2\pi\sigma^2) + \log \int_{-\infty}^{-\alpha-\beta y_{t-1}-\gamma z_{t-1}} e^{-u^2/2\sigma^2} du\right)\mathbf{1}(y_t = 0), \end{aligned}$$

is continuous at every θ with probability 1 and, because parameters are restricted to a compact set, $\mathbb{E}\sup_\theta |g(y_t, y_{t-1}, z_{t-1}, \theta)|$ is finite.

Finally, we can apply Proposition 7.3 (Consistency of M-estimators with compact parameter space) from Hayashi (2000). Our model satisfies all the conditions of the proposition. Thus, the MLE estimate $\hat{\theta}$ is consistent. □

(Anna Bykhovskaya) Yale University
E-mail address: anna.bykhovskaya@yale.edu