

GENERALIZED AUTOREGRESSIVE MULTIVARIATE MODELS: FROM BINARY TO POISSON

ANNA BYKHOVSKAYA AND NOUR MEDDAHI

ABSTRACT. This paper presents a framework for binary autoregressive time series in which each observation is a Bernoulli variable whose success probability evolves with past outcomes and probabilities, in the spirit of GARCH-type dynamics, accommodating nonlinearities, network interactions, and cross-sectional dependence in the multivariate case. Existence and uniqueness of a stationary solution is established via a coupling argument tailored to the discontinuities inherent in binary data. A key theoretical result, further supported by our empirical illustration on S&P 100 data, shows that, under a rare-events scaling, aggregates of such binary processes converge to a Poisson autoregression, providing a micro-foundation for this widely used count model. Maximum likelihood estimation is proposed and illustrated empirically.

Keywords: Binary data; Interactions; GARCH; Aggregation; Poisson autoregression

1. INTRODUCTION

Binary time series taking values in $\{0, 1\}$ are central to modeling dynamic economic and financial phenomena in which outcomes represent discrete events or decisions rather than continuous quantities. In economics, they arise naturally in dynamic binary choice models, where agents repeatedly decide whether to take actions such as purchasing, investing, or entering a market (Rust, 1987; Heckman, 1981). In finance, binary indicators capture rare but economically important events, such as extreme return days and bear markets (Nyberg, 2013), as well as corporate defaults and credit downgrades (Duffie *et al.*, 2009). Macroeconomic forecasting provides another key application, where the variable of interest is binary, such as the onset of a recession (Kauppi and Saikkonen, 2008).

Beyond individual processes, multivariate binary time series offer a natural framework for modeling interdependent or networked decisions, where outcomes for one unit depend on others. Such models are useful for analyzing financial contagion, correlated defaults, and evolving networks of economic linkages. Allowing the network structure to vary over time further enriches the analysis by capturing changing patterns of influence, information diffusion, or systemic risk propagation (cf., dynamic network formation of Graham 2016).

Aggregating binary outcomes across a large panel produces a count of successes at each date. We show that, under a rare-events scaling, these cross-sectional interactions provide a natural microfoundation for Poisson autoregression, a leading model for count time series.

We develop a general framework for multivariate binary time series in which each observation follows a Bernoulli distribution with a time-varying success probability that depends on both past outcomes and past success probabilities, in the spirit of the GARCH-type dynamics of Bollerslev (1986). Unlike the models of Kedem and Fokianos (2005) and De Jong and Woutersen (2011), our formulation allows current success probabilities to depend on latent past success probabilities, rather than only on observed outcomes. This additional layer accommodates richer persistence and feedback dynamics. Under general and potentially nonlinear dynamics for the success probabilities, we establish the existence of a stationary solution to our model. Under a contraction condition on the probability-update function, we further establish its uniqueness.

Our model fits into several important classes. When specialized to a linear form, our model corresponds to the generalized autoregressive score (GAS) framework of Creal *et al.* (2013). However, most of the existing results for GAS models (e.g., Blasques *et al.*, 2014, 2022) do not apply in our setting. The reason is that Bernoulli variables, expressible as indicators that a uniform random shock falls below a success probability, generate discontinuities that violate the smoothness and Lipschitz assumptions crucial for the GAS theory. Our model is also a member of the class of positive-valued time series with exponential-family conditional distributions studied, in the univariate case, by Davis and Liu (2016); Aknouche and Francq (2021) and, in the multivariate case with one lag of dependent variables, by Lee *et al.* (2023). Our analysis departs from those contributions by focusing specifically on cross-sectional interactions and their effect on the aggregate number of successes, a question those papers do not address.

The stochastic properties of related binary models have been studied by Moysiadis and Fokianos (2014) and Fokianos and Moysiadis (2017), and some of our structural results, particularly the existence and uniqueness of a stationary solution, are analogous in spirit to results obtained there, reflecting the robustness of the underlying probabilistic argument. The present paper departs from that work in two important respects. First, we consider an N -dimensional panel with cross-sectional interactions. Second, and more substantially, we target the rare-events asymptotic regime, which requires success probabilities to be close to zero and motivates working directly with $[0, 1]^N$ rather than reparametrizing the probabilities via an inverse distribution function. The cited papers apply a smooth monotone transformation, such as the logistic or normal CDF, that maps probabilities to an unbounded domain. Under these transformations, small probabilities are mapped to values near $-\infty$, making

analysis in this region problematic. Moreover, when combined additively, these highly negative transformed values decrease further, behaving more like products of probabilities than sums, which complicates interpretation.

Our central result provides a micro-level foundation for the integer-valued GARCH (IN-GARCH) model (Rydberg and Shephard, 2000; Streett, 2000; Ferland *et al.*, 2006), also known as Poisson autoregression (Fokianos *et al.*, 2009), one of the most widely used models for count time series. Under a rare-events scaling in which individual success probabilities vanish as the cross-sectional dimension N grows, we show that the aggregate number of successes converges, in finite-dimensional distributions, to a Poisson autoregression.

The practical motivation is clear: many economically important phenomena, such as defaults in large credit portfolios, the exercise of binary options across many positions, or episodes of heightened systemic risk, are individually rare, yet their cross-sectional aggregate incidence is of primary interest. This contrasts with the original motivation of Rydberg and Shephard (2000), which arose from discretizing time in a Poisson process.

Our setting yields a richer family of Poisson autoregressions than the parametric specifications considered in Fokianos *et al.* (2009). In particular, our recursion for the Poisson intensity admits genuinely nonlinear functions of both the aggregate count and the limiting intensity. We also show that the result extends to network interaction structures: when the network is doubly stochastic and sufficiently dense, heterogeneous interaction weights wash out asymptotically, and the same Poisson autoregressive limit obtains. This illustrates a broader phenomenon in which network topology becomes irrelevant at the aggregate level once connectivity is sufficiently rich.

Finally, we propose maximum likelihood estimation (MLE) and establish consistency and asymptotic normality under a boundedness condition on the success probabilities, which rules out the rare-events regime. In that regime the Poisson MLE applied to the aggregate count is the natural substitute, and under structural assumptions such as homogeneity, the aggregation theory supplies an explicit mapping from the Poisson parameters back to those of the underlying binary system.

We illustrate the framework using daily returns on S&P 100 constituents over 2012–2025, defining tail events as idiosyncratic returns falling below the 5th percentile. The binary and Poisson parameter estimates are close, consistent with the aggregation theory, and the filtered Poisson intensity tracks the sum of individual success probabilities closely in sample. In out-of-sample forecasting, the heterogeneous binary model outperforms the Poisson-calibrated specification, indicating that the granularity of the binary likelihood remains empirically meaningful.

The remainder of the paper is organized as follows. Section 2 introduces the model, provides motivating examples, and establishes the existence and uniqueness of a stationary solution. Section 3 analyzes the model's aggregate behavior, focusing on the total number of successes within a given period. Section 4 derives the consistency and asymptotic normality of the maximum likelihood estimator. Section 5 extends the model by allowing for additional covariates. Section 6 applies the proposed framework to S&P 100 data. Finally, Section 7 concludes. All proofs appear in the appendices.

2. SETTING

Consider an N -dimensional binary vector $\mathbf{y}_t = (y_{1,t}, \dots, y_{N,t})^\top$, where the superscript $\top$ denotes transposition. Here time is discrete and we deal with two cases: either two-sided $t \in \mathbb{Z}$ or one-sided $t \in \{t_0, t_0 + 1, t_0 + 2, \dots\}$ for some $t_0 \in \mathbb{Z}$. Let $\mathcal{F}_t$ be a filtration of σ -fields, so that $\mathbf{y}_t$ is $\mathcal{F}_t$ -measurable. The components of $\mathbf{y}_t$ are assumed to be independent conditional on $\mathcal{F}_{t-1}$.¹ Given $\mathcal{F}_{t-1}$, each component $y_{i,t}$, $i = 1, \dots, N$, is assumed to be a Bernoulli random variable with success probability $p_{i,t}$, measurable with respect to $\mathcal{F}_{t-1}$,

$$(1) \quad y_{i,t} \mid \mathcal{F}_{t-1} \sim B(p_{i,t}) \text{ or equivalently } y_{i,t} = I(u_{i,t} \leq p_{i,t}), \quad u_{i,t} \mid \mathcal{F}_{t-1} \sim i.i.d. \ U[0, 1].$$

The vector of success probabilities $\mathbf{p}_t = (p_{1,t}, \dots, p_{N,t})^\top$ may depend on its own past values $\mathbf{p}_\tau$, $\tau = t - 1, \dots, t - s$, as well as on past realizations $\mathbf{y}_\tau$, $\tau = t - 1, \dots, t - q$. Formally,

$$(2) \quad \mathbf{p}_t = \mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}),$$

where s and q are finite integers controlling the memory of the process, and $\mathbf{g}(\cdot)$ is a continuous mapping from $[0, 1]^{Ns} \times \{0, 1\}^{Nq}$ to $[0, 1]^N$. When the time is one-sided, $t \in \{t_0, \dots\}$, the evolution (2) should be supplemented with initial values of $\mathbf{p}_t$ and the equation is only valid once all the indices in (2) are at least t_0 .

The general form of $\mathbf{g}(\cdot)$ allows for nonlinearities and interactions, whereby $y_{i,t}$ depends not only on its own past but also on the past values and success probabilities of all N variables. The dependence of $\mathbf{p}_t$ on both $\mathbf{y}_{t-1}$ and $\mathbf{p}_{t-1}$ mirrors the GARCH structure of Bollerslev (1986). We, therefore, refer to the dynamics in (1)–(2) as the *Generalized Autoregressive Binary Process* (GAB).

The Bernoulli distribution belongs to the one-parameter exponential family and, thus, the GAB model fits in the class of models considered in Davis and Liu (2016); Aknouche and Francq (2021); Lee *et al.* (2023).

¹Contemporaneous conditional dependence can be accommodated without affecting the stationarity results. The aggregation results, however, require an appropriate form of weak cross-sectional dependence. Such dependence can be introduced by replacing the i.i.d. $U[0, 1]$ innovations below with innovations that are conditionally marginally uniform on $[0, 1]$ but cross-sectionally dependent.

2.1. **Examples.** Let us introduce some useful examples of the function $\mathbf{g}(\cdot)$. We begin with a simpler setting where $N = 1$, so that $\mathbf{p}_t = p_{1,t} \equiv p_t$ and $\mathbf{y}_t = y_{1,t} \equiv y_t$.

Example 1: Linear GAB(1, 1)

$$(3) \quad p_t = \omega + \alpha y_{t-1} + \beta p_{t-1}, \quad \omega, \alpha, \beta \geq 0, \quad \omega + \alpha + \beta \leq 1.$$

The conditions on the parameters ensure that p_t stays in $[0, 1]$. This framework resembles a classical GARCH model, where instead of conditional variance we specify conditional success probability as a linear function of its own past and a past variable realization. Moreover, the model (3) is a special case of generalized autoregressive score (GAS) models of Creal *et al.* (2013), where, in the notations of (Creal *et al.*, 2013), $s_t = y_t - p_t$, $S_t = p_t(1 - p_t)$.

Example 2: Linear GAB(s, q)

$$(4) \quad p_t = \omega + \sum_{j=1}^q \alpha_j y_{t-j} + \sum_{j=1}^s \beta_j p_{t-j}, \quad \omega, \alpha_j, \beta_j \geq 0, \quad \omega + \sum_{j=1}^q \alpha_j + \sum_{j=1}^s \beta_j \leq 1.$$

This framework extends the previous one by allowing for multiple lags of both probabilities and outcomes.

Example 3: Logit GAB(1, 1). Consider a function from $[0, 1] \times \{0, 1\}$ to $[0, 1]$, defined for any parameter values ω, α, β ,

$$\mathbf{g}(p, y) = \left(1 + \exp(-\omega - \alpha y) \left(\frac{1-p}{p} \right)^\beta \right)^{-1}.$$

The above function does not require additional parameter restrictions to ensure that probabilities remain in $[0, 1]$. This function is continuous and maps $(0, 0)$ to 0 and $(1, 1)$ to 1 when $\beta > 0$, while for $\beta < 0$ it maps $(0, 0)$ to 1 and $(1, 1)$ to 0. In contrast to the logistic transformation of Moysiadis and Fokianos (2014); Fokianos and Moysiadis (2017), the above form still allows for small probabilities. In terms of our original variables, the evolution can be written as

$$(5) \quad p_t = \left(1 + \exp(-\omega - \alpha y_{t-1}) \left(\frac{1-p_{t-1}}{p_{t-1}} \right)^\beta \right)^{-1}.$$

Equivalently, one has

$$\ln \left(\frac{p_t}{1-p_t} \right) = \omega + \alpha y_{t-1} + \beta \ln \left(\frac{p_{t-1}}{1-p_{t-1}} \right)$$

or, for $x_t = \ln \left(\frac{p_t}{1-p_t} \right) \in \mathbb{R} \cup \{\pm\infty\}$, so that $p_t = (1 + \exp(-x_t))^{-1}$,

$$(6) \quad x_t = \omega + \alpha y_{t-1} + \beta x_{t-1}.$$

Eq. (6) corresponds to Moysiadis and Fokianos (2014); Fokianos and Moysiadis (2017), except that they restrict $x_t \in \mathbb{R}$, ruling out $p_t = 0$.

Example 4: Nonlinear GAB(1, 1)

$$p_t = \omega + \alpha y_{t-1} + f(p_{t-1}),$$

where $f(\cdot)$ is a continuous function such that $\forall p \in [0, 1]$,

$$-\omega - \min\{\alpha, 0\} \leq f(p) \leq 1 - \omega - \max\{\alpha, 0\}.$$

This framework allows for nonlinear dependence between p_t and p_{t-1} . Notice that, since y_t is binary, we do not need to apply any nonlinear function to it.

Figures 1a and 1b illustrate nonlinear GAB(1, 1) models with $\omega = 0.05$ and $\alpha = 0.4$, using $f(p) = 2p(p - 0.5)^2$ and $f(p) = (p + 0.7)(p - 5/6)^2$, respectively. The key difference lies in the magnitude of the nonlinear component. In Figure 1a, the nonlinear term remains small, especially when $p \approx 0$, resulting in oscillatory dynamics with prolonged low-probability periods punctuated by short-lived spikes following realizations of $y_t = 1$. Overall, the path of p_t stays close to the Markovian component $\omega + \alpha y_{t-1}$. In contrast, Figure 1b exhibits a much stronger nonlinear effect, leading to substantially different dynamics and pronounced deviations from the Markovian benchmark.

All of the above examples can be straightforwardly generalized to $N > 1$, where each time series i has its own $p_{i,t}$, whose evolution involves only quantities indexed by i and no quantities indexed by $j \neq i$. More interesting cases arise when the dynamics of $p_{i,t}$ depend on the outcomes or probabilities of other series. A few examples are provided below.

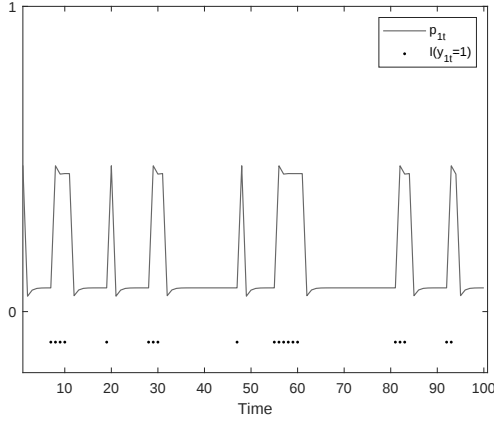
Example 5: Exchangeable GAB(1, 1). All N components share a common probability of success, driven by the cross-sectional average of past outcomes,

$$(7) \quad p_{i,t} \equiv p_t = \omega + \gamma \frac{1}{N} \sum_{j=1}^N y_{j,t-1} + \beta p_{t-1}, \quad \omega, \beta, \gamma \geq 0, \quad \omega + \beta + \gamma \leq 1.$$

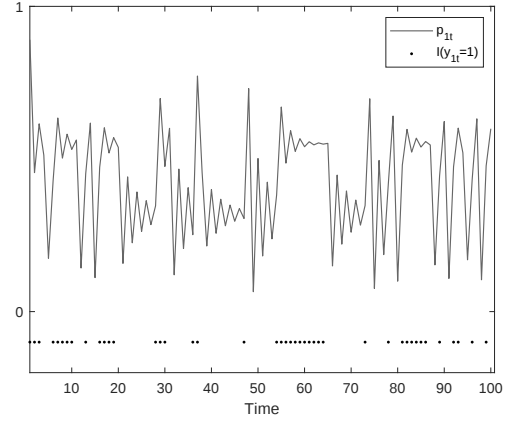
Under this exchangeable specification, the sum $\sum_{j=1}^N y_{j,t}$, conditional on $\mathcal{I}_{t-1}$, follows a binomial distribution with parameters N and p_t .

Example 6: Interactive GAB(1, 1). Each component has its own probability of success, driven by both its own past outcome and the cross-sectional average,

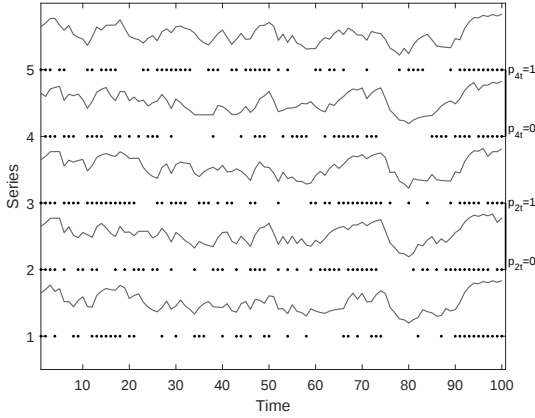
$$(8) \quad p_{i,t} = \omega_i + \alpha_i y_{i,t-1} + \gamma_i \frac{1}{N} \sum_{j=1}^N y_{j,t-1} + \beta_i p_{i,t-1}, \quad \omega_i, \alpha_i, \gamma_i, \beta_i \geq 0, \quad \omega_i + \alpha_i + \gamma_i + \beta_i \leq 1.$$



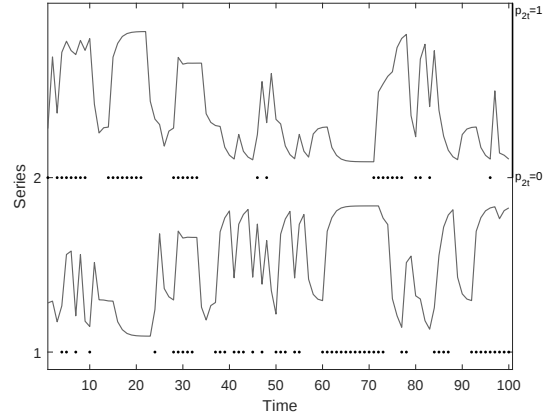
(A) Nonlinear single-series model: $p_{1,0} = 1/12$, $p_{1,t} = 0.05 + 0.4y_{1,t-1} + 2p_{1,t-1}(p_{1,t-1} - 0.5)^2$.



(B) Nonlinear single-series model: $p_{1,0} = 1/12$, $p_{1,t} = 0.05 + 0.4y_{1,t-1} + (p_{1,t-1} + 0.7)(p_{1,t-1} - 5/6)^2$.



(C) Interactive model (8) with $N = 5$, $p_{i,0} = 0.5$, and $\omega_i = 0.05$, $\alpha_i = 0.1$, $\beta_i = 0.6$, $\gamma_i = 0.2$ for all i .



(D) Two-series model with negative interaction: $N = 2$, $p_{1,0} = p_{2,0} = 3/7$, $p_{i,t} = 0.3 + 0.4y_{i,t-1} + 0.2p_{i,t-1} - 0.3p_{j,t-1}$, $i \neq j$.

FIGURE 1. Raster plots with probability trajectories. Black dots indicate success periods ($y_{i,t} = 1$). Continuous lines represent the corresponding success probabilities $p_{i,t}$.

When $\gamma_i = 0$ for all i , the model (8) reduces to N trajectories of GAB(1, 1) models (3) that are independent conditional on $\mathcal{I}_{t-1}$. When, for all i , $\omega_i \equiv \omega$, $\alpha_i \equiv 0$, $\beta_i \equiv \beta$, $\gamma_i \equiv \gamma$, the model (8) reduces to a single probability model (7).

Example 7: Network GAB(1, 1). Each component has its own probability of success, driven by a network-weighted average of past outcomes,

$$(9) \quad p_{i,t} = \omega_i + \alpha_i y_{i,t-1} + \gamma_i \sum_{j=1}^N W_{ij} y_{j,t-1} + \beta_i p_{i,t-1}, \quad \omega_i, \alpha_i, \gamma_i, \beta_i \geq 0, \quad \omega_i + \alpha_i + \gamma_i + \beta_i \leq 1,$$

where W is an $N \times N$ row-normalized adjacency matrix governing the relationships among individuals. Eq. (8) corresponds to the special case $W_{ij} = 1/N$ for all i, j , i.e., the complete graph.

All of the above multivariate examples can be readily extended to incorporate multiple lags of outcomes and probabilities, analogous to the transition from $\text{GAB}(1, 1)$ to $\text{GAB}(s, q)$. They can also be generalized to allow for nonlinearities that aggregate $p_{i,t}$ and/or $y_{i,t}$ across i , for instance, in the spirit of the peer-effect interaction terms of Bykhovskaya (2022), where a nonlinear function of neighbors' past outcomes enters the dynamics.

Figures 1c and 1d visualize multivariate GAB dynamics for $y_{i,t}$ and $p_{i,t}$: dots at row i indicate realizations of successes for series i , while the continuous trajectory above each row represents the corresponding success probability $p_{i,t}$. Figure 1c corresponds to the interactive model (8) with $N = 5$ and strictly positive coefficients γ_i . The figure reveals pronounced co-movement in the success probabilities induced by the common component $\frac{1}{N} \sum_{j=1}^N y_{j,t-1}$, while the series-specific term $\alpha_i y_{i,t-1}$ introduces heterogeneity across series.

Figure 1d illustrates, using two series, how the opposite effect can arise, with the success probabilities moving in opposite directions. In particular, when the success probability $p_{j,t}$ increases, the negative term $-0.3p_{j,t}$ becomes more pronounced,² which in turn reduces the success probability $p_{i,t+1}$ for $i \neq j$.

2.2. Stationarity. We next analyze the stationarity properties of the GAB model (1)–(2) in the asymptotic regime $T \rightarrow \infty$, with the number of time series N held fixed.

2.2.1. Existence.

Theorem 1. *For any continuous $\mathbf{g}(\cdot)$, there exists a probability space supporting $\mathcal{F}_t$ and strictly stationary sequence $(\mathbf{y}_t, \mathbf{p}_t)_{t \in \mathbb{Z}}$, such that (1)–(2) holds.*

Notice that no additional conditions are imposed, implying that all GAB models admit a stationary solution regardless of parameter values. The key features ensuring this property are the bounded support of all variables and the continuity of the probability function $\mathbf{g}$. Consequently, all examples from the preceding subsection, including the nonlinear Example 3 and the interactive Examples 6 and 7, possess stationary solutions. In the case of the logit GAB model (Example 3), note in particular that it is not necessary to impose the classical condition $|\beta| < 1$.

Since all variables are bounded by construction (outcomes in $\{0, 1\}$ and probabilities in $[0, 1]$), the process has finite moments of all orders. Hence, strict stationarity implies weak stationarity, leading to the following corollary.

²The coefficients are chosen to ensure that $p_{i,t} \in [0, 1]$.

Corollary 2. *The multivariate binary time series model (1)–(2), $t \in \mathbb{Z}$, has a weakly stationary solution.*

The law of iterated expectations and (1) imply that $\mathbb{E}y_{i,t} = \mathbb{E}p_{i,t}$ and $\mathbb{E}_{t-1}y_{i,t} = p_{i,t}$, where the subscript $t-1$ denotes conditioning on $\mathcal{F}_{t-1}$. Assuming $\sum_{j=1}^q \alpha_j + \sum_{j=1}^s \beta_j < 1$, the linear GAB(s, q) yields

$$(10) \quad \mu := \mathbb{E}y_{i,t} = \mathbb{E}p_{i,t} = \frac{\omega}{1 - \sum_{j=1}^q \alpha_j - \sum_{j=1}^s \beta_j},$$

and, assuming $\gamma + \beta < 1$, its N -dimensional variation (7) yields

$$\mu := \mathbb{E}y_{i,t} = \mathbb{E}p_{i,t} = \frac{\omega}{1 - \gamma - \beta},$$

which matches the unconditional variance formula for classical GARCH(s, q) and GARCH(1, 1).

In the interactive model (8), assuming $\alpha_i + \beta_i < 1$ for all i and $\frac{1}{N} \sum_{i=1}^N \frac{\gamma_i}{1 - \alpha_i - \beta_i} < 1$,

$$\mu_i := \mathbb{E}y_{i,t} = \mathbb{E}p_{i,t} = \omega_i + \alpha_i \mu_i + \gamma_i \frac{1}{N} \sum_{j=1}^N \mu_j + \beta_i \mu_i = \frac{\omega_i}{1 - \alpha_i - \beta_i} + \frac{\gamma_i}{N(1 - \alpha_i - \beta_i)} \sum_{j=1}^N \mu_j,$$

so that

$$(11) \quad \sum_{i=1}^N \mu_i = \sum_{i=1}^N \frac{\omega_i}{1 - \alpha_i - \beta_i} + \frac{1}{N} \sum_{j=1}^N \mu_j \sum_{i=1}^N \frac{\gamma_i}{1 - \alpha_i - \beta_i} = \frac{\sum_{i=1}^N \frac{\omega_i}{1 - \alpha_i - \beta_i}}{1 - \frac{1}{N} \sum_{i=1}^N \frac{\gamma_i}{1 - \alpha_i - \beta_i}}$$

and

$$(12) \quad \mu_i = \frac{\omega_i}{1 - \alpha_i - \beta_i} + \frac{\gamma_i}{N(1 - \alpha_i - \beta_i)} \frac{\sum_{j=1}^N \frac{\omega_j}{1 - \alpha_j - \beta_j}}{1 - \frac{1}{N} \sum_{j=1}^N \frac{\gamma_j}{1 - \alpha_j - \beta_j}}.$$

In the network GAB(1, 1) model (9), a similar argument yields, in matrix form with $A := \text{diag}(\alpha_i + \beta_i)$, $\Gamma := \text{diag}(\gamma_i)$, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_N)^\top$, and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)^\top$,

$$(13) \quad (I - A - \Gamma W) \boldsymbol{\mu} = \boldsymbol{\omega}.$$

Since W is row-stochastic, the row sums of $A + \Gamma W$ equal $\alpha_i + \beta_i + \gamma_i \leq 1 - \omega_i$, so for $\omega_i > 0$ the matrix $I - A - \Gamma W$ is strictly row-diagonally dominant and hence invertible, giving $\boldsymbol{\mu} = (I - A - \Gamma W)^{-1} \boldsymbol{\omega}$. For general W this solution does not simplify further. The interactive model (8) corresponds to $W_{ij} = 1/N$, the complete graph, in which case the uniform structure allows $\sum_i \mu_i$ to be solved explicitly, yielding the closed-form expressions (11)–(12).

The linear GAB(s, q) model with $\sum_{j=1}^q \alpha_j + \sum_{j=1}^s \beta_j = 1$, and hence $\omega = 0$, admits multiple stationary solutions. In particular, both $y_{it} = p_{it} \equiv 0$ and $y_{it} = p_{it} \equiv 1$ constitute trivial stationary paths. This is in contrast to the IGARCH case, where the unit-root condition destroys weak stationarity; here, bounded support guarantees that stationarity is preserved even on the boundary, though at the cost of uniqueness.

2.2.2. Uniqueness.

Assumption I. Suppose that there exist non-negative coefficients $\{\alpha_{ij}^\tau\}$, $\{\beta_{ij}^\tau\}$, with $\beta_{ij}^\tau = 0$ for $\tau > s$ and $\alpha_{ij}^\tau = 0$ for $\tau > q$, such that the function $\mathbf{g}(\cdot) = (g_1(\cdot), \dots, g_N(\cdot))^\top$ satisfies the Lipschitz bound

$$(14) \quad \begin{aligned} & |g_i(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}) - g_i(\{\mathbf{p}'_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}'_\tau\}_{\tau=t-1, \dots, t-q})| \\ & \leq \sum_{j=1}^N \sum_{\tau=1}^s |p_{j,t-\tau} - p'_{j,t-\tau}| \beta_{ij}^\tau + \sum_{j=1}^N \sum_{\tau=1}^q |y_{j,t-\tau} - y'_{j,t-\tau}| \alpha_{ij}^\tau \quad \forall i. \end{aligned}$$

Let Φ be the associated $N \max(s, q) \times N \max(s, q)$ companion matrix, where the first N rows of Φ are given by

$$\left(\alpha_{i1}^1 + \beta_{i1}^1, \dots, \alpha_{iN}^1 + \beta_{iN}^1, \dots, \alpha_{i1}^{\max(s,q)} + \beta_{i1}^{\max(s,q)}, \dots, \alpha_{iN}^{\max(s,q)} + \beta_{iN}^{\max(s,q)} \right), \quad i = 1, \dots, N,$$

the bottom-left $N(\max(s, q) - 1) \times N(\max(s, q) - 1)$ block of Φ is an identity matrix, and the bottom-right $N(\max(s, q) - 1) \times N$ block is zero. Suppose that the spectral radius $\rho(\Phi) < 1$.

Assumption I requires that the Lipschitz coefficients of $\mathbf{g}(\cdot)$, when arranged into the companion matrix Φ , yield a stable system in the sense that the spectral radius $\rho(\Phi) < 1$. This is the direct analogue of the stationarity condition for a vector autoregression of order $\max(s, q)$. A simple sufficient condition (and equivalent for $N = 1$) is the row-sum bound $\sum_{j=1}^N \sum_{\tau=1}^s \beta_{ij}^\tau + \sum_{j=1}^N \sum_{\tau=1}^q \alpha_{ij}^\tau \leq K < 1$ for all i , which implies $\rho(\Phi) < 1$. For the network GAB(1, 1) model (9), the companion matrix reduces to the single $N \times N$ block $A + \Gamma W$, so Assumption I specializes to $\rho(A + \Gamma W) < 1$, which also implies invertibility of $I - A - \Gamma W$ and hence the existence of the closed-form mean (13).

Alternatively, the contraction assumption can be stated in an aggregate form:

Assumption II. There exists $0 \leq K < \frac{1}{s+q}$ such that

$$(15) \quad \begin{aligned} & \sum_{i=1}^N |g_i(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}) - g_i(\{\mathbf{p}'_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}'_\tau\}_{\tau=t-1, \dots, t-q})| \\ & \leq K \left(\sum_{i=1}^N \sum_{\tau=1}^s |p_{i,t-\tau} - p'_{i,t-\tau}| + \sum_{i=1}^N \sum_{\tau=1}^q |y_{i,t-\tau} - y'_{i,t-\tau}| \right). \end{aligned}$$

Neither Assumption I nor Assumption II implies the other. For example, when $N = 2$ and $s = q = 1$, let $g_1(\mathbf{p}, \mathbf{y}) = \frac{2}{3}p_1$ and $g_2(\mathbf{p}, \mathbf{y}) = 0$, then Assumption I holds, but Assumption II fails since $2/3 > 1/2$. Conversely, let $g_i(\mathbf{p}, \mathbf{y}) = f_i(p_1) + f_i(p_2) + f_i(y_1) + f_i(y_2)$, where $f_1(x) = \min\{0.2, 0.4x\}$, $f_2(x) = \max\{0, 0.4(x - 0.5)\}$. Then $g_i \in [0, 0.8]$ and each f_i is Lipschitz with constant 0.4. Moreover, the intervals on which f_1 and f_2 are nonconstant do not overlap, so bound (15) holds with $K = 0.4$ (and no smaller constant is possible). However, under Assumption I, we must have $\alpha_{ij} = \beta_{ij} = 0.4$, implying $\Phi = 0.8(1, 1)^\top(1, 1)$ and hence $\rho(\Phi) = 1.6$.

Theorem 3. *Suppose that $\mathbf{g}(\cdot)$ satisfies either Assumption I or II. The distribution of a strictly stationary sequence $(\mathbf{y}_t, \mathbf{p}_t)_{t \in \mathbb{Z}}$ solving (1)–(2) is unique. If the time is one-sided, $t \in \{t_0, \dots\}$, then the time-series forgets its initial condition geometrically fast and converges in distribution to the unique stationary solution as $t \rightarrow \infty$.*

More precisely, by forgetting the initial condition we mean that for any two realizations $(\mathbf{y}_t, \mathbf{p}_t)$ and $(\mathbf{y}'_t, \mathbf{p}'_t)$, $t \geq t_0$, of the process driven by the same innovation sequence $u_{i,t}$, but initialized from different values, we have

$$(16) \quad \mathbb{E} \sum_{i=1}^N |y_{i,t} - y'_{i,t}| = \mathbb{E} \sum_{i=1}^N |p_{i,t} - p'_{i,t}| \leq C \varrho^t, \quad t \geq t_0,$$

for some constants $C < \infty$ and $0 \leq \varrho < 1$.

While we could not locate Theorem 3 in the literature, related results have appeared in Moysiadis and Fokianos (2014); Davis and Liu (2016); Fokianos and Moysiadis (2017); Aknouche and Francq (2021); Lee *et al.* (2023). The nontrivial aspect in the binary setting arises from the discontinuous nature of $y_{i,t}$. A jump in $y_{i,t}$ from 0 to 1 can lead to markedly different values of future probabilities $\mathbf{p}_{t+1}$. In particular, given the information set at time $t - 1$, since $y_{i,t} = I(u_{i,t} \leq p_{i,t})$, there is a discontinuity at $u_{i,t} = p_{i,t}$, which induces discontinuous and non-Lipschitz behavior in $\mathbf{p}_{t+1}$. The key to overcoming this difficulty lies in a suitable coupling construction.

3. AGGREGATION

In this section, we study the behavior of the total number of successes at a given time, $X_t(N) := y_{1,t} + \dots + y_{N,t}$. While the previous section analyzed the limit of the system as $T \rightarrow \infty$ with N fixed, we now investigate the complementary asymptotic regime in which $N \rightarrow \infty$. To initialize the dynamics, we consider an initial condition $\{\mathbf{p}_{1-\max\{s,q\}}, \dots, \mathbf{p}_{-1}, \mathbf{p}_0\}$. For $t = 1 - \max\{s, q\}, \dots, 0$, we supplement this initial condition by setting $y_{i,t} = I(u_{i,t} \leq p_{i,t})$, where the variables $u_{i,t}$ are $U[0, 1]$ and independent of one another (and independent of all $u_{i,t}$

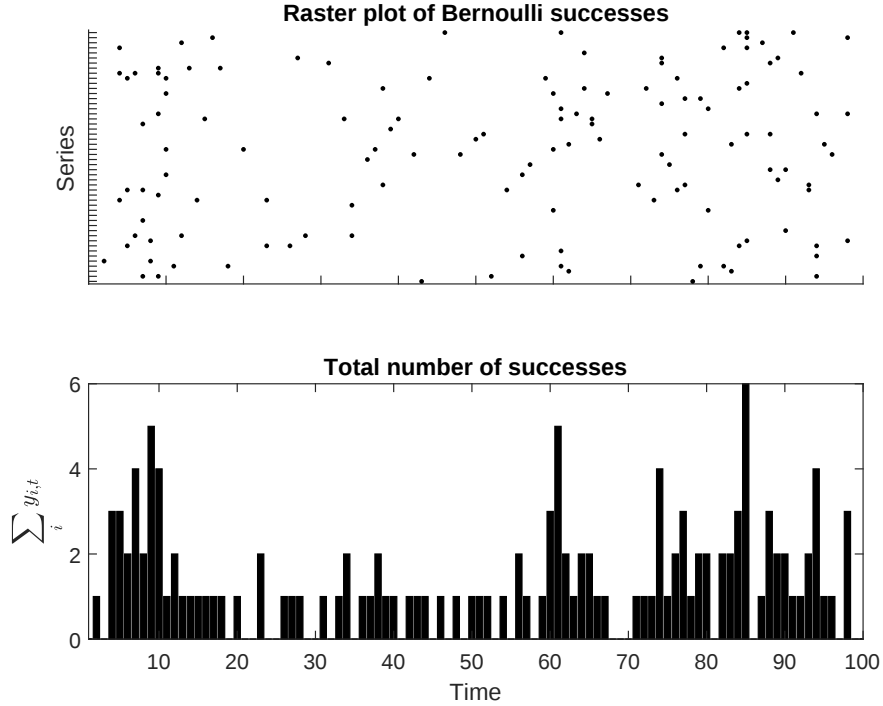


FIGURE 2. Individual and aggregate successes for $N = 50$ series generated from the interactive model (8) with $\omega_i = 0.005$, $\alpha_i = 0.01$, $\beta_i = 0.6$, $\gamma_i = 0.2$, and $p_{i,0} = 1/38$ for all i . The top panel shows the raster of individual successes; the bottom panel reports the total number of successes at each time period.

appearing in Eq. (1)). The initial probabilities $\mathbf{p}_{1-\max\{s,q\}}, \dots, \mathbf{p}_0$ are allowed to be random. In this case, all errors $u_{i,t}$ are assumed to be independent of the initial condition.

Our focus is on the case of rare events, where each $p_{i,t}$ tends to zero as $N \rightarrow \infty$. Studying such rare events is important because many applications involve outcomes that occur infrequently but whose aggregate frequency carries significant information. Examples include default counts in large credit portfolios, the number of extreme returns across assets, or spikes in systemic risk indicators. In these settings aggregation not only smooths out idiosyncratic noise but also highlights the limiting distributional behavior of the system, providing a tractable framework for inference and risk assessment.

Figure 2 illustrates the behavior of the system when N is large and successes are rare. Although the total number of series is $N = 50$, the aggregate number of successes remains small throughout the sample. By contrast, Figure 1c considers a setting with only $N = 5$ series but substantially higher success probabilities. Despite these differences in cross-sectional size and individual success rates, the total number of successes is of comparable magnitude across the two data-generating processes.

As we show below, aggregation naturally leads to a Poisson autoregression/INGARCH, providing a micro-level theoretical foundation for the applicability of such models. In the linear case, the conditional mean of the aggregate count follows a GARCH-type recursion, recovering exactly the model of Ferland *et al.* (2006); Fokianos *et al.* (2009). More generally, our nonlinear Theorem 9 shows that richer micro-level interactions survive aggregation in a precise sense: the limiting intensity λ_t inherits a genuinely nonlinear structure, yielding a broader class of Poisson autoregressions than the parametric nonlinear specifications considered in Fokianos *et al.* (2009, Section 2.3).

Let us define a linear Poisson autoregression of order 1, $[X_t, \lambda_t]_{t=0,1,\dots}$. Here $X_t \in \mathbb{Z}_{\geq 0}$, $\lambda_t \in \mathbb{R}_{\geq 0}$, and we denote by $\mathcal{I}_t^X$ the σ -field generated by $\{\lambda_0, X_0, \dots, \lambda_t, X_t, \lambda_{t+1}\}$. Then, for each $t \geq 0$, X_t follows a Poisson distribution, $Pois(\lambda_t)$, conditional on $\mathcal{I}_{t-1}^X$. For each $t > 0$, λ_t satisfies for deterministic $c^P, \gamma^P, \beta^P \geq 0$

$$(17) \quad \lambda_t = c^P + \gamma^P X_{t-1} + \beta^P \lambda_{t-1}.$$

Dynamics (17) implies that $\sigma(\lambda_0, X_0, \dots, \lambda_t, X_t, \lambda_{t+1}) = \sigma(\lambda_0, X_0, X_1, \dots, X_t)$.

Theorem 4. *Consider a multivariate model (8). Suppose that $\omega_i \equiv \omega_{i,N} = \frac{c_i}{N}$, $\alpha_i \equiv \alpha_{i,N} = \frac{a_i}{N^\kappa}$, $\kappa > 0$, $0 \leq c_i, a_i \leq C < \infty$, $\beta_i \equiv \beta$, $i = 1, \dots, N$. Additionally suppose that there exist limits $\bar{c} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N c_i$ and $\bar{\gamma} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \gamma_i$. Finally, suppose that the initial condition satisfies*

$$(18) \quad \sum_{i=1}^N p_{i,0} \xrightarrow[N \rightarrow \infty]{d} \lambda_0, \quad \sum_{i=1}^N \mathbb{E} p_{i,0} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_0, \quad \max_{i=1,\dots,N} p_{i,0} \xrightarrow[N \rightarrow \infty]{p} 0,$$

where λ_0 is some random variable. Then as $N \rightarrow \infty$, $\left[\sum_{i=1}^N y_{i,t}, \sum_{i=1}^N p_{i,t} \right]_{t=0,1,\dots}$ converges in finite-dimensional distributions to a Poisson autoregression of order 1, $[X_t, \lambda_t]_{t=0,1,\dots}$ with dynamics (17), $c^P = \bar{c}$, $\gamma^P = \bar{\gamma}$, $\beta^P = \beta$.

Theorem 4 shows that different specifications for $\mathbf{p}_t$ can generate identical aggregate dynamics. For example, suppose that in both models $\alpha_i \equiv 0$, and in the first model $\gamma_i = \frac{1}{3}$ for all i , while in the second model $\gamma_i = \frac{2}{3}$ for half of the units and $\gamma_i = 0$ for the rest, with all other coefficients are identical across the two models. In both cases, the average coefficient is $\bar{\gamma} = \frac{1}{3}$, implying identical aggregate dynamics. However, at the panel level, the two models differ substantially: in the second model, half of the probabilities evolve deterministically given $\mathbf{p}_0$.

Remark 5. *One can extend the model to allow for heterogeneous β_i , as long as $\beta_i = \beta + o(1)$.*

Remark 6. Condition $\alpha_i \rightarrow 0$ as $N \rightarrow \infty$ is essential for the Poisson–limit result. Although the rate of convergence may vary across i , keeping α_i fixed would change the asymptotic behavior of the total sum $\sum_{i=1}^N y_{i,t}$. If α_i does not vanish, with positive probability we get $y_{i,t} = 1$, so that $p_{i,t+1} \geq \alpha_i$. Therefore, at least one unit retains a non-negligible success probability, violating the small-probability condition required by the Poisson limit theorem. Given that $\alpha_i + \beta_i + \gamma_i < \bar{C} < 1$ for all i , we expect that the effect of $y_{i,t} = 1$ will decay over time. Consequently, as the system evolves, the aggregate process behaves approximately like a Poisson process with an additional serially correlated component induced by past feedback effects.

For example, if $\alpha_i \equiv \alpha > 0$, $c_i \equiv c > 0$, and $\beta = \gamma_i = 0$, then, conditionally on $X_t = k$, at $t + 1$ there are k non-negligible probabilities $p_{i,t+1} = c/N + \alpha \approx \alpha$ and $N - k$ probabilities $p_{i,t+1} = c/N$. Thus, aggregating the outcomes of the former, we get $\text{Bin}(k, \alpha)$, while aggregating the outcomes of the latter, we get an independent $\text{Pois}(c)$ component, and

$$X_{t+1} \stackrel{d}{=} \text{Pois}(c) + \text{Bin}(X_t, \alpha).$$

Remark 7. If $\lambda_0 \sim \text{Gamma}(\mathbf{m}, b)$, where $\mathbf{m}$ is a shape parameter and b is a rate parameter, then the unconditional distribution of X_0 is negative binomial, $\text{NB}(\mathbf{m}, \frac{b}{1+b})$. If additionally, $\bar{c} = \bar{\gamma} = 0$, so that the evolution of the Poisson intensity (17) is deterministic given λ_0 , then $\lambda_t \equiv \beta^t \lambda_0 \sim \text{Gamma}(\mathbf{m}, b\beta^{-t})$. Thus, the unconditional distribution of X_t is again negative binomial, $\text{NB}(\mathbf{m}, \frac{b}{b+\beta^t})$.

There are many cases when the initial condition (18) is satisfied. For example, if $p_{i,0} = \frac{r_i}{N}$, where r_i are i.i.d. with finite mean. Then $\frac{1}{N} \max_{i=1,\dots,N} r_i \xrightarrow[N \rightarrow \infty]{P} 0$ and $\sum_{i=1}^N p_{i,0} = \frac{1}{N} \sum_{i=1}^N r_i \xrightarrow[N \rightarrow \infty]{P} \mathbb{E} r_1$. Moreover, it is satisfied by the stationary solution, as shown below.

Proposition 8. If $\beta + \gamma_i < \bar{C} < 1$ for all i , then the unique stationary solution of the multivariate model (8) satisfies assumption on the initial condition (18).

Theorem 4 can be extended to allow for multiple lags and various nonlinearities. Consider a nonlinear interactive model, which generalizes model (8):

$$(19) \quad \begin{aligned} p_{i,t} = & \frac{c_i}{N} + \frac{1}{N\kappa} f_{\alpha,i}(y_{i,t-1}, \dots, y_{i,t-s}) + \sum_{\tau=1}^s \beta_\tau p_{i,t-\tau} \\ & + f_{\gamma,i} \left(\frac{1}{N} \sum_{j=1}^N y_{j,t-1}, \dots, \frac{1}{N} \sum_{j=1}^N y_{j,t-s}, \frac{1}{N} \sum_{j=1}^N p_{j,t-1}, \dots, \frac{1}{N} \sum_{j=1}^N p_{j,t-s} \right) \\ & + \frac{1}{N} \gamma_i^\top f_\gamma \left(\sum_{j=1}^N y_{j,t-1}, \dots, \sum_{j=1}^N y_{j,t-s}, \sum_{j=1}^N p_{j,t-1}, \dots, \sum_{j=1}^N p_{j,t-s} \right), \end{aligned}$$

where $\kappa > 0$, $\beta_\tau \geq 0$, $0 \leq c_i \leq C$, $\gamma_i \in \mathbb{R}_+^k$ with $\|\gamma_i\|_1 \leq C$, $f_{\alpha,i} : \{0,1\}^s \rightarrow \mathbb{R}_+$, $f_{\gamma,i} : [0,1]^{2s} \rightarrow \mathbb{R}_+$, and $f_\gamma : \mathbb{R}_+^{2s} \rightarrow \mathbb{R}_+^k$. We assume

$$f_{\alpha,i}(0, \dots, 0) = f_{\gamma,i}(0, \dots, 0) = 0, \quad f_\gamma(0, \dots, 0) = 0.$$

We introduce the nonlinear Poisson autoregression counterpart of the interactive model (19), $[X_t, \lambda_t]_{t > -s}$. Denote by $\mathcal{I}_t^X$ the σ -field generated by $\{\lambda_{-s+1}, X_{-s+1}, \dots, \lambda_t, X_t, \lambda_{t+1}\}$. Then, for each $t > -s$, X_t follows a Poisson distribution, $Pois(\lambda_t)$, conditional on $\mathcal{I}_{t-1}^X$. For each $t > 0$, λ_t satisfies for deterministic $c^P \geq 0$, $\gamma^P \in \mathbb{R}_+^k$, $\beta^P \in \mathbb{R}_+^{2s}$

$$(20) \quad \lambda_t = c^P + (X_{t-1}, \dots, X_{t-s}, \lambda_{t-1}, \dots, \lambda_{t-s})\beta^P + [f_\gamma(X_{t-1}, \dots, X_{t-s}, \lambda_{t-1}, \dots, \lambda_{t-s})]^\top \gamma^P.$$

Dynamics (20) implies that

$$\sigma(\lambda_{-s+1}, X_{-s+1}, \dots, \lambda_t, X_t, \lambda_{t+1}) = \sigma(\lambda_{-s+1}, X_{-s+1}, \dots, \lambda_0, X_0, X_1, \dots, X_t).$$

Theorem 9. *Consider a nonlinear interactive model (19). Suppose that $\kappa, C, A, L_\gamma, M_1, M_2$ are finite constants independent of N . Suppose that, for all i ,*

$$0 \leq f_{\alpha,i}(y_1, \dots, y_s) \leq A \sum_{\tau=1}^s y_\tau.$$

Suppose that, for all i , $f_{\gamma,i}$ is twice continuously differentiable and with uniformly bounded gradient and Hessian,

$$\sup_{x \in [0,1]^{2s}} \|\nabla f_{\gamma,i}(x)\|_1 \leq M_1, \quad \sup_{x \in [0,1]^{2s}} \|\nabla^2 f_{\gamma,i}(x)\| \leq M_2.$$

Suppose that f_γ is Lipschitz,

$$\|f_\gamma(x) - f_\gamma(x')\|_\infty \leq L_\gamma \|x - x'\|_\infty, \quad x, x' \in \mathbb{R}_+^{2s},$$

Suppose that $\sum_{\tau=1}^s \beta_\tau + M_1 + CL_\gamma < 1$ and there exist limits

$$\bar{c} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N c_i, \quad \bar{\gamma} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \gamma_i, \quad \bar{\mathbf{f}} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \nabla f_{\gamma,i}(0, \dots, 0).$$

Finally, suppose that the initial condition satisfy

$$\sum_{i=1}^N p_{i,\tau} \xrightarrow[N \rightarrow \infty]{d} \lambda_\tau, \quad \sum_{i=1}^N \mathbb{E} p_{i,\tau} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_\tau, \quad \max_{i=1, \dots, N} p_{i,\tau} \xrightarrow[N \rightarrow \infty]{p} 0,$$

where $(\lambda_{1-s}, \dots, \lambda_0)$ is a random vector and convergence is joint over $\tau = -s+1, \dots, 0$. Then as $N \rightarrow \infty$, $\left[\sum_{i=1}^N y_{i,t}, \sum_{i=1}^N p_{i,t} \right]_{t=-s+1, -s+2, \dots}$ converges in finite-dimensional distributions to a nonlinear Poisson autoregression of order s , $[X_t, \lambda_t]_{t=-s+1, -s+2, \dots}$ with dynamics (20), $c^P = \bar{c}$, $\gamma^P = \bar{\gamma}$, $\beta^P = \bar{\mathbf{f}} + (0, \dots, 0, \beta_1, \dots, \beta_s)^\top$.

Condition $\sum_{\tau=1}^s \beta_\tau + M_1 + CL_\gamma < 1$ implies that, starting from N large enough, the right-hand side of (19) belongs to $[0, 1]$ for every i and t . Conditions $f_{\gamma,i}(\cdot) \geq 0$ and $f_{\gamma,i}(0, \dots, 0) = 0$ imply that the gradient $\nabla f_{\gamma,i}(0, \dots, 0) \in \mathbb{R}_+^{2s}$, so that $\bar{\mathbf{f}} \in \mathbb{R}_+^{2s}$.

When $s = k = 1$, $f_{\alpha,i}(y) = a_i y$, $f_{\gamma,i} \equiv 0$, $f_\gamma(y, p) = y$, Theorem 9 reduces to Theorem 4. Alternatively, one can set $f_\gamma \equiv 0$ and $f_{\gamma,i}(y, p) = \gamma_i y$. The linear multi-lag extension of Theorem 4 can be obtained by setting $k = 1$, $f_{\alpha,i}(y_1, \dots, y_s) = \sum_{\tau=1}^s a_{i,\tau} y_\tau$, $f_{\gamma,i}(y, p) \equiv 0$, $f_\gamma(y_1, \dots, y_s, p_1, \dots, p_s) = \sum_{\tau=1}^s (\delta_\tau^y y_\tau + \delta_\tau^p p_\tau)$, where $\delta_\tau^y, \delta_\tau^p \in \mathbb{R}_+$. Alternatively, one can set $f_\gamma \equiv 0$ and $f_{\gamma,i}(y_1, \dots, y_s, p_1, \dots, p_s) = \gamma_i \sum_{\tau=1}^s (\delta_\tau^y y_\tau + \delta_\tau^p p_\tau)$. Both representations recover the same limiting intensity λ_t , because for linear functions the prefactor $1/N$ can be taken out of the argument.

The term $\frac{1}{N^\kappa} f_{\alpha,i}(y_{i,t-1}, \dots, y_{i,t-s})$ is a nonlinear and multi-lag generalization of $\frac{a_i}{N^\kappa} y_{i,t-1}$. For example, it allows for across-time interactions $y_{i,t-1} y_{i,t-2}$. This term disappears in the limit due to the prefactor $N^{-\kappa}$. Without the prefactor the term can occasionally take non-negligible values, violating the requirement of the Poisson approximation that all probabilities are small. The term $f_{\gamma,i}(\cdot)$ is a nonlinear and multi-lag generalization of the interactions, which are governed by the average number of successes. The averages are $o_p(1)$, leading to the average gradient of $f_{\gamma,i}$ in the limit as the outcome of the first-order Taylor approximation. Thus, these nonlinearities asymptotically become linear. Finally, the term $\frac{1}{N} \gamma_i^\top f_\gamma(\cdot)$ is the one which remains nonlinear in the limit, due to its arguments converging to non-degenerate random variables. Unlike the $f_{\alpha,i}(\cdot)$ and $f_{\gamma,i}(\cdot)$ terms, where individual heterogeneity enters through i -indexed coefficients that average out in the limit, $f_\gamma(\cdot)$ is common across all i : every coordinate i is exposed to the same aggregate quantity $f_\gamma\left(\sum_{j=1}^N y_{j,t-1}, \dots, \sum_{j=1}^N y_{j,t-s}, \sum_{j=1}^N p_{j,t-1}, \dots, \sum_{j=1}^N p_{j,t-s}\right)$, with individual heterogeneity entering only through γ_i .

A distinctive feature of the aggregation in Theorems 4 and 9 is that each coordinate i interacts with the equal-weight cross-sectional aggregates $\sum_{j=1}^N y_{j,t}$ and $\sum_{j=1}^N p_{j,t}$, leaving no room for heterogeneous interaction weights. In some special cases, such as the network GAB model, this restriction can be relaxed, as illustrated in Theorem 10 below.

Theorem 10. *Consider a multivariate network model (9). Suppose that the network matrix W is doubly stochastic and all out-degrees satisfy $d_i^{\text{out}}/\log N \geq d, d > 0$, where $d_i^{\text{out}} := \sum_{j=1}^N I(W_{ij} \neq 0)$. Suppose that $\omega_i \equiv \omega_{i,N} = \frac{c_i}{N}$, $\alpha_i \equiv \alpha_{i,N} = \frac{a_i}{N^\kappa}$, $\kappa > 0$, $0 \leq c_i, a_i \leq C < \infty$, $\beta_i \equiv \beta$, $\gamma_i \equiv \gamma$, $i = 1, \dots, N$, and there exists limit $\bar{c} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N c_i$. Finally, suppose that*

the initial condition satisfies

$$\sum_{i=1}^N p_{i,0} \xrightarrow[N \rightarrow \infty]{d} \lambda_0, \quad \sum_{i=1}^N \mathbb{E} p_{i,0} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_0, \quad \max_{i=1, \dots, N} p_{i,0} \xrightarrow[N \rightarrow \infty]{p} 0,$$

where λ_0 is some random variable. Then as $N \rightarrow \infty$, $\left[\sum_{i=1}^N y_{i,t}, \sum_{i=1}^N p_{i,t} \right]_{t=0,1,\dots}$ converges in finite-dimensional distributions to a Poisson autoregression of order 1, $[X_t, \lambda_t]_{t=0,1,\dots}$ with dynamics (17), $c^P = \bar{c}$, $\gamma^P = \gamma$, $\beta^P = \beta$.

A row-normalized network matrix W is doubly stochastic if, for example, for each i , $d_i^{\text{in}} = d_i^{\text{out}} = d$, where $d_i^{\text{in}} = \sum_{j=1}^N I(W_{ji} \neq 0)$, $d_i^{\text{out}} = \sum_{j=1}^N I(W_{ij} \neq 0)$ are in- and out-degrees of a vertex i . That means, each individual has d incoming and d outgoing links, i.e. the network is d -regular. The doubly stochastic structure, combined with homogeneity of $\gamma_i = \gamma$, ensures that the network weights wash out in the limit. Thus, the limiting intensity equation in Theorem 10 is identical to that of Theorem 4 with $\gamma_i = \gamma$.

The condition on the out-degree growing with N ensures that terms $\sum_j W_{ij} y_{j,t-1}$ are of order $o_p(1)$, mimicking the scaling of the equal-weight average $\frac{1}{N} \sum_{j=1}^N y_{j,t}$ and thereby ensuring the Poisson approximation applies.

4. ESTIMATION

We begin by discussing estimation when success probabilities are bounded away from 0 and 1, the regime where traditional maximum likelihood methods apply, such as Moysiadis and Fokianos (2014); Davis and Liu (2016); Fokianos and Moysiadis (2017); Aknouche and Francq (2021); Lee *et al.* (2023). We then briefly discuss what can be done in rare-event scenarios.

Suppose that the success probability function $\mathbf{g}(\cdot)$ depends on an unknown parameter $\theta_0 \in \mathbb{R}^m$, such as the coefficients in the models discussed in Section 2.1. Suppose time periods $t = 1, \dots, T$ are observed. Then the parameter θ_0 can be estimated by maximum likelihood. Since the innovations $u_{i,t}$ are independent and uniformly distributed on $[0, 1]$, the likelihood and log-likelihood functions conditional on the initial values $\mathbf{y}_0, \dots, \mathbf{y}_{-q+1}, \mathbf{p}_0, \dots, \mathbf{p}_{-s+1}$ take the form

$$\mathcal{L}_T(\theta) = \prod_{t=1}^T \prod_{i=1}^N (\tilde{p}_{i,t})^{y_{i,t}} (1 - \tilde{p}_{i,t})^{1-y_{i,t}}, \quad \text{where } \tilde{\mathbf{p}}_t \text{ is recursively defined by}$$

$$\tilde{\mathbf{p}}_t = \begin{cases} \mathbf{g}(\{\tilde{\mathbf{p}}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta), & t > 0, \\ \mathbf{p}_t, & t \leq 0, \end{cases}$$

$$\mathcal{Q}_T(\theta) := \log \mathcal{L}_T(\theta) = \sum_{t=1}^T \sum_{i=1}^N [y_{i,t} \log \tilde{p}_{i,t} + (1 - y_{i,t}) \log (1 - \tilde{p}_{i,t})].$$

In empirical applications the first q realizations of the binary process $\mathbf{y}_t$ can be used as initial values. The corresponding probabilities $\mathbf{p}_t$, however, are unobserved. A practical approach is, therefore, to initialize them using the sample means of $\mathbf{y}_t$. In this section we impose Lipschitz-type contraction of Assumptions I, II (in which the constants are assumed to be θ -independent). As shown in the proof of Theorem 11, this guarantees that the effect of the probability initialization on the normalized log-likelihood vanishes uniformly over θ .

Assumption III.

- (i) $\theta_0 \in \Theta$ for some compact set $\Theta \subset \mathbb{R}^m$;
- (ii) $\mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta)$ is continuous in $\theta \in \Theta$;
- (iii) $\exists \varepsilon > 0$ such that for all $i = 1, \dots, N$, $\theta \in \Theta$, $g_i(\cdot; \theta) \in [\varepsilon, 1 - \varepsilon]$;
- (iv) Let $\mathbf{y}_t, \mathbf{p}_t$ be the stationary solution of Theorem 3. If

$$\text{Prob}(\mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta) = \mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta_0)) = 1,$$

then $\theta = \theta_0$.

In all linear GAB examples, we do not lose any generality by imposing the compactness condition, because $\mathbf{g}$ maps to $[0, 1]^N$. For nonlinear $\mathbf{g}$, compactness of the parameter space can be relaxed if, with positive probability, $\mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta)$ goes to 0 or 1 as $\theta \rightarrow \infty$. This ensures that the log-likelihood is never maximized at unbounded values of θ .

Assumption III(iii) is introduced to keep the log-likelihood and its derivatives uniformly bounded. It resembles the standard GARCH condition that the intercept be bounded away from zero (see, e.g., Francq and Zakoian (2019, Example 7.1)). In our setup, when the functions g_i are linear, this means that all intercepts are separated from zero and that the sums of all coefficients in g_i are separated from one. Note that if, for all i , $g_i(0, \dots, 0; \theta_0) = 0$, then $y_{i,t} = p_{i,t} = 0$ is a degenerate stationary solution. Similarly, if, for all i , $g_i(1, \dots, 1; \theta_0) = 1$, then $y_{i,t} = p_{i,t} = 1$ is a degenerate stationary solution. In both cases, the log-likelihood becomes infinite. By imposing III(iii), we rule out such degenerate settings.

Assumption III(iv) requires the existence of a unique parameter value θ that is compatible with the data, i.e., identification of θ_0 .

The maximum likelihood estimator solves

$$\hat{\theta}^{MLE} = \arg \max_{\theta \in \Theta} \mathcal{Q}_T(\theta).$$

Theorem 11. *Suppose that the multivariate binary time series model (1)–(2) satisfies Assumption III and either Assumption I or II, in which the constants are assumed to be θ -independent. Then $\hat{\theta}^{MLE} \xrightarrow[T \rightarrow \infty]{p} \theta_0$.*

Assumption I or II is imposed to ensure ergodicity and uniqueness of a stationary solution.

Assumption IV.

- (i) θ_0 is an interior point of $\Theta \in \mathbb{R}^m$;
- (ii) There exists a compact neighborhood $\Theta_0 \subset \Theta$ of θ_0 such that, for every fixed $(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-q})$, the function $\mathbf{g}$ is twice continuously differentiable jointly in $(\mathbf{p}_{t-1}, \dots, \mathbf{p}_{t-s}, \theta)$ on $[0, 1]^{N_s} \times \Theta_0$;
- (iii) The matrix

$$H_0 = \mathbb{E} \sum_{i=1}^N \frac{1}{g_{i,t,\theta_0}(1 - g_{i,t,\theta_0})} \left(\frac{d}{d\theta} g_{i,t,\theta_0} \right) \left(\frac{d}{d\theta} g_{i,t,\theta_0} \right)^\top$$

is nonsingular, where

$$g_{i,t,\theta_0} := g_i(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}; \theta_0)$$

and $\frac{d}{d\theta} g_{i,t,\theta_0}$ denotes the total derivative evaluated at θ_0 , accounting for the dependence of the lagged probabilities on θ through the recursion. More precisely, if $D_t \in \mathbb{R}^{N \times m}$ has i th row $(dg_{i,t,\theta_0}/d\theta)^\top$, then D_t is the unique bounded stationary solution of the recursion

$$D_t = G_{\theta,t} + \sum_{j=1}^s G_{p_j,t} D_{t-j},$$

where

$$G_{\theta,t} := \frac{\partial}{\partial \theta^\top} \mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}; \theta) \Big|_{\theta=\theta_0},$$

and

$$G_{p_j,t} := \frac{\partial}{\partial \mathbf{p}_{t-j}^\top} \mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}; \theta) \Big|_{\theta=\theta_0}.$$

Theorem 12. *Suppose that the multivariate binary time series model (1)–(2) satisfies Assumptions III, IV, and either I or II, in which the constants are assumed to be θ -independent. Then $\sqrt{T}(\hat{\theta}^{MLE} - \theta_0) \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, H_0^{-1})$.*

Note that the boundedness condition in Assumption III(iii) rules out rare events, where $p_{i,t}$ is close to zero. Both consistency and asymptotic normality rely on this condition: it ensures the log-likelihood is well-defined and uniformly dominated, and keeps the score and Hessian bounded. As we illustrated in Section 3, when N is large and $p_{i,t}$ is small, the aggregate

count $\sum_{i=1}^N y_{i,t}$ is approximately Poisson with mean $\sum_{i=1}^N p_{i,t}$. Thus, a natural alternative for estimation in such settings is the Poisson autoregressive model, which replaces the binary observation equation with a Poisson likelihood.

In the framework of Theorem 4, when $\beta < 1$, $\gamma_i = \nu c_i$ for all i (that is, when some form of homogeneity is present), the following intuition can be used to identify the Poisson log-likelihood in the MLE maximization in (4). First, in line with $\max_i p_{i,t} \xrightarrow[N \rightarrow \infty]{p} 0$, we expect

$$\left| \sum_i y_{i,t} \log(1 - g_{i,t,\theta}) \right| \leq |\log(1 - \max_i g_{i,t,\theta})| \sum_i y_{i,t} \xrightarrow[N \rightarrow \infty]{p} 0.$$

Therefore,

$$\begin{aligned} & \sum_{i=1}^N y_{i,t} \log g_{i,t,\theta} + (1 - y_{i,t}) \log(1 - g_{i,t,\theta}) \\ &= \sum_{i=1}^N \log(1 - g_{i,t,\theta}) + \sum_{i=1}^N y_{i,t} \log g_{i,t,\theta} - \sum_{i=1}^N y_{i,t} \log(1 - g_{i,t,\theta}) \\ &\approx - \sum_{i=1}^N g_{i,t,\theta} + \left(\sum_{i=1}^N y_{i,t} \right) \log \left(\sum_{j=1}^N g_{j,t,\theta} \right) + \sum_{i=1}^N y_{i,t} \log \left(g_{i,t,\theta} / \sum_{j=1}^N g_{j,t,\theta} \right), \end{aligned}$$

so that the first two terms operate on aggregate quantities and exactly match the Poisson log-likelihood. The last term is close to being independent of aggregate quantities and can therefore be maximized separately. To see this, applying iterative back-substitution yields

$$\begin{aligned} p_{i,t} &= \frac{c_i}{N(1-\beta)} + \beta^t \left(p_{i,0} - \frac{c_i}{N(1-\beta)} \right) + \frac{a_i}{N^\kappa} \sum_{\tau=0}^{t-1} \beta^\tau y_{i,t-1-\tau} + \frac{\nu c_i}{N} \sum_{\tau=0}^{t-1} \beta^\tau \sum_{j=1}^N y_{j,t-1-\tau}, \\ \sum_i p_{i,t} &= \frac{\sum_i c_i}{N(1-\beta)} + \beta^t \left(\sum_i p_{i,0} - \frac{\sum_i c_i}{N(1-\beta)} \right) \\ &\quad + \sum_i \frac{a_i}{N^\kappa} \sum_{\tau=0}^{t-1} \beta^\tau y_{i,t-1-\tau} + \left(\frac{\nu \sum_i c_i}{N} \right) \sum_{\tau=0}^{t-1} \beta^\tau \sum_{j=1}^N y_{j,t-1-\tau}. \end{aligned}$$

When t is large, $\beta^t \approx 0$, and the second terms in both expressions above are negligible. Since $\kappa > 0$, the third terms are also negligible, leaving

$$\frac{g_{i,t,\theta}}{\sum_{j=1}^N g_{j,t,\theta}} \approx \frac{\frac{c_i}{N} \left(\frac{1}{1-\beta} + \nu \sum_{\tau=0}^{t-1} \beta^\tau \sum_{j=1}^N y_{j,t-1-\tau} \right)}{\frac{\sum_j c_j}{N} \left(\frac{1}{1-\beta} + \nu \sum_{\tau=0}^{t-1} \beta^\tau \sum_{j=1}^N y_{j,t-1-\tau} \right)} = \frac{c_i}{\sum_j c_j}.$$

That is, only the relative parameters $\tilde{c}_i := c_i / \sum_j c_j$ remain. These lie on a simplex and are independent of the aggregate values, so the original maximization problem splits into two separate problems, the first of which corresponds to a Poisson likelihood.

5. COVARIATES

The model can be extended to include exogenous covariates $\mathbf{z}_t \in \mathbb{R}^k$, where the covariate process $\{\mathbf{z}_t\}$ is independent of the innovation process $\{\mathbf{u}_t\}$, so that

$$p_{i,t} = g_i(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}, \mathbf{z}_t).$$

Provided $\mathbf{z}_t$ is strictly stationary and ergodic, Theorems 1 and 3 hold unchanged under Assumptions I–II as stated. This follows because in the coupling construction both trajectories share the same realization of $\mathbf{z}_t$, so it cancels identically from the Lipschitz bounds (14) and (15). Similarly, estimation Theorems 11 and 12 continue to hold.

The more interesting question is how to incorporate covariates in the aggregation framework. If k is fixed, then $\mathbf{z}_t$ must enter each equation for $p_{i,t}$ with a coefficient vanishing in N . Otherwise, the finite order of $\mathbf{z}_t$ would prevent $p_{i,t}$ from being of order $1/N$. Paralleling the treatment of the intercept ω_i , we therefore add a linear term $\frac{1}{N}\delta_i^\top \mathbf{z}_t$ to each $p_{i,t}$ equation in interactive and network GAB models. Letting $\bar{\delta} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i$, this contributes an extra term $\bar{\delta}^\top \mathbf{z}_t$ to the right-hand side of the evolution equation for λ_t .

If k grows with N , the covariates must themselves aggregate to a limiting quantity. We decompose $\mathbf{z}_t = (\mathbf{z}_t^{\text{com}}, \mathbf{z}_t^{\text{ind}})$, where $\mathbf{z}_t^{\text{com}} \in \mathbb{R}^{r_c}$ is a common component of fixed dimension affecting all units, and $\mathbf{z}_t^{\text{ind}} = ((\mathbf{z}_{1,t}^{\text{ind}})^\top, \dots, (\mathbf{z}_{N,t}^{\text{ind}})^\top)^\top \in \mathbb{R}^{Nr}$ stacks unit-specific components with $\mathbf{z}_{i,t}^{\text{ind}} \in \mathbb{R}^r$, so that $k = Nr + r_c$. The common component is handled exactly as in the fixed k case. For the individual components, $p_{i,t}$ may depend on $\mathbf{z}_{i,t}^{\text{ind}}$ directly and on the cross-sectional average $\frac{1}{N} \sum_{j=1}^N \mathbf{z}_{j,t}^{\text{ind}}$ (or the weighted average $\sum_{j=1}^N W_{ij} \mathbf{z}_{j,t}^{\text{ind}}$ in the network GAB case, as in Theorem 10). The direct effect of $\mathbf{z}_{i,t}^{\text{ind}}$ on $p_{i,t}$ must be scaled by $1/N$, for the same order- $1/N$ reason as above, and must appear with a homogeneous coefficient η , so that the aggregation yields $\eta \frac{1}{N} \sum_{i=1}^N \mathbf{z}_{i,t}^{\text{ind}}$. The cross-sectional average must carry an additional factor of $1/N$, since unlike $y_{i,t}$ and $p_{i,t}$, the scale of the exogenous $\mathbf{z}_{i,t}^{\text{ind}}$ is not reduced by aggregation; the corresponding term in $p_{i,t}$ is therefore $\frac{\zeta_i}{N} \frac{1}{N} \sum_{j=1}^N \mathbf{z}_{j,t}^{\text{ind}}$ (or nonlinear bounded³ function $\frac{\zeta_i}{N} f_z \left(\frac{1}{N} \sum_{j=1}^N \mathbf{z}_{j,t}^{\text{ind}} \right)$, so that f_z survives in the limit analogously to the f_γ term in Theorem 9). Thus, assuming that $\frac{1}{N} \sum_{i=1}^N \mathbf{z}_{i,t}^{\text{ind}} \xrightarrow[N \rightarrow \infty]{p} \bar{\mathbf{z}}_t$, $\mathbb{E} \frac{1}{N} \sum_{i=1}^N \mathbf{z}_{i,t}^{\text{ind}} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \bar{\mathbf{z}}_t < \infty$, and $\max_i \|\mathbf{z}_{i,t}^{\text{ind}}\|/N \xrightarrow[N \rightarrow \infty]{p} 0$ for each t , the aggregated covariate $\bar{\mathbf{z}}_t$ appears on the right-hand side of the limiting evolution equation for λ_t .

³Boundedness guarantees that $p_{i,t} \in [0, 1]$ and that $\mathbb{E} f_z \left(\frac{1}{N} \sum_{j=1}^N \mathbf{z}_{j,t}^{\text{ind}} \right) \xrightarrow[N \rightarrow \infty]{} \mathbb{E} f_z(\bar{\mathbf{z}}_t)$.

6. EMPIRICAL ILLUSTRATION

In this section, we apply the interactive GAB model to study periods of unusually low stock returns among the constituents of the S&P 100. Such episodes tend to coincide with market stress and elevated financial risk. We use binary indicators of tail events to study how downside risks emerge, cluster over time, and propagate across assets, defining a tail event for stock i as its return falling below the 5th percentile of its empirical distribution.

Our sample consists of daily returns for 94 S&P 100 constituents available over the full period 01.01.2012–31.12.2025, yielding 3,517 time series observations. The 94 stocks are those for which data are available throughout the entire span; we also retain only one of the two Google share classes. Further details on the sample composition are provided in Section D of the Appendix.

We subtract the risk-free rate and regress returns on the Fama-French five factors, then work with the idiosyncratic residuals. The Fama-French returns, which are reported in percent, are converted to decimal units before the regression. The final three years of data are reserved for forecast evaluation, so estimation uses observations up to 01.01.2023, giving $T = 2,768$. The factor regressions are estimated on this estimation sample only. The 5th percentile threshold for each stock is also computed on this estimation sample, and $y_{i,t} = 1$ whenever stock i 's idiosyncratic return falls below its threshold, and $y_{i,t} = 0$ otherwise. Figure 3 summarizes the resulting binary panel. As expected, tail events cluster around periods of market stress, including COVID and the Fed's monetary policy tightening. This clustering motivates the inclusion of the cross-sectional interaction term as in Eq. (8), which captures the market-wide stress and allows individual tail probabilities to adjust.

We first estimate the interactive GAB model (8). The resulting estimates are shown in Figure 4. In line with the aggregation framework, both ω_i and α_i are relatively small, with $\text{mean}(\omega_i) = 0.004$ and $\text{mean}(\alpha_i) = 0.03$. The averages of the remaining coefficients are $\text{mean}(\gamma_i) = 0.18$ and $\text{mean}(\beta_i) = 0.71$.

For comparison with the aggregation results, we also re-estimate the model imposing $\alpha_i = 0$, since this coefficient vanishes under aggregation, and compare the results with the Poisson MLE based on the aggregated data. The results are summarized in Table 1. The GAB and Poisson estimates display a similar persistence pattern, although finite- N differences remain, particularly for the intercept. When $\alpha_i = 0$ is imposed, the intercept absorbs the mean contribution previously attributed to α_i , leading to $\sum_i \omega_i$ larger than $\bar{c}$. The comparison is further illustrated by the filtered Poisson intensity λ_t and the sum of GAB probabilities $\sum_i p_{i,t}$, shown in Figure 5a.

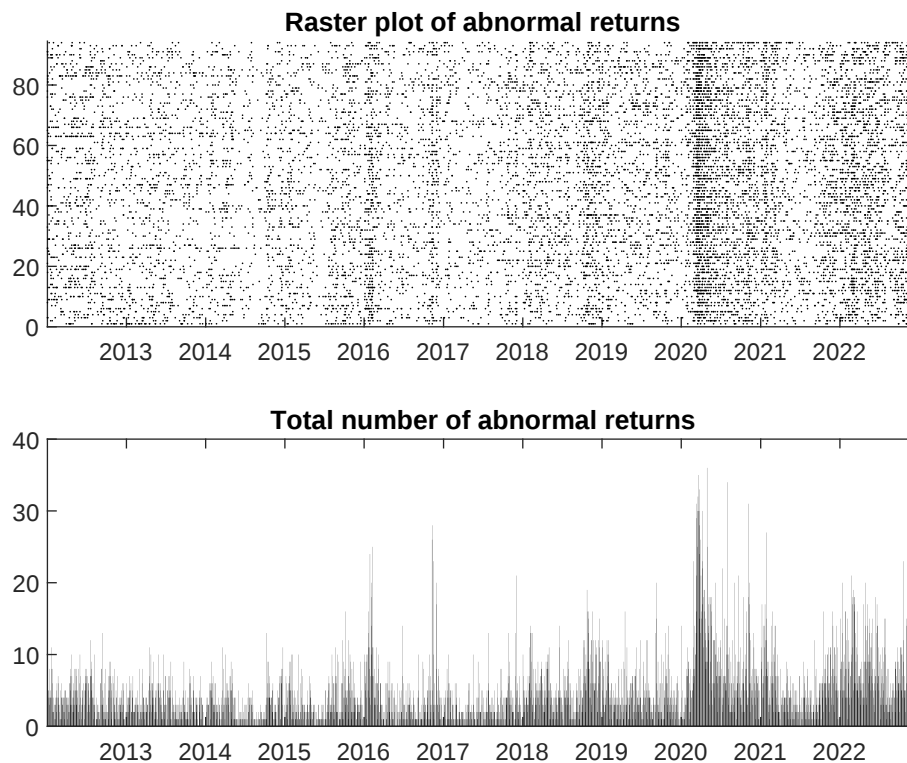


FIGURE 3. Abnormally low (below 5%) returns in S&P 100.

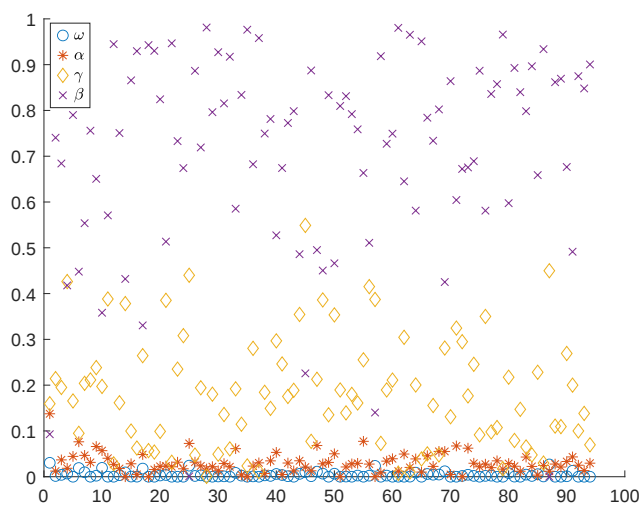


FIGURE 4. Interactive GAB estimates for 94 stocks.

Finally, we compare the forecasting performance of the GAB model using the three-year sample that was excluded from the estimation stage. The out-of-sample binary data are

	$\sum_i \omega_i$	$\text{mean}(\omega_i)$	$\text{mean}(\alpha_i)$	$\text{mean}(\gamma_i)$	$\text{mean}(\beta_i)$
Interactive GAB	0.37	0.004	0.03	0.18	0.71
Interactive GAB with $\alpha_i = 0$	0.51	0.005	–	0.24	0.65
	$\bar{c}$			$\bar{\gamma}$	β
Aggregated Poisson	0.2	–	–	0.2	0.75

TABLE 1. Estimation Results

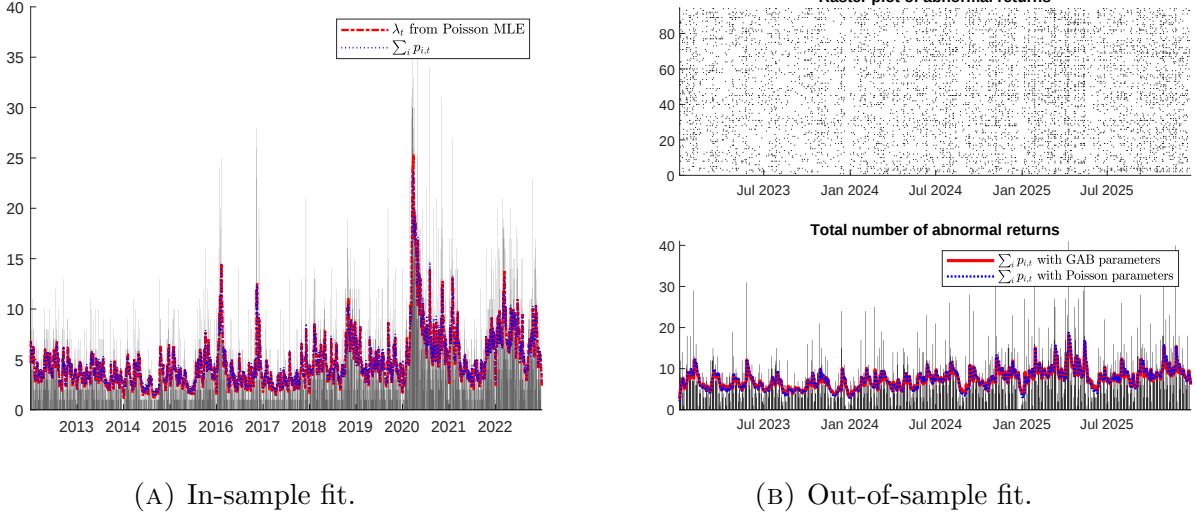


FIGURE 5. Aggregate GAB probabilities $\sum_i p_{i,t}$ and Poisson intensity λ_t . The in-sample GAB path uses the restricted specification with $\alpha_i = 0$ for comparison with the aggregation result, whereas the out-of-sample GAB path uses the unrestricted specification.

constructed by comparing returns to the same 5th percentile threshold used previously and are shown in Figure 5b. The out-of-sample period features a higher fraction of abnormally low returns than the in-sample 5% benchmark (8.7% of the data points), indicating elevated uncertainty. The tail events nevertheless remain dispersed across stocks: even on the most concentrated out-of-sample day, 43.6% of stocks are classified as tail events. We generate recursive 1-step-ahead forecasts of $y_{i,t}$ over the out-of-sample period ($T + 1, \dots, T + 749$, three years of trading days) based on four models.

The first model is the unrestricted interactive GAB specification (heterogeneous parameters). The second is an interactive GAB model with homogeneous parameters calibrated from the Poisson maximum likelihood estimates, i.e., $\omega_i = \bar{c}/N$, $\alpha_i = 0$, $\gamma_i = \bar{\gamma}$, and $\beta_i = \beta$. The third is a constant forecast $\hat{y}_{i,t} = 0.05$, reflecting the unconditional 5% probability of observing a success. The fourth is a persistence-based forecast $\hat{y}_{i,t} = y_{i,t-1}$.

We report the mean squared errors of these forecasts in Table 2. The results show that the first model, the unrestricted interactive GAB with heterogeneous parameters, achieves the

Model 1	Model 2	Model 3	Model 4
0.0789	0.0791	0.0805	0.154

TABLE 2. Out-of-sample MSE (averaged across stocks and the out-of-sample period). Model 1: unrestricted interactive (heterogeneous) GAB. Model 2: Poisson (homogeneous GAB). Model 3: 5% probability of success. Model 4: tomorrow is the same as today.

best forecasting performance. Although the GAB and Poisson estimation results in Table 1 are close, the second (homogeneous) model performs slightly worse. A Diebold–Mariano test based on the time series of daily cross-sectional mean squared losses, with a Newey–West long-run variance estimator using five lags, gives $DM = -2.08$ and $p = 0.004$ for Model 1 versus Model 2, so the forecasting improvement is statistically significant. Both naive benchmarks are substantially outperformed.⁴

Figure 5b also presents the paths of $\sum_i p_{i,t}$ for the first two models. For the second model, this path coincides with the trajectory of the Poisson intensity λ_t . The superior forecasting performance of Models 1 and 2 relative to the naive benchmarks highlights the importance of the interaction term $\frac{1}{N} \sum_i y_{i,t}$ in capturing overall market conditions, a term absent from Models 3 and 4.

7. CONCLUSION

This paper develops a general class of multivariate binary autoregressive models, in which each observation is a Bernoulli variable whose success probability evolves according to a GARCH-type recursion driven by past outcomes and past probabilities. The framework accommodates nonlinear dynamics, cross-sectional interactions, and network-weighted dependence structures.

Three sets of results are established. The bounded support of binary outcomes and probabilities guarantees, without any parameter restrictions, the existence of a strictly stationary solution, a feature with no counterpart in continuous-valued GARCH models, while uniqueness and ergodicity follow from a coupling argument that exploits a Lipschitz contraction condition on the probability update function. The most distinctive results concern aggregation: under a rare-events scaling in which individual success probabilities are of order $1/N$, the aggregate number of successes converges to a Poisson autoregression, providing a rigorous

⁴This is consistent with the MSE results. Suppose that the observations are i.i.d. Bernoulli. Then, for the fourth model, the squared forecast error is equal to 0 when two consecutive realizations coincide and 1 when they differ. The latter occurs with probability $2p(1-p) \approx 0.16$ for $p \approx 0.087$. For the third model, the squared forecast error takes values 0.05^2 and 0.95^2 , and $MSE \approx 0.913 \cdot 0.05^2 + 0.087 \cdot 0.95^2 \approx 0.08$. For the first two models, $MSE \approx 0.079$ is close to the optimal value, which corresponds to the forecast $\hat{y}_{i,t} \approx 0.087$ yielding $MSE = p(1-p)^2 + (1-p)p^2 \approx 0.079$.

micro-foundation for this commonly used model. The network extension shows further that, under sufficient connectivity, heterogeneous interaction weights wash out asymptotically, leaving a limiting intensity equation that does not depend on the network topology. Finally, in the moderate-probability regime, the MLE is consistent and asymptotically normal with asymptotic variance equal to the inverse Fisher information; in the rare-events regime, the Poisson likelihood applied to the aggregate count is the natural substitute, and the aggregation theory supplies the mapping from its parameters back to those of the underlying binary system.

Several directions remain open. On the theoretical side, a unified asymptotic theory in which both T and N grow simultaneously would constitute a valuable generalization. The extension of asymptotic normality for the MLE to settings where N grows with T is a natural complement. At the intersection of the two regimes, a composite likelihood that exploits both the individual binary structure and the aggregate Poisson approximation may yield efficiency gains relative to either likelihood alone. On the empirical side, the binary panel framework is well suited to modeling correlated binary decisions in large administrative datasets: voting records, loan defaults, and insurance claims. The network GAB model in particular provides a tractable vehicle for studying the role of connectivity in propagating financial or economic shocks.

APPENDIX A. PROOFS: EXISTENCE AND UNIQUENESS

Proof of Theorem 1. We use Meyn and Tweedie (2009, Theorem 12.0.1 (Markov–Krylov–Bogolyubov theorem)), which guarantees the existence of a stationary solution. Since the Markov chain $(\mathbf{p}_{t-1}, \dots, \mathbf{p}_{t-s}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-q})$ is defined on a compact space $X = [0, 1]^{Ns} \times \{0, 1\}^{Nq}$, the “bounded in probability on average” condition is satisfied. It only remains to check that the chain is weak Feller.

Let $C(X)$ be the class of bounded continuous functions from X to $\mathbb{R}$. To show that the chain is weak Feller we need to check that its transition probability kernel maps $C(X)$ to $C(X)$.

Take any bounded and continuous $f : X \rightarrow \mathbb{R}$. The transition kernel maps

$$\begin{aligned} & (Pf)(\mathbf{p}_{t-1}, \dots, \mathbf{p}_{t-s}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-q}) \\ &= \sum_{\mathbf{y} \in \{0,1\}^N} \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{1-y_i} f(\mathbf{p}, \mathbf{p}_{t-1}, \dots, \mathbf{p}_{t-s+1}, \mathbf{y}, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-q+1}), \\ & \mathbf{p} = (p_1, \dots, p_N) = \mathbf{g}(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}), \quad \mathbf{y} = (y_1, \dots, y_N). \end{aligned}$$

Since both f and $\mathbf{g}$ are continuous, their composition in (Pf) is also continuous. Function $\mathbf{g}$ is bounded since its image is $[0, 1]^N$, and so the composition (Pf) is also bounded. Thus, we are left with a weak Feller chain on a compact set, which by Meyn and Tweedie (2009, Theorem 12.0.1 (Markov–Krylov–Bogolyubov theorem)), has an invariant probability measure (i.e., stationary solution). $\square$

Proof of Theorem 3. Let $(\mathbf{y}_t, \mathbf{p}_t)_{t \in \mathbb{Z}}$ be a stationary solution given by Theorem 1 and let $\mathcal{F}_t$ be the corresponding σ -algebras. Extending the σ -algebras, if necessary, we assume that they support i.i.d. uniform random variables $u_{i,t}$ from (1).

Let $(\mathbf{y}'_t, \mathbf{p}'_t)_{t \geq t_0}$ be (possibly another) solution of (1)–(2) constructed by the same sequence $u_{i,t}$, but potentially different starting values $\mathbf{p}' := \{\mathbf{p}'_{t_0+s-1}, \dots, \mathbf{p}'_{t_0}\}$, $\mathbf{y}' := \{\mathbf{y}'_{t_0+q-1}, \dots, \mathbf{y}'_{t_0}\}$; we need another extension of σ -algebras to support these starting values. Observe that

$$\begin{aligned}
 \mathbb{E}_{t-1} |y_{i,t} - y'_{i,t}| &= \mathbb{E}_{t-1} |I(u_{i,t} \leq p_{i,t}) - I(u_{i,t} \leq p'_{i,t})| \\
 &= \mathbb{E}_{t-1} I(p'_{i,t} < u_{i,t} \leq p_{i,t} \text{ or } p_{i,t} < u_{i,t} \leq p'_{i,t}) \\
 &= \mathbb{E}_{t-1} I(\min(p_{i,t}, p'_{i,t}) < u_{i,t} \leq \max(p_{i,t}, p'_{i,t})) \\
 &= \max(p_{i,t}, p'_{i,t}) - \min(p_{i,t}, p'_{i,t}) = |p_{i,t} - p'_{i,t}|.
 \end{aligned}
 \tag{21}$$

Thus, assuming Assumption I holds,

$$\begin{aligned}
 \mathbb{E} |p_{i,t} - p'_{i,t}| &= \mathbb{E} |g_i(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}) - g_i(\{\mathbf{p}'_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}'_\tau\}_{\tau=t-1, \dots, t-q})| \\
 &\leq \mathbb{E} \left(\sum_{j=1}^N \sum_{\tau=1}^s |p_{j,t-\tau} - p'_{j,t-\tau}| \beta_{ij}^\tau + \sum_{j=1}^N \sum_{\tau=1}^q |y_{j,t-\tau} - y'_{j,t-\tau}| \alpha_{ij}^\tau \right) \\
 &= \sum_{j=1}^N \sum_{\tau=1}^s \mathbb{E} |p_{j,t-\tau} - p'_{j,t-\tau}| \beta_{ij}^\tau + \sum_{j=1}^N \sum_{\tau=1}^q \mathbb{E} |p_{j,t-\tau} - p'_{j,t-\tau}| \alpha_{ij}^\tau \\
 &= \sum_{j=1}^N \sum_{\tau=1}^{\max(s,q)} \mathbb{E} |p_{j,t-\tau} - p'_{j,t-\tau}| (\alpha_{ij}^\tau + \beta_{ij}^\tau),
 \end{aligned}
 \tag{22}$$

where we treat $\alpha_{ij}^\tau = 0$ for $\tau > q$ and $\beta_{ij}^\tau = 0$ for $\tau > s$.

Consider an $N \max(s, q)$ -dimensional vector

$$\eta_t = (\mathbb{E} |p_{1,t} - p'_{1,t}|, \dots, \mathbb{E} |p_{N,t} - p'_{N,t}|, \dots, \mathbb{E} |p_{N,t-\max(s,q)+1} - p'_{N,t-\max(s,q)+1}|)^\top.$$

Then Eq. (22) implies

$$\eta_t \leq \Phi \eta_{t-1} \leq \Phi^{t-t_0-\max(s,q)} \eta_{t_0+\max(s,q)}.
 \tag{23}$$

By Assumption I, all eigenvalues of Φ lie inside the unit circle. Thus, $\Phi^t \rightarrow 0$ as $t \rightarrow \infty$, so that $\mathbb{E}|p_{i,t} - p'_{i,t}| \rightarrow 0$ geometrically fast for all i as $t \rightarrow \infty$. Thus, we must have $\mathbf{p}_t - \mathbf{p}'_t \xrightarrow[t \rightarrow \infty]{p} 0$.

Similarly, if instead we assume that Assumption II holds,

$$\begin{aligned}
 (24) \quad \mathbb{E} \sum_{i=1}^N |p_{i,t} - p'_{i,t}| &= \mathbb{E} \sum_{i=1}^N |g_i(\{\mathbf{p}_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}) - g_i(\{\mathbf{p}'_\tau\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}'_\tau\}_{\tau=t-1, \dots, t-q})| \\
 &\leq K \mathbb{E} \left(\sum_{i=1}^N \sum_{\tau=1}^s |p_{i,t-\tau} - p'_{i,t-\tau}| + \sum_{i=1}^N \sum_{\tau=1}^q |y_{i,t-\tau} - y'_{i,t-\tau}| \right) \\
 &= K \sum_{\tau=1}^s \mathbb{E} \sum_{i=1}^N |p_{i,t-\tau} - p'_{i,t-\tau}| + K \sum_{\tau=1}^q \mathbb{E} \sum_{i=1}^N |p_{i,t-\tau} - p'_{i,t-\tau}|.
 \end{aligned}$$

Consider a $\max(s, q)$ -dimensional vector

$$\tilde{\eta}_t = \left(\mathbb{E} \sum_{i=1}^N |p_{i,t} - p'_{i,t}|, \dots, \mathbb{E} \sum_{i=1}^N |p_{i,t-\max(s,q)+1} - p'_{i,t-\max(s,q)+1}| \right)$$

and a $\max(s, q) \times \max(s, q)$ matrix $\tilde{\Phi}$, where

$$\tilde{\Phi}_{1i} = 2K, \quad i = 1, \dots, \min(s, q), \quad \tilde{\Phi}_{1i} = K, \quad i = \min(s, q) + 1, \dots, \max(s, q),$$

the bottom-left $(\max(s, q) - 1) \times (\max(s, q) - 1)$ block of $\tilde{\Phi}$ is an identity matrix, and the bottom-right $(\max(s, q) - 1) \times 1$ block is zero. Then the L_1 norm of the first row of $\tilde{\Phi}$ equals $K(s + q) < 1$ and, again, all eigenvalues of $\tilde{\Phi}$ lie inside the unit circle. Thus,

$$\tilde{\eta}_t \leq \tilde{\Phi} \tilde{\eta}_{t-1} \leq \tilde{\Phi}^{t-t_0-\max(s,q)} \tilde{\eta}_{t_0+\max(s,q)} \rightarrow 0 \quad \text{as } t \rightarrow \infty$$

and again $\mathbb{E}|p_{i,t} - p'_{i,t}| \rightarrow 0$ geometrically fast and $\mathbf{p}_t - \mathbf{p}'_t \xrightarrow[t \rightarrow \infty]{p} 0$.

Geometric decay of $\mathbb{E}|p_{i,t} - p'_{i,t}|$ and boundedness of $p_{i,t}, p'_{i,t}$ imply (16). The convergence in probability $\mathbf{p}_t - \mathbf{p}'_t \xrightarrow[t \rightarrow \infty]{p} 0$ implies that the Lévy-Prokhorov distance between the distributions of $(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-q}, \mathbf{p}_{t-1}, \dots, \mathbf{p}_{t-s})$ and $(\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-q}, \mathbf{p}'_{t-1}, \dots, \mathbf{p}'_{t-s})$ goes to 0 as $t \rightarrow \infty$. Hence, if $(\mathbf{y}'_t, \mathbf{p}'_t)$ is another stationary solution, then, by t -independence of the distribution, we should have the distributional match $(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-q}, \mathbf{p}_{t-1}, \dots, \mathbf{p}_{t-s}) \stackrel{d}{=} (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-q}, \mathbf{p}'_{t-1}, \dots, \mathbf{p}'_{t-s})$ for all t . Using the Markov chain representation from the proof of Theorem 1, this implies that the distributions of any two stationary solutions coincide. $\square$

The following corollary will be useful for estimation theorems.

Corollary 13. *Suppose that $\mathbf{g}(\cdot)$ satisfies either Assumption I or II, and let $\{(\mathbf{y}_t, \mathbf{p}_t)\}_{t \in \mathbb{Z}}$ denote the unique strictly stationary solution of (1)–(2). Define the augmented Markov state*

$$Z_t := (\mathbf{p}_t, \dots, \mathbf{p}_{t-s+1}, \mathbf{y}_t, \dots, \mathbf{y}_{t-q+1})$$

and let π denote its stationary distribution. Then, for every continuous function h on the state space of Z_t ,

$$(25) \quad \frac{1}{T} \sum_{t=1}^T h(Z_t) \xrightarrow[T \rightarrow \infty]{p} \mathbb{E}_\pi[h(Z_0)].$$

Proof. Let P denote the transition kernel of the Markov chain $\{Z_t\}$. By Theorem 3, for every initial state z ,

$$(26) \quad P^n(z, \cdot) \xrightarrow[n \rightarrow \infty]{} \pi.$$

Since the state space is compact, every continuous function h is bounded. Hence Eq. (26) implies

$$P^n h(z) := \int h(z') P^n(z, dz') \longrightarrow \int h(z') \pi(dz') =: \pi(h)$$

for every z .

Since the process is stationary,

$$\text{Cov}_\pi(h(Z_0), h(Z_n)) = \int h(z) P^n h(z) \pi(dz) - \pi(h)^2.$$

Moreover,

$$|h(z) P^n h(z)| \leq \|h\|_\infty^2.$$

Therefore, by dominated convergence,

$$(27) \quad \text{Cov}_\pi(h(Z_0), h(Z_n)) \rightarrow 0.$$

Let

$$\gamma_n := \text{Cov}_\pi(h(Z_0), h(Z_n)).$$

By stationarity,

$$\mathbb{V} \left(\frac{1}{T} \sum_{t=1}^T h(Z_t) \right) = \frac{\gamma_0}{T} + \frac{2}{T^2} \sum_{n=1}^{T-1} (T-n) \gamma_n = \frac{\gamma_0}{T} + \frac{2}{T} \sum_{n=1}^{T-1} \left(1 - \frac{n}{T}\right) \gamma_n \rightarrow 0.$$

Since, by stationarity,

$$\mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T h(Z_t) \right] = \pi(h),$$

we conclude that

$$\frac{1}{T} \sum_{t=1}^T h(Z_t) \xrightarrow[T \rightarrow \infty]{p} \pi(h). \quad \square$$

APPENDIX B. PROOFS: AGGREGATION

Lemma 14. *Consider a sequence of random probabilities $q_i, i = 1, \dots, N$, such that*

$$(28) \quad \sum_{i=1}^N q_i \xrightarrow[N \rightarrow \infty]{d} \xi, \quad \max_{i=1, \dots, N} q_i \xrightarrow[N \rightarrow \infty]{p} 0,$$

where ξ is some random variable. Given $q_i, i = 1, \dots, N$, consider a sequence of independent Bernoulli random variables $y_i \sim B(q_i)$. As $N \rightarrow \infty$, $\left[\sum_{i=1}^N q_i, \sum_{i=1}^N y_i \right]$ converges in distribution to $[\xi, Y]$, where the random variable Y is conditionally Poisson distributed with random intensity ξ .

Proof. Letting $S_N := \sum_{i=1}^N y_i$, the conditional (on the realization $\{q_i\}_i$) probability generating function of S_N satisfies

$$(29) \quad G_N(z) := \mathbb{E}(z^{S_N} \mid q_1, \dots, q_N) = \prod_{i=1}^N (q_i z + (1 - q_i)) = \prod_{i=1}^N (1 + q_i(z - 1))$$

so that

$$\begin{aligned} \log(G_N(z)) &= \sum_{i=1}^N \log(1 + q_i(z - 1)) = \sum_{i=1}^N \left(q_i(z - 1) - \frac{1}{2} q_i^2 (z - 1)^2 + o(q_i^2) \right) \\ &= (z - 1) \sum_{i=1}^N q_i - \frac{1}{2} (z - 1)^2 \sum_{i=1}^N q_i^2 + o\left(\sum_{i=1}^N q_i^2 \right). \end{aligned}$$

Since $\sum_{i=1}^N q_i^2 \leq (\max_i q_i) \sum_{i=1}^N q_i \xrightarrow[N \rightarrow \infty]{p} 0$, as $N \rightarrow \infty$, $\log(G_N(z)) \xrightarrow[N \rightarrow \infty]{d} (z - 1)\xi$, so that $G_N(z)$ converges in distribution to the generating function of the Poisson distribution, $\exp((z - 1)\xi)$. Thus, $S_N \xrightarrow[N \rightarrow \infty]{d} Y$ with $Y \sim \text{Pois}(\xi)$ jointly with $\sum_{i=1}^N q_i \xrightarrow[N \rightarrow \infty]{d} \xi$. $\square$

Proof of Theorem 4. We show by induction that, for all $T = 0, 1, \dots$, $\left[\sum_{i=1}^N p_{i,t}, \sum_{i=1}^N y_{i,t} \right]_{t=0, \dots, T}$ converges both in distribution and in expectation to a Poisson autoregression $[\lambda_t, X_t]_{t=0, \dots, T}$ and, additionally, $\max_i p_{i,T} \xrightarrow[N \rightarrow \infty]{p} 0$.

At time $t = 0$ (base of induction), the process satisfies conditions of Lemma 14, thus,

$$(30) \quad \left[\sum_{i=1}^N p_{i,0}, \sum_{i=1}^N y_{i,0} \right] \xrightarrow[N \rightarrow \infty]{d} [\lambda_0, X_0], \text{ with } X_0 \sim \text{Pois}(\lambda_0).$$

By (18), $\max_i p_{i,0} \xrightarrow[N \rightarrow \infty]{p} 0$ and $\mathbb{E} \sum_{i=1}^N p_{i,0} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_0$. Applying conditional expectation,

$$\mathbb{E} \sum_{i=1}^N y_{i,0} = \mathbb{E} \sum_{i=1}^N p_{i,0} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_0 = \mathbb{E} [\mathbb{E}(X_0 \mid \lambda_0)] = \mathbb{E} X_0.$$

Let us show that $T = 1$ follows from $T = 0$. The argument for the general induction step is the same.

Let us add $\sum_{i=1}^N p_{i,1}$ to the joint convergence. By Markov inequality for any $\varepsilon > 0$,

$$(31) \quad \text{Prob} \left(\frac{1}{N^\kappa} \sum_{j=1}^N y_{j,0} > \varepsilon \right) \leq \frac{1}{\varepsilon} \mathbb{E} \frac{1}{N^\kappa} \sum_{j=1}^N y_{j,0} = \frac{1}{N^\kappa \varepsilon} \sum_{j=1}^N \mathbb{E} p_{j,0} \xrightarrow[N \rightarrow \infty]{} 0.$$

Thus, summing Eq. (8) with respect to i for $t = 1$,

$$\sum_{i=1}^N p_{i,1} = \frac{1}{N} \sum_{i=1}^N c_i + \frac{1}{N^\kappa} \sum_{i=1}^N a_i y_{i,0} + \beta \sum_{i=1}^N p_{i,0} + \frac{1}{N} \sum_{j=1}^N y_{j,0} \sum_{i=1}^N \gamma_i \xrightarrow[N \rightarrow \infty]{d} \bar{c} + \beta \lambda_0 + \bar{\gamma} X_0 = \lambda_1,$$

and we get

$$\left[\sum_{i=1}^N p_{i,0}, \sum_{i=1}^N y_{i,0}, \sum_{i=1}^N p_{i,1} \right] \xrightarrow[N \rightarrow \infty]{d} [\lambda_0, X_0, \lambda_1].$$

Let us show that the conditions (28) are valid for $q_i = p_{i,1}$. The first one is already proved. To show the second one, note that by Eq. (8),

$$p_{i,1} = \frac{c_i}{N} + \frac{a_i}{N^\kappa} y_{i,0} + \beta p_{i,0} + \gamma_i \frac{1}{N} \sum_{j=1}^N y_{j,0}.$$

Thus, since by assumptions, $\max_i p_{i,0} \rightarrow 0$ and $0 \leq \gamma_i \leq 1$, and, by (31) with $\kappa = 1$, $\frac{1}{N} \sum_{j=1}^N y_{j,0} \xrightarrow[N \rightarrow \infty]{p} 0$,

$$0 \leq \max_i p_{i,1} \leq \frac{C}{N} + \frac{C}{N^\kappa} + \beta \max_i p_{i,0} + \frac{1}{N} \sum_{j=1}^N y_{j,0} \xrightarrow[N \rightarrow \infty]{} 0.$$

Direct application of Lemma 14 gives

$$(32) \quad \left[\sum_{i=1}^N p_{i,1}, \sum_{i=1}^N y_{i,1} \right] \xrightarrow[N \rightarrow \infty]{d} [\lambda_1, X_1], \text{ with } X_1 \sim \text{Pois}(\lambda_1).$$

We need to combine the above with distributional convergence (30), which requires an extension of Lemma 14. By adding $\{p_{i,0}, y_{i,0}\}_{i=1}^N$ to the conditioning in (29), and repeating the

same proof, the convergence in (32) extends to conditional convergence. Hence, we arrive at

$$\left[\sum_{i=1}^N p_{i,0}, \sum_{i=1}^N y_{i,0}, \sum_{i=1}^N p_{i,1}, \sum_{i=1}^N y_{i,1} \right] \xrightarrow[N \rightarrow \infty]{d} [\lambda_0, X_0, \lambda_1, X_1].$$

It remains to prove convergence in expectations. For $t = 0$ we have already shown it. For $t = 1$,

$$\begin{aligned} \sum_{i=1}^N \mathbb{E} p_{i,1} &= \frac{1}{N} \sum_{i=1}^N c_i + \frac{1}{N^\kappa} \sum_{i=1}^N a_i \mathbb{E} y_{i,0} + \beta \sum_{i=1}^N \mathbb{E} p_{i,0} + \frac{1}{N} \sum_{j=1}^N \mathbb{E} y_{j,0} \sum_{i=1}^N \gamma_i \\ &= \frac{1}{N} \sum_{i=1}^N c_i + \frac{1}{N^\kappa} \sum_{i=1}^N a_i \mathbb{E} p_{i,0} + \beta \sum_{i=1}^N \mathbb{E} p_{i,0} + \frac{1}{N} \sum_{j=1}^N \mathbb{E} y_{j,0} \sum_{i=1}^N \gamma_i \\ &\xrightarrow[N \rightarrow \infty]{} \bar{c} + \beta \mathbb{E} \lambda_0 + \bar{\gamma} \mathbb{E} X_0 = \mathbb{E} \lambda_1, \\ \sum_{i=1}^N \mathbb{E} y_{i,1} &= \sum_{i=1}^N \mathbb{E} p_{i,1} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_1 = \mathbb{E} [\mathbb{E}(X_1 \mid \lambda_1)] = \mathbb{E} X_1. \end{aligned}$$

Repeating the same steps for general T , we arrive at the desired joint convergence. $\square$

Proof of Proposition 8. Consider a stationary distribution and let us denote it as $\mathbf{p}^* = (p_1^*, \dots, p_N^*)^\top$ with corresponding realizations $\mathbf{y}^* = (y_1^*, \dots, y_N^*)^\top$. Let us start with the mean of $\sum_{i=1}^N p_{i,t}^*$. Plugging $\beta_i = \beta$, $\omega_i = \frac{c_i}{N}$, $\alpha_i = \frac{a_i}{N^\kappa}$ in Eq. (11), we get

$$\sum_{i=1}^N \mu_i = \frac{\frac{1}{N} \sum_{i=1}^N \frac{c_i}{1 - \beta - a_i/N^\kappa}}{1 - \frac{1}{N} \sum_{i=1}^N \frac{\gamma_i}{1 - \beta - a_i/N^\kappa}}.$$

Taylor expanding, we get

$$\begin{aligned} \gamma_i (1 - \beta - a_i/N^\kappa)^{-1} &\approx \gamma_i (1 - \beta)^{-1} \left(1 + \frac{a_i}{(1 - \beta)N^\kappa} \right), \\ c_i (1 - \beta - a_i/N^\kappa)^{-1} &\approx c_i (1 - \beta)^{-1} \left(1 + \frac{a_i}{(1 - \beta)N^\kappa} \right) \end{aligned}$$

and

$$(33) \quad \sum_{i=1}^N \mu_i = \frac{\frac{1}{N} \sum_{i=1}^N c_i}{1 - \beta - \frac{1}{N} \sum_{i=1}^N \gamma_i} + O(N^{-\kappa}) \xrightarrow[N \rightarrow \infty]{} \frac{\bar{c}}{1 - \beta - \bar{\gamma}}.$$

Similarly, Eq. (12) implies, uniformly in i ,

$$(34) \quad \mu_i = \frac{c_i}{N(1 - \beta - a_i N^{-\kappa})} + \frac{\gamma_i}{N(1 - \beta - a_i N^{-\kappa})} \sum_{j=1}^N \mu_j = O(N^{-1}).$$

We begin by checking that $\max_{i=1, \dots, N} p_{i,t}^* \xrightarrow[N \rightarrow \infty]{p} 0$. First, note that

$$\begin{aligned} 0 &\leq \sum_i p_{i,t}^* = \frac{1}{N} \sum_i c_i + \frac{1}{N^\kappa} \sum_i a_i y_{i,t-1}^* + \beta \sum_i p_{i,t-1}^* + \frac{1}{N} \sum_i \gamma_i \sum_i y_{i,t-1}^* \\ &\leq C + \beta \sum_i p_{i,t-1}^* + \left(\frac{1}{N} \sum_i \gamma_i + \frac{C}{N^\kappa} \right) \sum_i y_{i,t-1}^*. \end{aligned}$$

Thus, by Minkowski's inequality (or Cauchy–Schwarz inequality), letting $\tilde{\gamma}_N := \frac{1}{N} \sum_i \gamma_i + \frac{C}{N^\kappa}$,

$$\begin{aligned} (35) \quad \sqrt{\mathbb{E} \left(\sum_i p_{i,t}^* \right)^2} &\leq C + \sqrt{\mathbb{E} \left(\beta \sum_i p_{i,t-1}^* + \tilde{\gamma}_N \sum_i y_{i,t-1}^* \right)^2} \\ &= C + \sqrt{\mathbb{E} \left((\beta + \tilde{\gamma}_N) \sum_i p_{i,t-1}^* + \tilde{\gamma}_N \sum_i (y_{i,t-1}^* - p_{i,t-1}^*) \right)^2} \\ &= C + \sqrt{(\beta + \tilde{\gamma}_N)^2 \mathbb{E} \left(\sum_i p_{i,t-1}^* \right)^2 + \tilde{\gamma}_N^2 \mathbb{E} \left(\sum_i p_{i,t-1}^* - \sum_i (p_{i,t-1}^*)^2 \right)} \\ &\leq C + (\beta + \tilde{\gamma}_N) \sqrt{\mathbb{E} \left(\sum_i p_{i,t}^* \right)^2} + \tilde{\gamma}_N \sqrt{\sum_i \mu_i} \end{aligned}$$

where the third line follows from $\mathbb{E}_{t-2}(y_{i,t-1}^* - p_{i,t-1}^*) = \mathbb{E}_{t-2}(y_{i,t-1}^* - p_{i,t-1}^*)(y_{j,t-1}^* - p_{j,t-1}^*) = 0$ for any $i \neq j$, and the last line uses stationarity. Since $\beta + \gamma_i < \bar{C} < 1$ for all i , $1 - \beta - \tilde{\gamma}_N > 1 - \bar{C} - \frac{C}{N^\kappa} > 0$ for all N large enough. Thus, Eq. (35) combined with the bound (33), imply that

$$(36) \quad \sup_N \mathbb{E} \left(\sum_i p_{i,t}^* \right)^2 < \infty.$$

Second, again using $\mathbb{E}_{t-1}(y_{i,t}^* - p_{i,t}^*) = \mathbb{E}_{t-1}(y_{i,t}^* - p_{i,t}^*)(y_{j,t}^* - p_{j,t}^*) = 0$,

(37)

$$\begin{aligned} \mathbb{E} \left(\sum_i y_{i,t}^* \right)^2 &= \mathbb{E} \left(\sum_i (y_{i,t}^* - p_{i,t}^*) + \sum_i p_{i,t}^* \right)^2 = \mathbb{E} \left(\sum_i (y_{i,t}^* - p_{i,t}^*) \right)^2 + \mathbb{E} \left(\sum_i p_{i,t}^* \right)^2 \\ &= \mathbb{E} \sum_i (p_{i,t}^* - (p_{i,t}^*)^2) + \mathbb{E} \left(\sum_i p_{i,t}^* \right)^2 \leq \sum_i \mu_i + \mathbb{E} \left(\sum_i p_{i,t}^* \right)^2 = O(1). \end{aligned}$$

Third, again using stationarity and Minkowski's inequality for the dynamic of $p_{i,t}^*$,

$$\sqrt{\mathbb{E} (p_{i,t}^*)^2} \leq \frac{C}{N} + \frac{C}{N^\kappa} \sqrt{\mu_i} + \beta \sqrt{\mathbb{E} (p_{i,t}^*)^2} + \frac{1}{N} \sqrt{\mathbb{E} \left(\sum_i y_{i,t}^* \right)^2}.$$

Combining the above bound with Eq. (34), (37), we get, uniformly in i ,

$$(1 - \beta) \sqrt{\mathbb{E} (p_{i,t}^*)^2} \leq \frac{C}{N} + \frac{C}{N^\kappa} \sqrt{\mu_i} + \frac{1}{N} \sqrt{\mathbb{E} \left(\sum_i y_{i,t}^* \right)^2} = O(N^{-1} + N^{-\kappa-1/2})$$

and

$$\mathbb{E} (p_{i,t}^*)^2 = O(N^{-2} + N^{-2\kappa-1}), \quad \sum_{i=1}^N \mathbb{E} (p_{i,t}^*)^2 = O(N^{-1} + N^{-\kappa}).$$

Thus, for any $\varepsilon > 0$,

$$\begin{aligned} \text{Prob} \left(\max_{i=1,\dots,N} p_{i,t}^* > \varepsilon \right) &\leq \sum_{i=1}^N \text{Prob}(p_{i,t}^* > \varepsilon) = \sum_{i=1}^N \text{Prob}((p_{i,t}^*)^2 > \varepsilon^2) \\ &\leq \frac{1}{\varepsilon^2} \sum_{i=1}^N \mathbb{E}(p_{i,t}^*)^2 = \frac{1}{\varepsilon^2} O(N^{-1} + N^{-\kappa}) \xrightarrow{N \rightarrow \infty} 0. \end{aligned}$$

It remains to check convergence of $\sum_i p_{i,t}^*$. Consider its probability generating function, where $\bar{\gamma}_N := \frac{1}{N} \sum_i \gamma_i$,

$$\begin{aligned} G_N(z) &:= \mathbb{E} z^{\sum_i p_{i,t}^*} = \mathbb{E} z^{\frac{1}{N} \sum_i c_i + \sum_i \left(\frac{a_i}{N^\kappa} + \bar{\gamma}_N \right) y_{i,t-1}^* + \beta \sum_i p_{i,t-1}^*} \\ (38) \quad &= \mathbb{E} \left\{ z^{\frac{1}{N} \sum_i c_i + \beta \sum_i p_{i,t-1}^*} \prod_{i=1}^N \left[1 - p_{i,t-1}^* \left(1 - z^{\frac{a_i}{N^\kappa} + \bar{\gamma}_N} \right) \right] \right\}. \end{aligned}$$

The product in (38) can be analyzed by means of

$$\begin{aligned}
\log \prod_{i=1}^N \left[1 - p_{i,t-1}^* \left(1 - z^{\frac{a_i}{N^k} + \bar{\gamma}_N} \right) \right] &= \sum_i \log \left[1 - p_{i,t-1}^* \left(1 - z^{\frac{a_i}{N^k} + \bar{\gamma}_N} \right) \right] \\
&= - \sum_i p_{i,t-1}^* \left(1 - z^{\frac{a_i}{N^k} + \bar{\gamma}_N} \right) + o \left(\sum_i p_{i,t-1}^* \left(1 - z^{\frac{a_i}{N^k} + \bar{\gamma}_N} \right) \right) \\
&= - \sum_i p_{i,t-1}^* (1 - z^{\bar{\gamma}_N}) + o \left(\sum_i p_{i,t-1}^* \right) + o \left(\sum_i p_{i,t-1}^* \left(1 - z^{\frac{a_i}{N^k} + \bar{\gamma}_N} \right) \right).
\end{aligned}$$

For each fixed $z \in (0, 1]$,

$$0 \leq \sum_i p_{i,t-1}^* \left(1 - z^{\frac{a_i}{N^k} + \bar{\gamma}_N} \right) \leq \tilde{C} \sum_i p_{i,t-1}^* = O_p(1),$$

where the boundedness in probability follows from (36) by Markov inequality. Therefore, for each fixed $z \in (0, 1]$ we can rewrite (38) as

(39)

$$\begin{aligned}
G_N(z) &= \mathbb{E} z^{\frac{\sum_i c_i}{N} + \beta \sum_i p_{i,t-1}^*} \exp \left[- \sum_i p_{i,t-1}^* (1 - z^{\bar{\gamma}_N}) + o(1) \right] \\
&= z^{\frac{\sum_i c_i}{N}} \mathbb{E} \left[z^\beta \exp\{z^{\bar{\gamma}_N} - 1\} \right]^{\sum_i p_{i,t-1}^*} (1 + o(1)) = z^{\frac{\sum_i c_i}{N}} G_N(z^\beta \exp\{z^{\bar{\gamma}_N} - 1\}) + o(1).
\end{aligned}$$

Uniform integrability (36) implies tightness and, thus, that $\sum_i p_{i,t}^*$ has a convergent subsequence. Denote the limit of this subsequence by λ_0 and let $G_z = \mathbb{E} z^{\lambda_0}$. For this subsequence we must have, for each $z \in (0, 1]$, $G_N(z) \xrightarrow{N \rightarrow \infty} G(z)$. Moreover, $z^{\frac{\sum_i c_i}{N}} \xrightarrow{N \rightarrow \infty} z^{\bar{c}}$ and $z^\beta \exp\{z^{\bar{\gamma}_N} - 1\} \xrightarrow{N \rightarrow \infty} z^\beta \exp\{z^{\bar{\gamma}} - 1\}$. Thus, taking the limit in (39) and using the continuity and monotonicity of generating functions, we get for any fixed $z \in (0, 1)$

$$(40) \quad G(z) = z^{\bar{c}} G(z^\beta \exp\{z^{\bar{\gamma}} - 1\}).$$

We claim that there is a unique solution to (40). Fix $z_0 \in (0, 1)$ and let $z_{n+1} := z_n^\beta \exp\{z_n^{\bar{\gamma}} - 1\}$ for $n = 0, 1, \dots$. Then $z_n \in (0, 1)$ and, using the bound $1 - \exp -z \leq z$ for $z \geq 0$,

$$\begin{aligned}
(41) \quad 0 \leq -\log z_{n+1} &= -\beta \log z_n + 1 - z_n^{\bar{\gamma}} = -\beta \log z_n + 1 - \exp\{\bar{\gamma} \log z_n\} \\
&\leq -(\beta + \bar{\gamma}) \log z_n \leq -(\beta + \bar{\gamma})^n \log z_0 \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

Thus, $z_{n+1} \xrightarrow{n \rightarrow \infty} 1$, so that $G(z_n) \xrightarrow{n \rightarrow \infty} G(1) \equiv 1$. Iterating (40), we get

$$G(z_0) = z_0^{\bar{c}} G(z_1) = z_0^{\bar{c}} z_1^{\bar{c}} G(z_2) = \prod_{i=0}^{\infty} z_i^{\bar{c}} = \exp \left\{ \bar{c} \sum_{i=0}^{\infty} \log z_i \right\},$$

and the sum $\sum_{i=0}^{\infty} \log z_i$ is finite, since by (41) it is bounded by the sum of a geometric sequence. Thus, the limiting probability generating function is unique. Probability generating function uniquely determines random variable. Therefore, the limit λ_0 is unique, and $\sum_{i=1}^N p_{i,t}^* \xrightarrow[N \rightarrow \infty]{d} \lambda_0$. Combined with uniform integrability (36), the distributional convergence also implies $\mathbb{E} \sum_{i=1}^N p_{i,t}^* \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_0$. $\square$

Proof of Theorem 9. We show by induction that, for all $T \geq -s + 1$, $[\sum_{i=1}^N p_{i,t}, \sum_{i=1}^N y_{i,t}]_{t=-s+1, \dots, T}$ converges both in distribution and in expectation to a nonlinear Poisson autoregression $[\lambda_t, X_t]_{t=-s+1, \dots, T}$ and, additionally, $\max_i p_{i,T} \xrightarrow[N \rightarrow \infty]{p} 0$. The proof follows the same steps as the proof of Theorem 4, with the only difference being that we now have multiple nonlinear terms. At $t = -s + 1, \dots, 0$, we are exactly in the setting of Lemma 14, so the desired (joint) convergence is satisfied. Moreover, applying conditional expectation, for $\tau = -s + 1, \dots, 0$,

$$\mathbb{E} \sum_{i=1}^N y_{i,\tau} = \mathbb{E} \sum_{i=1}^N p_{i,\tau} \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \lambda_\tau = \mathbb{E} [\mathbb{E}(X_\tau \mid \lambda_{-s+1}, X_{-s+1}, \dots, X_{\tau-1}, \lambda_\tau)] = \mathbb{E} X_\tau.$$

Let us show that $T = 1$ follows from $T = 0$. The argument for the general induction step is the same.

Let us show that $\max_i p_{i,1} \xrightarrow[N \rightarrow \infty]{p} 0$. By Markov inequality,

$$\frac{1}{N^m} \sum_{j=1}^N y_{j,\tau} \xrightarrow[N \rightarrow \infty]{p} 0, \quad \frac{1}{N^m} \sum_{j=1}^N p_{j,\tau} \xrightarrow[N \rightarrow \infty]{p} 0 \quad \tau = 0, \dots, -s + 1, m > 0.$$

Because

$$|f_{\alpha,i}(y_{i,t-1}, \dots, y_{i,t-s})| \leq A \sum_{\tau=1}^s y_{i,t-\tau} \leq sA, \quad \sum_{i=1}^N |f_{\alpha,i}(y_{i,t-1}, \dots, y_{i,t-s})| \leq A \sum_{\tau=1}^s \sum_{i=1}^N y_{i,t-\tau},$$

we obtain, uniformly in i ,

$$(42) \quad \frac{1}{N^\kappa} f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) \xrightarrow[N \rightarrow \infty]{p} 0, \quad \frac{1}{N^\kappa} \sum_{i=1}^N f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) \xrightarrow[N \rightarrow \infty]{p} 0.$$

Since $f_{\gamma,i}(0, \dots, 0) = 0$ and $\nabla f_{\gamma,i}$ is uniformly bounded,

$$(43) \quad \left| f_{\gamma,i} \left(\frac{1}{N} \sum_{j=1}^N y_{j,t-1}, \dots, \frac{1}{N} \sum_{j=1}^N y_{j,t-s}, \frac{1}{N} \sum_{j=1}^N p_{j,t-1}, \dots, \frac{1}{N} \sum_{j=1}^N p_{j,t-s} \right) \right| \\ \leq M_1 \left(\frac{1}{N} \sum_{j=1}^N y_{j,t-1} + \dots + \frac{1}{N} \sum_{j=1}^N y_{j,t-s} + \frac{1}{N} \sum_{j=1}^N p_{j,t-1} + \dots + \frac{1}{N} \sum_{j=1}^N p_{j,t-s} \right) \xrightarrow[N \rightarrow \infty]{p} 0.$$

Finally, because f_γ is continuous,

$$f_\gamma \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \xrightarrow[N \rightarrow \infty]{d} f_\gamma(X_0, \dots, X_{-s+1}, \lambda_0, \dots, \lambda_{-s+1}),$$

so that combining with $\|\gamma_i\|_1 < C$, we get, uniformly in i ,

$$(44) \quad \left| \frac{1}{N} \gamma_i^\top f_\gamma \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \right| \\ \leq \frac{C}{N} \left\| f_\gamma \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \right\|_1 \xrightarrow[N \rightarrow \infty]{p} 0.$$

Combining Eq. (42), (43), (44), and the assumption on the initial conditions $\max_i p_{i,\tau} \rightarrow 0$, $\tau = 0, \dots, -s+1$, we deduce

$$0 \leq \max_i p_{i,1} \xrightarrow[N \rightarrow \infty]{p} 0.$$

Next, let us add $\sum_{i=1}^N p_{i,1}$ to the joint convergence. Summing Eq. (19) with respect to i for $t = 1$, and using Eq. 42 and the fact that

$$\sum_{i=1}^N \|\nabla^2 f_{\gamma,i}\| \left(\frac{1}{N} \sum_{j=1}^N y_{j,0} + \dots + \frac{1}{N} \sum_{j=1}^N y_{j,-s+1} + \frac{1}{N} \sum_{j=1}^N p_{j,0} + \dots + \frac{1}{N} \sum_{j=1}^N p_{j,-s+1} \right)^2 \\ \leq M_2 \left(\sum_{j=1}^N y_{j,0} + \dots + \sum_{j=1}^N y_{j,-s+1} + \sum_{j=1}^N p_{j,0} + \dots + \sum_{j=1}^N p_{j,-s+1} \right) \\ \cdot \left(\frac{1}{N} \sum_{j=1}^N y_{j,0} + \dots + \frac{1}{N} \sum_{j=1}^N y_{j,-s+1} + \frac{1}{N} \sum_{j=1}^N p_{j,0} + \dots + \frac{1}{N} \sum_{j=1}^N p_{j,-s+1} \right) \xrightarrow[N \rightarrow \infty]{p} 0,$$

we obtain

$$\sum_{i=1}^N p_{i,1} = \frac{1}{N} \sum_{i=1}^N c_i + \frac{1}{N\kappa} \sum_{i=1}^N f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) + \sum_{\tau=1}^s \beta_\tau \sum_{i=1}^N p_{i,1-\tau} \\ + \sum_{i=1}^N f_{\gamma,i} \left(\frac{1}{N} \sum_{j=1}^N y_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N y_{j,-s+1}, \frac{1}{N} \sum_{j=1}^N p_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N p_{j,-s+1} \right) \\ + \left(\frac{1}{N} \sum_{i=1}^N \gamma_i^\top \right) f_\gamma \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \\ \xrightarrow[N \rightarrow \infty]{d} \bar{c} + \sum_{\tau=1}^s \beta_\tau \lambda_{1-\tau} + (X_0, \dots, X_{-s+1}, \lambda_0, \dots, \lambda_{-s+1}) \bar{\mathbf{f}} \\ + \bar{\gamma}^\top f_\gamma(X_0, \dots, X_{-s+1}, \lambda_0, \dots, \lambda_{-s+1}) = \lambda_1$$

and

$$\left[\sum_{i=1}^N p_{i,-s+1}, \sum_{i=1}^N y_{i,-s+1}, \dots, \sum_{i=1}^N p_{i,0}, \sum_{i=1}^N y_{i,0}, \sum_{i=1}^N p_{i,1} \right] \xrightarrow[N \rightarrow \infty]{d} [\lambda_{-s+1}, X_{-s+1}, \dots, \lambda_0, X_0, \lambda_1].$$

Direct application of Lemma 14 with $q_i = p_{i,1}$ gives

$$(45) \quad \left[\sum_{i=1}^N p_{i,1}, \sum_{i=1}^N y_{i,1} \right] \xrightarrow[N \rightarrow \infty]{d} [\lambda_1, X_1], \text{ with } X_1 \sim \text{Pois}(\lambda_1).$$

By adding $\{p_{i,t}, y_{i,t}\}_{i=1,\dots,N, t=-s+1,\dots,0}$ to the conditioning in (29), and repeating the same proof, the convergence in (45) extends to conditional convergence. Hence, we arrive at

$$\left[\sum_{i=1}^N p_{i,t}, \sum_{i=1}^N y_{i,t} \right]_{t=-s+1,\dots,1} \xrightarrow[N \rightarrow \infty]{d} [\lambda_t, X_t]_{t=-s+1,\dots,1}.$$

It remains to prove convergence in expectations. For $t \leq 0$ we have already shown it. Let us separately analyze all nonlinear terms in $t = 1$ dynamics. First,

$$(46) \quad 0 \leq \frac{1}{N^\kappa} \sum_{i=1}^N \mathbb{E} f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) \leq \frac{A}{N^\kappa} \sum_{\tau=1}^s \sum_{i=1}^N \mathbb{E} y_{i,1-\tau} \xrightarrow[N \rightarrow \infty]{} 0.$$

Next, let $f_\gamma = (f_\gamma^{[1]}, \dots, f_\gamma^{[k]})^\top$ and consider $r = 1, \dots, k$,

$$0 \leq f_\gamma^{[r]} \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \leq L_\gamma \sum_{\tau=1}^s \left(\sum_{j=1}^N y_{j,1-\tau} + \sum_{j=1}^N p_{j,1-\tau} \right)$$

Each term on the right converges in distribution and in expectation to its nonnegative limit and is therefore uniformly integrable (for nonnegative random variables, convergence in distribution together with convergence of first moments implies uniform integrability).

Therefore, $f_\gamma^{[r]} \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right)$ is also uniformly integrable.

Combined with convergence in distribution, this gives convergence of expectations:

$$(47) \quad \mathbb{E} f_\gamma^{[r]} \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \xrightarrow[N \rightarrow \infty]{} \mathbb{E} f_\gamma^{[r]}(X_0, \dots, X_{-s+1}, \lambda_0, \dots, \lambda_{-s+1}).$$

It remains to consider $f_{\gamma,i}$. A second-order Taylor expansion around zero gives

$$\begin{aligned} & \left| \sum_{i=1}^N \mathbb{E} f_{\gamma,i} \left(\frac{1}{N} \sum_{j=1}^N y_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N y_{j,-s+1}, \frac{1}{N} \sum_{j=1}^N p_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N p_{j,-s+1} \right) \right. \\ & \quad \left. - \mathbb{E} \left[\left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \frac{1}{N} \sum_{i=1}^N \nabla f_{\gamma,i}(0, \dots, 0) \right] \right| \\ & \leq \frac{M_2}{2N} \mathbb{E} \left[\left(\sum_{j=1}^N y_{j,0} \right)^2 + \dots + \left(\sum_{j=1}^N y_{j,-s+1} \right)^2 + \left(\sum_{j=1}^N p_{j,0} \right)^2 + \dots + \left(\sum_{j=1}^N p_{j,-s+1} \right)^2 \right]. \end{aligned}$$

Since each sum over j is bounded by N , for every $\varepsilon > 0$,

$$\begin{aligned} \frac{1}{N} \mathbb{E} \left(\sum_{i=1}^N y_{i,\tau} \right)^2 & \leq \varepsilon \mathbb{E} \sum_{i=1}^N y_{i,\tau} + \mathbb{E} \left[\sum_{i=1}^N y_{i,\tau} I \left(\sum_{i=1}^N y_{i,\tau} > \varepsilon N \right) \right], \\ \frac{1}{N} \mathbb{E} \left(\sum_{i=1}^N p_{i,\tau} \right)^2 & \leq \varepsilon \mathbb{E} \sum_{i=1}^N p_{i,\tau} + \mathbb{E} \left[\sum_{i=1}^N p_{i,\tau} I \left(\sum_{i=1}^N p_{i,\tau} > \varepsilon N \right) \right]. \end{aligned}$$

The second terms converge to zero as $N \rightarrow \infty$ by uniform integrability, while the first terms are bounded by a constant times ε . Thus, as $N \rightarrow \infty$ followed by $\varepsilon \rightarrow 0$, the second-order Taylor remainder goes to 0 and we obtain

$$\begin{aligned} (48) \quad & \sum_{i=1}^N \mathbb{E} f_{\gamma,i} \left(\frac{1}{N} \sum_{j=1}^N y_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N y_{j,-s+1}, \frac{1}{N} \sum_{j=1}^N p_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N p_{j,-s+1} \right) \\ & \xrightarrow{N \rightarrow \infty} \mathbb{E}(X_0, \dots, X_{-s+1}, \lambda_0, \dots, \lambda_{-s+1}) \bar{\mathbf{f}}. \end{aligned}$$

Summing up, Eq. (46), (47), (48) imply

$$\begin{aligned} \sum_{i=1}^N \mathbb{E} p_{i,1} & = \frac{1}{N} \sum_{i=1}^N c_i + \frac{1}{N\kappa} \sum_{i=1}^N \mathbb{E} f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) + \sum_{\tau=1}^s \beta_\tau \sum_{i=1}^N \mathbb{E} p_{i,1-\tau} \\ & \quad + \sum_{i=1}^N \mathbb{E} f_{\gamma,i} \left(\frac{1}{N} \sum_{j=1}^N y_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N y_{j,-s+1}, \frac{1}{N} \sum_{j=1}^N p_{j,0}, \dots, \frac{1}{N} \sum_{j=1}^N p_{j,-s+1} \right) \\ & \quad + \left(\frac{1}{N} \sum_{i=1}^N \gamma_i^\top \right) \mathbb{E} f_\gamma \left(\sum_{j=1}^N y_{j,0}, \dots, \sum_{j=1}^N y_{j,-s+1}, \sum_{j=1}^N p_{j,0}, \dots, \sum_{j=1}^N p_{j,-s+1} \right) \\ & \xrightarrow{N \rightarrow \infty} \bar{c} + \sum_{\tau=1}^s \beta_\tau \mathbb{E} \lambda_{1-\tau} + \mathbb{E}(X_0, \dots, X_{-s+1}, \lambda_0, \dots, \lambda_{-s+1}) \bar{\mathbf{f}} \\ & \quad + \bar{\gamma}^\top \mathbb{E} f_\gamma(X_0, \dots, X_{-s+1}, \lambda_0, \dots, \lambda_{-s+1}) = \mathbb{E} \lambda_1. \end{aligned}$$

Finally,

$$\sum_{i=1}^N \mathbb{E} y_{i,1} = \sum_{i=1}^N \mathbb{E} p_{i,1} \xrightarrow{N \rightarrow \infty} \mathbb{E} \lambda_1 = \mathbb{E} [\mathbb{E} (X_1 \mid \lambda_{-s+1}, X_{-s+1}, \dots, \lambda_0, X_0, \lambda_1)] = \mathbb{E} X_1.$$

Repeating the same steps for general T , we arrive at the desired joint convergence. $\square$

Proof of Theorem 10. The proof follows the lines of the previous proofs and proceeds by induction. The difference is that $\max_i p_{i,t}$ involves the term $\sum_{j=1}^N W_{ij} y_{j,t-1}$. We are going to show that this term still goes to 0 in probability.

$$\sum_{j=1}^N W_{ij} y_{j,t-1} \leq \frac{1}{d_i^{\text{out}}} \sum_{j=1}^N y_{j,t-1} \leq \frac{1}{d \log N} \sum_{j=1}^N y_{j,t-1} \xrightarrow[N \rightarrow \infty]{p} 0,$$

since $\sum_{j=1}^N y_{j,t-1} = O_p(1)$ by induction arguments.

Second, note that because W_{ij} is doubly stochastic, $\sum_{i=1}^N W_{ij} = 1$ and

$$\sum_{i=1}^N \sum_{j=1}^N W_{ij} y_{j,t-1} = \sum_{j=1}^N y_{j,t-1} \sum_{i=1}^N W_{ij} = \sum_{j=1}^N y_{j,t-1}.$$

Thus,

$$\sum_{i=1}^N p_{i,t} = \frac{1}{N} \sum_{i=1}^N c_i + \frac{1}{N\kappa} \sum_{i=1}^N a_i y_{i,t-1} + \beta \sum_{i=1}^N p_{i,t-1} + \gamma \sum_{i=1}^N y_{i,t-1}.$$

Therefore, repeating the steps of the proof of Theorem 4, we obtain the desired result. $\square$

APPENDIX C. PROOFS: ESTIMATION

Proof of Theorem 11. The proofs first defines the population analogue of the criterion function, then shows that the population criterion is maximized by θ_0 , and establishes the uniform convergence of the sample criterion to the population one. Finally, these three steps are combined to imply consistency.

Start by noting that both Assumptions I and II imply that if we choose different starting values at time r to initialize probabilities, but then use exactly the same realizations of $\mathbf{y}_t$ for two dynamics with the same parameter θ , then there exists $\varrho \in [0, 1)$ such that

$$(49) \quad \sup_{\theta \in \Theta} \|\mathbf{p}_t(\theta) - \mathbf{p}'_t(\theta)\|_1 \leq C \varrho^{t-r}, \quad t \geq r,$$

where the constant C can be chosen uniformly over the initial values since all probability coordinates belong to $[0, 1]$.

To define a proper population object, consider a unique stationary solution $\{\mathbf{y}_t, \mathbf{p}_t\}_{t \in \mathbb{Z}}$ of Theorem 3 with corresponding to the true parameter θ_0 . for every $\theta \in \Theta$, consider the recursion

$$(50) \quad \mathbf{p}_t^*(\theta) = \mathbf{g}(\{\mathbf{p}_\tau^*(\theta)\}_{\tau=t-1, \dots, t-s}, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta),$$

where the recursion is driven by the observations generated under the true parameter θ_0 . Let us show that there is a unique stationary solution to the recursion (50). Fix arbitrary initial values $\mathbf{p}^0 \in [0, 1]^{Ns}$ and consider the process $\mathbf{p}_t^{[r]}(\theta)$, which at time $t = r$ starts the recursion (50) from $\mathbf{p}^0$, i.e.,

$$\mathbf{p}_r^{[r]}(\theta) = \mathbf{g}(\mathbf{p}^0, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta),$$

and for $t > r$ continues as in the recursion by plugging $\mathbf{p}_t^{[r]}(\theta)$ in the right-hand side of the recursion (for $r < t < r + s$ use chosen initial values instead of the unobserved lags $\mathbf{p}_{t-t'}^{[r]}(\theta)$, $t - r < t' \leq s$). Then for $r' < r$, starting from $t = r$, $\mathbf{p}_t^{[r]}(\theta)$ and $\mathbf{p}_t^{[r']}(\theta)$ are driven by exactly the same observations $\mathbf{y}_t$. Thus, (49) implies

$$\sup_{\theta \in \Theta} \left\| \mathbf{p}_t^{[r]}(\theta) - \mathbf{p}_t^{[r']}(\theta) \right\|_1 \leq C \varrho^{t-r}.$$

Therefore, $\mathbf{p}_t^{[r]}(\theta)$ is uniformly Cauchy as $r \rightarrow -\infty$ and we can define, for every realization of $\{\mathbf{y}_\tau\}$

$$(51) \quad \mathbf{p}_t^*(\theta) = \lim_{r \rightarrow -\infty} \mathbf{p}_t^{[r]}(\theta).$$

The convergence is uniform over $\theta \in \Theta$, and the limit is independent of $\mathbf{p}^0$ (again, due to (49)). Since $\mathbf{g}$ is continuous and $\mathbf{p}_t^{[r]}(\theta)$ satisfies the recursion (50) for $t > r + s$, taking the limits, we obtain that $\mathbf{p}_t^*(\theta)$ satisfies the desired recursion for all t . Moreover, $\mathbf{p}_t^*(\theta)$ is measurable with respect to $\sigma(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots)$.

At the true parameter, $\mathbf{p}_t^*(\theta_0) = \mathbf{p}_t$, since both $\mathbf{p}_t$ and $\mathbf{p}_t^{[r]}(\theta_0)$ are driven by the same realizations $\mathbf{y}_\tau$ for $t > r$ and (49) implies

$$\left\| \mathbf{p}_t^{[r]}(\theta_0) - \mathbf{p}_t \right\|_1 \leq C \varrho^{t-r}.$$

Taking the limit $r \rightarrow -\infty$, we obtain $\mathbf{p}_t^*(\theta_0) = \mathbf{p}_t$.

Finally, because $\mathbf{g}$ satisfies either Assumptions I or II, it is continuous function of $\mathbf{p}_t$ and $\mathbf{y}_t$ (uniformly over θ). Thus, $\mathbf{p}_t^{[r]}(\theta)$ is a continuous function of θ for fixed $t > r$. Since $\mathbf{p}_t^{[r]}(\theta) \rightarrow \mathbf{p}_t^*(\theta)$ uniformly over θ , $\mathbf{p}_t^*(\theta)$ is also continuous function of θ .

To show population identification of the parameter, define

$$(52) \quad \ell_t^*(\theta) = \sum_{i=1}^N [y_{i,t} \log p_{i,t}^*(\theta) + (1 - y_{i,t}) \log(1 - p_{i,t}^*(\theta))]$$

and the population objective

$$(53) \quad Q(\theta) := \mathbb{E} \ell_t^*(\theta).$$

By Assumption III(iii),

$$(54) \quad \sup_{\theta \in \Theta} |\ell_t^*(\theta)| \leq N |\log \varepsilon|.$$

Let us show that $Q(\theta)$ has a unique maximum at θ_0 . Consider

$$(55) \quad \begin{aligned} Q(\theta) - Q(\theta_0) &= \sum_{i=1}^N \mathbb{E} \left[y_{i,t} \log \frac{p_{i,t}^*(\theta)}{p_{i,t}} + (1 - y_{i,t}) \log \frac{1 - p_{i,t}^*(\theta)}{1 - p_{i,t}} \right] \\ &= \sum_{i=1}^N \mathbb{E} \log \frac{(p_{i,t}^*(\theta))^{y_{i,t}} (1 - p_{i,t}^*(\theta))^{1-y_{i,t}}}{p_{i,t}^{y_{i,t}} (1 - p_{i,t})^{1-y_{i,t}}} \leq \sum_{i=1}^N \mathbb{E} \frac{(p_{i,t}^*(\theta))^{y_{i,t}} (1 - p_{i,t}^*(\theta))^{1-y_{i,t}}}{p_{i,t}^{y_{i,t}} (1 - p_{i,t})^{1-y_{i,t}}} - N \\ &= \sum_{i=1}^N \mathbb{E} \mathbb{E}_{t-1} \frac{(p_{i,t}^*(\theta))^{y_{i,t}} (1 - p_{i,t}^*(\theta))^{1-y_{i,t}}}{p_{i,t}^{y_{i,t}} (1 - p_{i,t})^{1-y_{i,t}}} - N \\ &= \sum_{i=1}^N \mathbb{E} \left(\frac{p_{i,t}^*(\theta)}{p_{i,t}} p_{i,t} + \frac{1 - p_{i,t}^*(\theta)}{1 - p_{i,t}} (1 - p_{i,t}) \right) - N \equiv N - N = 0, \end{aligned}$$

where we used the inequality $\log(a) \leq a - 1$ for any $a \geq 0$. Thus, θ_0 is a maximum of $Q(\theta)$. Because of (54), the value of $Q(\theta_0)$ is finite. Moreover, θ_0 is the unique maximum of $Q(\theta)$, because equality in (55) is only possible if $\frac{(p_{i,t}^*(\theta))^{y_{i,t}} (1 - p_{i,t}^*(\theta))^{1-y_{i,t}}}{p_{i,t}^{y_{i,t}} (1 - p_{i,t})^{1-y_{i,t}}} = 1$ with probability 1. So equality requires $p_{i,t}^*(\theta) = p_{i,t}$ for all i (because, conditional on $\mathcal{F}_{t-1}$, $y_{i,t}$ takes both values 0 and 1 with positive probability under Assumption III(iii)). This cannot hold by Assumption III(iv) if $\theta \neq \theta_0$.

Finally, because $\mathbf{p}_t^*(\theta)$ is a continuous function of θ and $g_i(\cdot) \in [\varepsilon, 1 - \varepsilon]$, for any $\delta > 0$,

$$(56) \quad \sup_{\theta \in \Theta \setminus B(\theta_0, \delta)} Q(\theta) < Q(\theta_0),$$

where $B(\theta_0, \delta)$ denotes an open ball of radius δ around θ_0 .

Next, we need to verify the uniform convergence of the sample criterion to the population one:

$$(57) \quad \sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^T \ell_t^*(\theta) - Q(\theta) \right| \xrightarrow[T \rightarrow \infty]{p} 0.$$

To establish (57), fix an integer $M \geq 1$ and consider $\mathbf{p}_t^{[t-M]}(\theta)$, that is, the probability filter initialized M periods before t at the fixed values $\mathbf{p}^0$. Define correspondingly

$$\ell_{t,M}(\theta) = \sum_{i=1}^N \left[y_{i,t} \log p_{i,t}^{[t-M]}(\theta) + (1 - y_{i,t}) \log \left(1 - p_{i,t}^{[t-M]}(\theta) \right) \right], \quad Q_M(\theta) := \mathbb{E} \ell_{t,M}(\theta).$$

By (49),

$$\sup_{\theta \in \Theta} \|\mathbf{p}_t^{[t-M]}(\theta) - \mathbf{p}_t^*(\theta)\|_1 \leq C \varrho^M.$$

Combining the above bound with Assumption III(iii),

$$(58) \quad \sup_{\theta \in \Theta} |\ell_{t,M}(\theta) - \ell_t^*(\theta)| \leq \frac{C}{\varepsilon} \varrho^M,$$

and hence

$$(59) \quad \sup_{\theta \in \Theta} |Q_M(\theta) - Q(\theta)| \leq \frac{C}{\varepsilon} \varrho^M.$$

For fixed M , $\ell_{t,M}(\theta)$ is a continuous function of θ and of a fixed finite block of the stationary Markov process $\{Z_t\}$. Therefore, by Corollary 13 applied to the corresponding finite-lag augmentation of Z_t , for every fixed $\theta \in \Theta$,

$$(60) \quad \frac{1}{T} \sum_{t=1}^T \ell_{t,M}(\theta) \xrightarrow[T \rightarrow \infty]{p} Q_M(\theta).$$

The convergence is uniform in θ , since $\ell_{t,M}(\theta)$ is continuous and the state-space of Z_t and Θ are compact, implying uniform continuity:

$$(61) \quad \sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^T \ell_{t,M}(\theta) - Q_M(\theta) \right| \xrightarrow[T \rightarrow \infty]{p} 0$$

for every fixed M .

Finally, (58)–(61) give (57),

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^T \ell_t^*(\theta) - Q(\theta) \right| \leq \sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^T \ell_{t,M}(\theta) - Q_M(\theta) \right| + \frac{2C}{\varepsilon} \varrho^M \rightarrow 0,$$

as first $T \rightarrow \infty$, and then $M \rightarrow \infty$.

To finish the proof, note that $\tilde{\mathbf{p}}_t(\theta)$ and $\mathbf{p}_t^*(\theta)$ are driven by the same observed sequence $\{\mathbf{y}_t\}$ and differ only in their probability initialization, so bound (49) gives

$$(62) \quad \sup_{\theta \in \Theta} \|\tilde{\mathbf{p}}_t(\theta) - \mathbf{p}_t^*(\theta)\|_1 \leq C \varrho^t.$$

Assumption III(iii) implies that for $\ell_{i,t}(p) = y_{i,t} \log p + (1 - y_{i,t}) \log(1 - p)$,

$$\sup_{p \in [\varepsilon, 1-\varepsilon]} \left| \frac{\partial \ell_{i,t}(p)}{\partial p} \right| \leq \frac{1}{\varepsilon}.$$

Combining the above bound with (62), we get

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \mathcal{Q}_T(\theta) - \frac{1}{T} \sum_{t=1}^T \ell_t^*(\theta) \right| \leq \frac{C}{T\varepsilon} \sum_{t=1}^T \varrho^t \leq \frac{C}{T\varepsilon(1-\varrho)} \xrightarrow{T \rightarrow \infty} 0,$$

which together with (57) implies

$$(63) \quad \sup_{\theta \in \Theta} \left| \frac{1}{T} \mathcal{Q}_T(\theta) - Q(\theta) \right| \xrightarrow[T \rightarrow \infty]{p} 0.$$

Identification (56) and uniform convergence (63) imply that for any $\delta > 0$, $\text{Prob}(\|\hat{\theta}^{MLE} - \theta_0\| > \delta) \xrightarrow[T \rightarrow \infty]{} 0$. That is, $\hat{\theta}^{MLE} \xrightarrow[T \rightarrow \infty]{p} \theta_0$. $\square$

Proof of Theorem 12. The proof starts with defining first and second order derivatives, which appear in the score and the Hessian. Then the CLT for the score and LLN for the Hessian are derived and combined to obtain asymptotic normality of $\hat{\theta}^{MLE}$.

Let $\mathbf{p}_t^*(\theta)$ denote the stationary probability, which satisfies recursion (50) constructed with the true realizations $\mathbf{y}_t$, but different θ , as in the proof of Theorem 11. We first need to show that $\mathbf{p}_t^*(\theta)$ is twice continuously differentiable. For this purpose, consider its finite-past approximations $\mathbf{p}_t^{[r]}(\theta)$ from fixed and independent of θ initial values $\mathbf{p}^0 \in [0, 1]^{N_s}$ and define

$$G_t(P; \theta) := \mathbf{g}(P, \{\mathbf{y}_\tau\}_{\tau=t-1, \dots, t-q}; \theta),$$

where P stacks the s probability lags. Since G_t is twice continuously differentiable, $\mathbf{p}_t^{[r]}(\theta)$ is twice continuously differentiable in θ . Writing its first derivative as an $N \times m$ Jacobian $D_t^{[r]}(\theta)$, differentiation of the filter recursion gives

$$(64) \quad \begin{aligned} D_t^{[r]}(\theta) &= G_{\theta,t}^{[r]}(\theta) + \sum_{j=1}^s G_{p_j,t}^{[r]}(\theta) D_{t-j}^{[r]}(\theta), \quad t \geq r, \\ D_{r-j}^{[r]}(\theta) &= 0, \quad j = 1, \dots, s, \end{aligned}$$

where the derivatives of G_t are evaluated along the finite-past filter $\mathbf{p}_t^{[r]}(\theta)$. The derivative recursion inherits the contraction property of Assumptions I and II. Under I, the absolute values of the entries of $G_{p_j,t}^{[r]}(\theta)$ are bounded by the corresponding probability Lipschitz coefficients. Hence, after stacking the s derivative lags, the homogeneous part of (66) is dominated by the companion matrix in Assumption I, whose spectral radius is strictly below one. Under Assumption II, the contractive bound is applied to the ℓ_1 -norms of each of m

columns (and, thus, to their maximum). Thus, under either Assumption I or II,

$$(65) \quad \sup_{r \in \mathbb{Z}} \sup_{t \geq r} \sup_{\theta \in \Theta_0} \|D_t^{[r]}(\theta)\|_1 < \infty.$$

If $r' < r$, subtract the derivative recursions corresponding to $\mathbf{p}_t^{[r]}(\theta)$ and $\mathbf{p}_t^{[r']}(\theta)$. Writing

$$\Delta_t(\theta) := D_t^{[r]}(\theta) - D_t^{[r']}(\theta),$$

we obtain

$$\Delta_t(\theta) = \sum_{j=1}^s G_{p_j, t}^{[r]}(\theta) \Delta_{t-j}(\theta) + R_t(\theta),$$

where

$$R_t(\theta) = G_{\theta, t}^{[r]}(\theta) - G_{\theta, t}^{[r']}(\theta) + \sum_{j=1}^s \left(G_{p_j, t}^{[r]}(\theta) - G_{p_j, t}^{[r']}(\theta) \right) D_{t-j}^{[r']}(\theta).$$

Since the second derivatives of G_t are uniformly bounded on the compact set Θ_0 , its first derivatives are uniformly Lipschitz in the probability arguments. Combining this with (49) and the uniform boundedness of $D_t^{[r']}(\theta)$ gives

$$\sup_{\theta \in \Theta_0} \|R_t(\theta)\| \leq C \varrho^{t-r}.$$

Since $\Delta_t(\theta)$ satisfies the same geometrically contracting recursion as the derivative filter itself (modulo different intercept), there exist $C < \infty$ and $\rho_1 \in [0, 1)$, such that

$$\sup_{\theta \in \Theta_0} \|D_t^{[r]}(\theta) - D_t^{[r']}(\theta)\| \leq C \rho_1^{t-r}.$$

Thus the derivatives are uniformly Cauchy as $r \rightarrow -\infty$. Since $\mathbf{p}_t^{[r]}(\theta) \rightarrow \mathbf{p}_t^*(\theta)$ uniformly in θ , $\mathbf{p}_t^*(\theta)$ is continuously differentiable and

$$D_t^*(\theta) := \frac{\partial \mathbf{p}_t^*(\theta)}{\partial \theta^\top} = \lim_{r \rightarrow -\infty} D_t^{[r]}(\theta) = \lim_{r \rightarrow -\infty} \frac{\partial \mathbf{p}_t^{[r]}(\theta)}{\partial \theta^\top}$$

uniformly over Θ_0 . Taking the limit in (66),

$$(66) \quad D_t^*(\theta) = G_{\theta, t}^*(\theta) + \sum_{j=1}^s G_{p_j, t}^*(\theta) D_{t-j}^*(\theta).$$

In particular, at $\theta = \theta_0$, because $\mathbf{p}_t^*(\theta_0) = \mathbf{p}_t$, (66) coincides with the recursion in Assumption IV(iii).

Differentiating (66) once more and using the same arguments as before, where second derivatives of $\mathbf{g}$ are continuous and bounded by Assumption IV(ii), shows that the second

derivative of $\mathbf{p}_t^*(\theta)$ exist and is uniformly bounded on Θ_0 :

$$\sup_t \sup_{\theta \in \Theta_0} \left\| \frac{\partial^2 \mathbf{p}_t^*(\theta)}{\partial \theta \partial \theta^\top} \right\| < \infty.$$

The same recursions give forgetting of the derivatives for the initialized filter $\tilde{\mathbf{p}}_t(\theta)$ used in the likelihood. If $\tilde{D}_t(\theta) = \partial \tilde{\mathbf{p}}_t(\theta) / \partial \theta^\top$, then, since G_t is twice differentiable, so its first derivative is Lipschitz, there exist $C < \infty$ and $\rho_2 \in [0, 1)$,

$$\sup_{\theta \in \Theta_0} \|\tilde{D}_t(\theta) - D_t^*(\theta)\| \leq C \rho_2^t.$$

Similarly, a (weaker) bound on the second derivative holds:

$$\sup_{\theta \in \Theta_0} \left\| \frac{\partial^2 \tilde{\mathbf{p}}_t(\theta)}{\partial \theta \partial \theta^\top} - \frac{\partial^2 \mathbf{p}_t^*(\theta)}{\partial \theta \partial \theta^\top} \right\| \xrightarrow[t \rightarrow \infty]{a.s.} 0.$$

Now we can analyze the score and the Hessian of the MLE problem. Recall the log-likelihood in (52),

$$\ell_t^*(\theta) = \sum_{i=1}^N [y_{i,t} \log p_{i,t}^*(\theta) + (1 - y_{i,t}) \log(1 - p_{i,t}^*(\theta))]$$

and define the score function

$$s_t^*(\theta) = \frac{\partial \ell_t^*(\theta)}{\partial \theta} = \sum_{i=1}^N \left[\frac{y_{i,t}}{p_{i,t}^*(\theta)} - \frac{1 - y_{i,t}}{1 - p_{i,t}^*(\theta)} \right] \frac{\partial}{\partial \theta} p_{i,t}^*(\theta) = \sum_{i=1}^N \frac{y_{i,t} - p_{i,t}^*(\theta)}{p_{i,t}^*(\theta)(1 - p_{i,t}^*(\theta))} \frac{\partial}{\partial \theta} p_{i,t}^*(\theta).$$

Since both $p_{i,t}^*$ and $\frac{\partial}{\partial \theta} p_{i,t}^*(\theta)$ are $\mathcal{F}_{t-1}$ measurable, $p_{i,t}^*(\theta_0) = p_{i,t}$, and $\mathbb{E}_{t-1} y_{i,t} = p_{i,t}$, $\{s_t^*(\theta_0), \mathcal{F}_t\}$ is a martingale difference sequence. Its conditional variance is

$$\begin{aligned} V_t &:= \mathbb{E}_{t-1} s_t(\theta_0) s_t(\theta_0)^\top = \sum_{i=1}^N \frac{\partial}{\partial \theta} p_{i,t}^*(\theta_0) \frac{\partial}{\partial \theta^\top} p_{i,t}^*(\theta_0) \mathbb{E}_{t-1} \left(\frac{y_{i,t} - p_{i,t}}{p_{i,t}(1 - p_{i,t})} \right)^2 \\ &\quad + \sum_{i \neq j} \frac{\partial}{\partial \theta} p_{i,t}^*(\theta_0) \frac{\partial}{\partial \theta^\top} p_{j,t}^*(\theta_0) \mathbb{E}_{t-1} \left(\frac{y_{i,t} - p_{i,t}}{p_{i,t}(1 - p_{i,t})} \right) \left(\frac{y_{j,t} - p_{j,t}}{p_{j,t}(1 - p_{j,t})} \right) \\ &= \sum_{i=1}^N \frac{1}{p_{i,t}(1 - p_{i,t})} \frac{\partial}{\partial \theta} p_{i,t}^*(\theta_0) \frac{\partial}{\partial \theta^\top} p_{i,t}^*(\theta_0). \end{aligned}$$

Let us show that

$$(67) \quad \frac{1}{T} \sum_{t=1}^T V_t \xrightarrow[T \rightarrow \infty]{p} H_0.$$

We proceed along the lines of the LLN part of the proof of Theorem 11. Fix M and initialize the probability recursion and its derivative recursion at time $t - M$ from fixed values $\mathbf{p}^0$.

Then define

$$V_{t,M} = \sum_{i=1}^N \frac{\left(D_{i,t}^{[t-M]}\right)^\top D_{i,t}^{[t-M]}}{p_{i,t}^{[t-M]} \left(1 - p_{i,t}^{[t-M]}\right)}.$$

Filter and its derivative past-forgetting imply

$$\sup_t \|V_{t,M} - V_t\| \xrightarrow{M \rightarrow \infty} 0.$$

For every fixed M , $V_{t,M}$ is a continuous function of a fixed finite block of the stationary Markov process $\{Z_t\}$. Therefore, Corollary 13, applied entrywise and to the corresponding finite-lag augmentation of Z_t , gives

$$\frac{1}{T} \sum_{t=1}^T V_{t,M} \xrightarrow[T \rightarrow \infty]{p} \mathbb{E} V_{t,M}.$$

Letting first $T \rightarrow \infty$ for fixed M , and then $M \rightarrow \infty$, proves LLN (67), since $\mathbb{E} V_t = H_0$.

Assumption III(iii) and the boundedness of the derivative (65) imply that $\|s_t^*(\theta_0)\| \leq C$. Hence the conditional Lindeberg condition is trivially satisfied and the martingale CLT gives

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t^*(\theta_0) \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, H_0).$$

For the initialized likelihood, let

$$\tilde{s}_t(\theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^N [y_{i,t} \log \tilde{p}_{i,t}(\theta) + (1 - y_{i,t}) \log(1 - \tilde{p}_{i,t}(\theta))].$$

The probability and derivative forgetting bounds, together with the fact that probabilities are bounded away from zero and one, imply

$$\|\tilde{s}_t(\theta_0) - s_t^*(\theta_0)\| \leq C \rho_1^t.$$

Consequently,

$$\frac{1}{\sqrt{T}} \left\| \sum_{t=1}^T (\tilde{s}_t(\theta_0) - s_t^*(\theta_0)) \right\| \rightarrow 0,$$

and hence

$$(68) \quad \frac{1}{\sqrt{T}} \frac{\partial \mathcal{Q}_T(\theta_0)}{\partial \theta} \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, H_0).$$

To analyze the Hessian, let

$$\begin{aligned} \mathcal{H}_t^*(\theta) &= \frac{\partial^2 \ell_t^*(\theta)}{\partial \theta \partial \theta^\top} = \sum_{i=1}^N \frac{y_{i,t} - p_{i,t}^*(\theta)}{p_{i,t}^*(\theta)(1 - p_{i,t}^*(\theta))} \frac{\partial^2}{\partial \theta \partial \theta^\top} p_{i,t}^*(\theta) \\ &\quad - \sum_{i=1}^N \left(\frac{y_{i,t}}{p_{i,t}^*(\theta)^2} + \frac{1 - y_{i,t}}{(1 - p_{i,t}^*(\theta))^2} \right) \left(\frac{\partial}{\partial \theta} p_{i,t}^*(\theta) \right) \left(\frac{\partial}{\partial \theta} p_{i,t}^*(\theta) \right)^\top. \end{aligned}$$

We need to show the uniform convergence of the Hessian to its mean, $\mathbb{E}\mathcal{H}_t^*(\theta)$. Note that

$$(69) \quad \mathbb{E}\mathcal{H}_t^*(\theta_0) = -H_0.$$

As in the proof of the uniform convergence of the criterion in Theorem 11, truncate the probability recursion together with its first and second derivative recursions M periods in the past, and let $\mathcal{H}_{t,M}^*(\theta)$ denote the resulting Hessian. Repeating the same steps: LLN for $\mathcal{H}_{t,M}^*(\theta)$ and forgetting of the initial value of $\mathcal{H}_{t,M}^*(\theta)$, we obtain

$$(70) \quad \sup_{\theta \in \Theta_0} \left\| \frac{1}{T} \sum_{t=1}^T \mathcal{H}_t^*(\theta) - \mathbb{E}\mathcal{H}_t^*(\theta) \right\| \xrightarrow[T \rightarrow \infty]{p} 0.$$

Let $\tilde{\mathcal{H}}_t(\theta)$ be the Hessian formed from the initialized filter. By probability, first-derivative, and second-derivative forgetting past,

$$\sup_{\theta \in \Theta_0} \|\tilde{\mathcal{H}}_t(\theta) - \mathcal{H}_t^*(\theta)\| \xrightarrow[t \rightarrow \infty]{a.s.} 0.$$

Therefore,

$$\sup_{\theta \in \Theta_0} \left\| \frac{1}{T} \sum_{t=1}^T \tilde{\mathcal{H}}_t(\theta) - \frac{1}{T} \sum_{t=1}^T \mathcal{H}_t^*(\theta) \right\| \leq \frac{1}{T} \sum_{t=1}^T \sup_{\theta \in \Theta_0} \|\tilde{\mathcal{H}}_t(\theta) - \mathcal{H}_t^*(\theta)\| \xrightarrow[T \rightarrow \infty]{a.s.} 0.$$

Combining the preceding with (70),

$$(71) \quad \sup_{\theta \in \Theta_0} \left\| \frac{1}{T} \frac{\partial^2 \mathcal{Q}_T(\theta)}{\partial \theta \partial \theta^\top} - \mathbb{E}\mathcal{H}_t^*(\theta) \right\| \xrightarrow[T \rightarrow \infty]{p} 0.$$

Finally, we can combine the score and the Hessian asymptotics to obtain the asymptotic normality of $\hat{\theta}^{MLE}$. By Theorem 11, $\hat{\theta}^{MLE} \xrightarrow[T \rightarrow \infty]{p} \theta_0$. Since θ_0 is an interior point of Θ , with probability approaching one $\hat{\theta}^{MLE} \in \Theta_0$ and satisfies

$$\frac{\partial \mathcal{Q}_T(\hat{\theta}^{MLE})}{\partial \theta} = 0.$$

By the mean value theorem,

$$0 = \frac{1}{\sqrt{T}} \frac{\partial \mathcal{Q}_T(\theta_0)}{\partial \theta} + \frac{1}{T} \frac{\partial^2 \mathcal{Q}_T(\tilde{\theta})}{\partial \theta \partial \theta^\top} \sqrt{T}(\hat{\theta}^{MLE} - \theta_0),$$

where each component of $\frac{\partial \mathcal{Q}_T(\theta_0)}{\partial \theta}$ corresponds to a separate value of $\tilde{\theta}$ on a segment connecting θ_0 and $\hat{\theta}^{MLE}$. Combining the score CLT (68), the Hessian limit (69) and uniform LLN (71), and consistency of $\hat{\theta}^{MLE}$, we get

$$\sqrt{T}(\hat{\theta}^{MLE} - \theta_0) = - \left(\frac{1}{T} \frac{\partial^2 \mathcal{Q}_T(\tilde{\theta})}{\partial \theta \partial \theta^\top} \right)^{-1} \frac{1}{\sqrt{T}} \frac{\partial \mathcal{Q}_T(\theta_0)}{\partial \theta} \xrightarrow[T \rightarrow \infty]{d} H_0^{-1} \mathcal{N}(0, H_0) = \mathcal{N}(0, H_0^{-1}). \quad \square$$

APPENDIX D. DATA

As of December 27, 2025, the S&P 100 consists of the following stock tickers: AAPL, ABBV, ABT, ACN, ADBE, AIG, AMD, AMGN, AMT, AMZN, AVGO, AXP, BA, BAC, BK, BKNG, BLK, BMY, BRKB, C, CAT, CL, CMCSA, COF, COP, COST, CRM, CSCO, CVS, CVX, DE, DHR, DIS, DUK, EMR, FDX, GD, GE, GM, GILD, GOOGL, GS, HD, HON, IBM, INTC, INTU, ISRG, JNJ, JPM, KO, LIN, LLY, LMT, LOW, MA, MCD, MDLZ, MDT, MET, META, MMM, MO, MRK, MS, MSFT, NEE, NFLX, NKE, NOW, NVDA, ORCL, PEP, PFE, PG, PM, PLTR, PYPL, QCOM, RTX, SBUX, SCHW, SO, SPG, T, TGT, TMO, TMUS, TSLA, TXN, UBER, UNH, UNP, UPS, USB, V, VZ, WFC, WMT, XOM.

Six of the above companies are not available for the entire period under consideration, 01.01.2012–12.31.2025. The tickers of those companies are: ABBV, META, NOW, PLTR, PYPL, and UBER.

REFERENCES

- AKNOUCHE, A. and FRANCO, C. (2021). Count and duration time series with equal conditional stochastic and mean orders. *Econometric Theory*, **37** (2), 248–280.
- BLASQUES, F., KOOPMAN, S. J. and LUCAS, A. (2014). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Electronic Journal of Statistics*, **8**, 1088–1112.
- , VAN BRUMMELEN, J., KOOPMAN, S. J. and LUCAS, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, **227** (2), 325–346.
- BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, **31** (3), 307–327.
- BYKHOVSKAYA, A. (2022). Time series approach to the evolution of networks: Prediction and estimation. *Journal of Business & Economic Statistics*, **41** (1), 170–183.
- CREAL, D., KOOPMAN, S. J. and LUCAS, A. (2013). Generalized autoregressive score models with applications. *Journal of applied econometrics*, **28** (5), 777–795.
- DAVIS, R. A. and LIU, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, pp. 1673–1707.

- DE JONG, R. M. and WOUTERSEN, T. (2011). Dynamic time series binary choice. *Econometric Theory*, **27** (4), 673–702.
- DUFFIE, D., ECKNER, A., HOREL, G. and SAITA, L. (2009). Frailty correlated default. *The Journal of Finance*, **64** (5), 2089–2123.
- FERLAND, R., LATOUR, A. and ORAICHI, D. (2006). Integer-valued GARCH process. *Journal of time series analysis*, **27** (6), 923–942.
- FOKIANOS, K. and MOYSIADIS, T. (2017). Binary time series models driven by a latent process. *Econometrics and Statistics*, **2**, 117–130.
- , RAHBEK, A. and TJØSTHEIM, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, **104** (488), 1430–1439.
- FRANCQ, C. and ZAKOIAN, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.
- GRAHAM, B. S. (2016). Homophily and transitivity in dynamic network formation. *working paper*.
- HECKMAN, J. J. (1981). Heterogeneity and state dependence. In *Studies in labor markets*, University of Chicago Press, pp. 91–140.
- KAUPPI, H. and SAIKKONEN, P. (2008). Predicting US recessions with dynamic binary response models. *The Review of Economics and Statistics*, **90** (4), 777–791.
- KEDEM, B. and FOKIANOS, K. (2005). *Regression models for time series analysis*. John Wiley & Sons.
- LEE, S., KIM, D. and KIM, B. (2023). Modeling and inference for multivariate time series of counts based on the INGARCH scheme. *Computational Statistics & Data Analysis*, **177**, 107579.
- MEYN, S. P. and TWEEDIE, R. L. (2009). *Markov chains and stochastic stability*. Cambridge University Press.
- MOYSIADIS, T. and FOKIANOS, K. (2014). On binary and categorical time series models with feedback. *Journal of Multivariate Analysis*, **131**, 209–228.
- NYBERG, H. (2013). Predicting bear and bull stock markets with dynamic binary time series models. *Journal of Banking & Finance*, **37** (9), 3351–3363.
- RUST, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, pp. 999–1033.
- RYDBERG, T. H. and SHEPHARD, N. (2000). A modelling framework for the prices and times of trades made on the New York stock exchange. In *Nonlinear and Nonstationary Signal Processing*, Cambridge University Press, pp. 217–246.
- STREETT, S. (2000). *Some observation driven models for time series of counts*. Ph.D. thesis, Colorado State University, Department of Statistics.

(Anna Bykhovskaya) DUKE UNIVERSITY

Email address: `anna.bykhovskaya@duke.edu`

(Nour Meddahi) TOULOUSE SCHOOL OF ECONOMICS

Email address: `nour.meddahi@tse-fr.eu`