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 microfoundation 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, including extreme return days or bear markets (Nyberg, 2013), corporate defaults and credit downgrades (Duffie *et al.*, 2009), and the exercise of binary options. 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).

Date: April 15, 2026.

As we show, such cross-sectional interactions, when aggregated across a large panel under a rare-events scaling, provide a natural micro-foundation for Poisson autoregression, a leading model for count time series.

We develop a general framework for multivariate binary time series in which each binary observation is treated as a realization of a Bernoulli variable whose success probability is time-varying and depends on both past outcomes and past success probabilities, in the spirit of the GARCH-type dynamics of Bollerslev (1986). For our model we establish, under general and potentially nonlinear dynamics for the success probabilities, the existence of a stationary solution and, under a contraction condition on the probability update function, its uniqueness. A distinctive feature of our formulation, compared to, for example, Kedem and Fokianos (2005) and De Jong and Woutersen (2011), is that the success probabilities may depend on latent past probabilities and not only on observed outcomes. This additional layer allows for richer dynamics and feedback effects that cannot be captured when probabilities evolve solely as functions of observed realizations.

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 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 the reparametrizing 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 (INGARCH) 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 2005–2025, 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 $\mathbf{y}_t = (y_{1,t}, \ldots, y_{N,t})^\mathsf{T}$, where T denotes transposition. Let $\mathcal{I}_t = \sigma(\mathbf{y}_\tau, \tau \leq t)$ denote the information available at time t, that is, the σ-field generated by $\{\mathbf{y}_t, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots\}$. The components of $\mathbf{y}_t$ are assumed to be independent conditional on $\mathcal{I}_{t-1}$. Given $\mathcal{I}_{t-1}$, each component $y_{i,t}$, $i = 1, \ldots, N$, is assumed to be a Bernoulli random variable with success probability $p_{i,t}$,

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

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

$$\mathbf{p}_t = \mathbf{g}\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\ldots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\ldots,t-q}\right), \tag{2}$$

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 $[0,1]^N$. 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).

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

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$(1,1)$

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

The conditions on the parameters ensure that p_t stays in $[0,1]$. This framework resembles 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)

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

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

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

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 $p_t = (1+\exp(-x_t))^{-1}$,

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

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 has its own $p_{i,t}$, independent of the others with $j \neq i$. 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,

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

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.

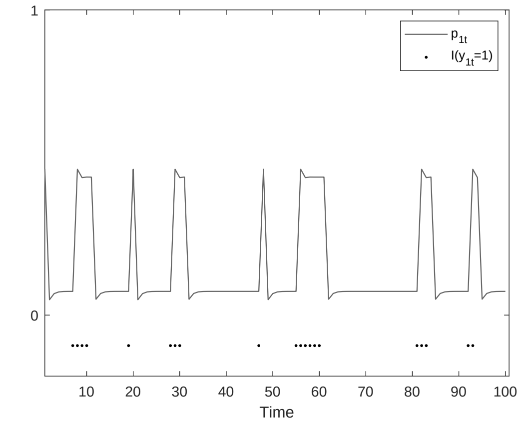
Example 6: Interactive GAB$(1,1)$. Each component has its own probability of success, driven by both its own past outcome and the cross-sectional average,

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

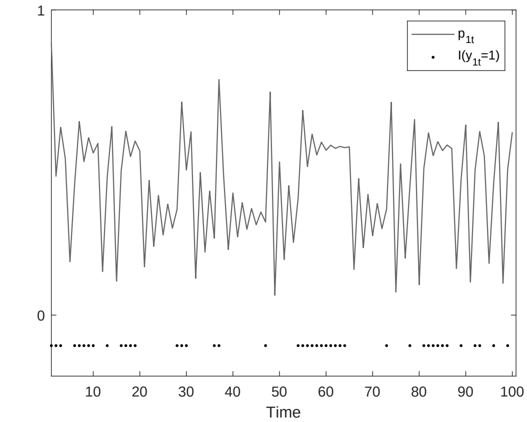
When $\gamma_i = 0$ for all i, the model (8) reduces to N trajectories of GAB$(1,1)$ models (3) that are independent conditional on $\mathcal{I}_{t-1}$. When, for all i, $\omega_i \equiv \omega$, $\alpha_i \equiv 0$, $\beta_i \equiv \beta$, $\gamma_i \equiv \gamma$, the model (8) reduces to a single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

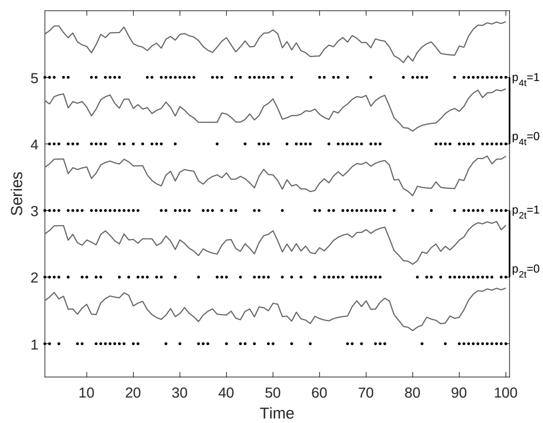
$$p_{i,t} = \omega_i + \alpha_i y_{i,t-1} + \gamma_i \sum_{j=1}^{N} W_{ij} y_{j,t-1} + \beta_i p_{i,t-1}, \quad \omega_i, \alpha_i, \gamma_i, \beta_i \geq 0, \quad \omega_i + \alpha_i + \gamma_i + \beta_i \leq 1, \tag{9}$$



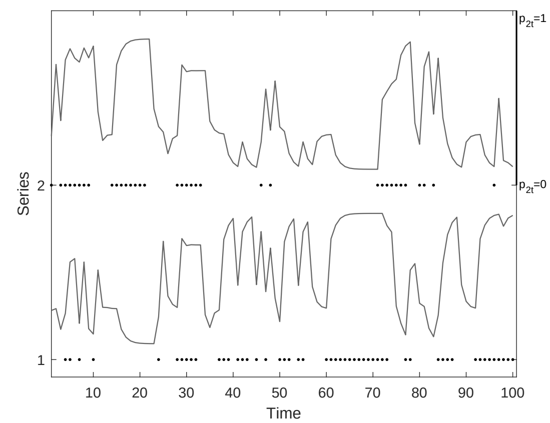
(A) Nonlinear single-series model: $p_{1,0} = 1/12$, $p_{1,t} = 0.05 + 0.4y_{1,t-1} + 2p_{1,t-1}(p_{1,t-1} - 0.5)^2$.



(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 $\omega_i = 0.05$, $\alpha_i = 0.1$, $\beta_i = 0.6$, $\gamma_i = 0.2$.



(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 $p_{i,t}$.

where W is an $N \times N$ row-normalized adjacency matrix governing the relationships among individuals $i = 1, \ldots, N$. 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 GAB$(1,1)$ to 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,[1] 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 $T \to \infty$, with the number of time series N held fixed.

2.2.1. *Existence.*

Theorem 1. *The multivariate binary time series model* (1)–(2) *has a strictly stationary solution.*

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 6,7, 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.

Corollary 2. *The multivariate binary time series model* (1)–(2) *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 the information available at time $t-1$, $\mathcal{I}_{t-1}$.

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

Assuming $\sum_{j=1}^{s}\alpha_j + \sum_{j=1}^{q}\beta_j < 1$, the linear GAB$(s,q)$ yields

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

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

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

and

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

In the network GAB$(1,1)$ model (9), similar argument yields, in matrix form with $A := \mathrm{diag}(\alpha_i+\beta_i)$ and $\Gamma := \mathrm{diag}(\gamma_i)$,

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

Since W is row-stochastic, the row sums of $A+\Gamma W$ equal $\alpha_i+\beta_i+\gamma_i \le 1-\omega_i$, so for $\omega_i>0$ the matrix $I-A-\Gamma W$ is strictly row-diagonally dominant and hence invertible, giving $\mu = (I-A-\Gamma W)^{-1}\omega$. 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 $\sum_{j=1}^{s}\alpha_j + \sum_{j=1}^{q}\beta_j = 1$, 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 strict 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* $\mathbf{g}(\cdot)=(g_1(\cdot),\dots,g_N(\cdot))$ *satisfies the Lipschitz bound*

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

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 $\sum_{j=1}^{N}\sum_{\tau=1}^{s}\beta_{ij}^{\tau}+\sum_{j=1}^{N}\sum_{\tau=1}^{q}\alpha_{ij}^{\tau}\le K<1$ for all i, which implies $\rho(\Phi)\le\|\Phi\|_\infty\le K<1$. 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* $0\le K<\frac{1}{s+q}$ *such that*

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

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 the coefficient 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 $\Phi = 0.8(1,1)^\mathsf{T}(1,1)$ and hence $\rho(\Phi) = 1.6$.

Theorem 3. *The multivariate binary time series model* (1)–(2) *with* $\mathbf{g}(\cdot)$ *satisfying either Assumption I or II has a unique strictly stationary solution and is geometrically ergodic.*

While we could not locate Theorem 3 in the literature, related results have appeared in Moysiadis and Fokianos (2014); 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 $y_{i,t} = I(u_{i,t} \le p_{i,t})$, 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, $X_t(N)$, given by

$$X_t(N) := y_{1,t} + \ldots + y_{N,t}. \tag{16}$$

While the previous section analyzed the limit of the system as $T \to \infty$ with N fixed, we now investigate the complementary asymptotic regime in which $N \to \infty$. To initialize the dynamics, we consider an initial condition

$$\{\mathbf{p}_0, \mathbf{p}_{-1}, \ldots, \mathbf{p}_{1-\max\{s,q\}}\}.$$

For $t = 0, \ldots, 1 - \max\{s, q\}$, we supplement this initial condition by setting $y_{i,t} = I(u_{i,t} \le p_{i,t})$, where the variables $u_{i,t}$ are Bernoulli and independent of one another (and independent of all $u_{i,t}$ appearing in Eq. (1)). The initial probabilities $\mathbf{p}_0, \mathbf{p}_{-1}, \ldots, \mathbf{p}_{1-\max\{s,q\}}$ 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 $N \to \infty$. 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,

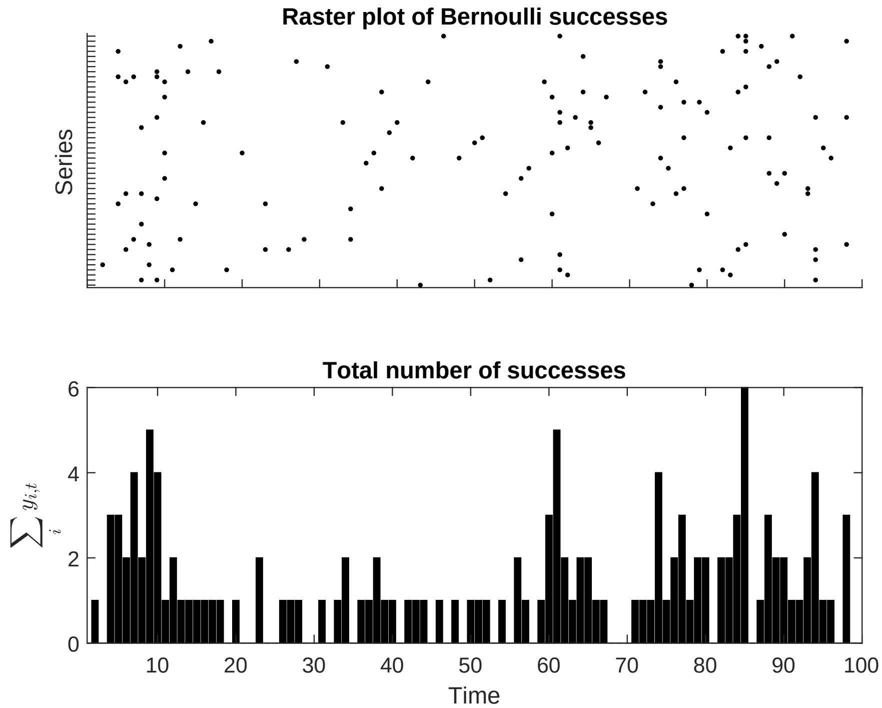


FIGURE 2. Individual and aggregate successes for $N = 50$ series generated from the interactive model (8). The top panel shows the raster of individual successes, while the bottom panel reports the total number of successes at each time period. Parameters are $\omega_i = 0.005$, $\alpha_i = 0.01$, $\beta_i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

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 8 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).

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

$$\sum_{i=1}^{N} p_{i,0} \xrightarrow[N\to\infty]{d} \lambda_0, \qquad \sum_{i=1}^{N} \mathbb{E} p_{i,0} \xrightarrow[N\to\infty]{} \mathbb{E}\lambda_0, \qquad \max_{i=1,\ldots,N} p_{i,0} \xrightarrow[N\to\infty]{p} 0, \tag{17}$$

where λ_0 *is some random variable. Then as* $N \to \infty$, $[X_t(N)]_{t=0,1,\ldots}$ *of Eq.* (16) *converges in finite-dimensional distributions to a process* $[X_t]_{t=0,1,\ldots}$.

Denote by $\mathcal{I}_t^X$ *the* σ*-field generated by* $\{X_t, \ldots, X_0, \lambda_0\}$*. Then, for each* $t \geq 0$, X_t *follows a Poisson distribution, Pois*(λ_t)*, conditional on* $\mathcal{I}_{t-1}^X$*. For each* $t \geq 0$, $\lambda_t = \lim_{N\to\infty} \sum_{i=1}^{N} p_{i,t}$ *and satisfies*

$$\lambda_t = \bar{c} + \bar{\gamma} X_{t-1} + \beta \lambda_{t-1}, \qquad t > 0. \tag{18}$$

Theorem 4 shows that different specifications for $\mathbf{p}_t$ can generate identical aggregate dynamics. For example, suppose that 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 remainder, with all other coefficients 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* $\alpha_i \to 0$ *as* $N \to \infty$ *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* $p_{i,t+1} \geq \alpha_i$*. Each unit therefore 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.*

There are many cases when the initial condition (17) is satisfied. For example, if $p_{i,0} = \frac{r_i}{N}$, where r_i are i.i.d. with finite mean. Then $\frac{1}{N} \max_{i=1,\ldots,N} r_i \xrightarrow[N\to\infty]{p} 0$ and

$\sum_{i=1}^N p_{i,0} = \frac{1}{N}\sum_{i=1}^N r_i \xrightarrow[N\to\infty]{p} \mathbb{E}r_1$. Moreover, it is satisfied by the stationary solution, as shown below.

Proposition 7. *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* (17).

Theorem 4 can be extended to allow for multiple lags and various nonlinearities.

Theorem 8. *Consider a nonlinear interactive model, which generalizes model* (8)*:*

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

where $\kappa > 0$, $0 \le c_i, \|\boldsymbol{\gamma}_i\| \le C < \infty$, $\boldsymbol{\gamma}_i \in \mathbb{R}_+^k$, $f_{\alpha,i} : \{0,1\}^s \to \mathbb{R}_+$, $f_{\gamma,i} : \times[0,1]^{2s} \to \mathbb{R}_+$, and $f_\gamma : \mathbb{R}_+^{2s} \to \mathbb{R}_+^k$. Suppose that, for all i, $f_{\gamma,i}$ is twice continuously differentiable, with uniformly bounded gradient and Hessian and with $f_{\gamma,i}(0,\dots,0) = 0$, $\nabla f_{\gamma,i}(0,\dots,0) \in \mathbb{R}_+^{2s}$. Suppose that, for all i, $|f_{\alpha,i}(y_{i,t-1},\dots,y_{i,t-s})| \le A\sum_{\tau=1}^s y_{i,t-\tau}$ for some finite A. Suppose that f_γ is either continuous and bounded or continuously differentiable with bounded Jacobian. Additionally suppose that there exist limits $\bar{c} := \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N c_i$, $\bar{\boldsymbol{\gamma}} := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \boldsymbol{\gamma}_i$ and $\bar{\mathbf{f}} := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \nabla f_{\gamma,i}(0,\dots,0)$. Finally, suppose that the initial conditions satisfy

$$\sum_{i=1}^N p_{i,\tau} \xrightarrow[N\to\infty]{d} \lambda_\tau, \qquad \sum_{i=1}^N \mathbb{E}p_{i,\tau} \xrightarrow[N\to\infty]{} \mathbb{E}\lambda_\tau, \qquad \max_{i=1,\dots,N} p_{i,\tau} \xrightarrow[N\to\infty]{p} 0, \qquad \tau = 0,\dots,-s+1, \tag{20}$$

where $\lambda_\tau, \tau = 0,\dots,-s+1$ are some random variable. Then as $N \to \infty$, $[X_t(N)]_{t=-s+1,\dots,0,1,\dots}$ of Eq. (16) *converges in finite-dimensional distributions to a process $[X_t]_{t=-s+1,\dots,0,1,\dots}$.*

Denote by $\mathcal{I}_t^X$ the σ-field generated by $\{X_t,\dots,X_0,\dots,X_{-s+1},\lambda_0,\dots,\lambda_{-s+1}\}$. Then, for each $t \ge -s+1$, X_t follows a Poisson distribution, $Pois(\lambda_t)$, conditional on $\mathcal{I}_{t-1}^X$. For each $t \ge -s+1$, $\lambda_t = \lim_{N\to\infty}\sum_{i=1}^N p_{i,t}$ and satisfies, for $t > 0$,

$$\lambda_t = \bar{c} + \sum_{\tau=1}^s \beta_\tau \lambda_{t-\tau} + (X_{t-1},\dots,X_{t-s},\lambda_{t-1},\dots,\lambda_{t-s})\bar{\mathbf{f}} + \bar{\boldsymbol{\gamma}}^\mathsf{T} f_\gamma(X_{t-1},\dots,X_{t-s},\lambda_{t-1},\dots\lambda_{t-s}). \tag{21}$$

When $s = k = 1$, $f_{\alpha,i}(y) = a_i y$, $f_{\gamma,i} \equiv 0$, $f_\gamma(y,p) = y$, Theorem 8 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 (e.g., $y_{i,t-1} = 1$), 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 $\frac{1}{N}\boldsymbol{\gamma}_i^\top f_\gamma(\cdot)$ 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 $\boldsymbol{\gamma}_i$.

A distinctive feature of the aggregation in Theorems 4 and 8 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 9 below.

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

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

where λ_0 is some random variable. Then as $N \to \infty$, $[X_t(N)]_{t=0,1,\ldots}$ of Eq. (16) *converges in finite-dimensional distributions to a process $[X_t]_{t=0,1,\ldots}$.*

Denote by $\mathcal{I}_t^X$ the σ-field generated by $\{X_t, \ldots, X_0, \lambda_0\}$. Then, for each $t \geq 0$, X_t follows a Poisson distribution, $Pois(\lambda_t)$, conditional on $\mathcal{I}_{t-1}^X$. For each $t \geq 0$, $\lambda_t = \lim_{N\to\infty} \sum_{i=1}^N p_{i,t}$ and satisfies

$$\lambda_t = \bar{c} + \gamma X_{t-1} + \beta \lambda_{t-1}, \qquad t > 0. \tag{22}$$

A row-normalized network matrix W is doubly stochastic if, for example, for each i, $d_i^{in} = d_i^{out} = d$, where $d_i^{in} = \sum_{j=1}^N \mathcal{I}(W_{ji} \neq 0)$, $d_i^{out} = \sum_{j=1}^N \mathcal{I}(W_{ij} \neq 0)$ 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 (22) 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); 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. 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 $\mathbf{y}_0, \ldots, \mathbf{y}_{-q+1}, \mathbf{p}_0, \ldots, \mathbf{p}_{-s+1}$ take the form

$$\begin{aligned}\mathcal{L}_T(\theta) =& \prod_{t=1}^T \prod_{i=1}^N g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\ldots,t-s},\, \{\mathbf{y}_\tau\}_{\tau=t-1,\ldots,t-q}; \theta\right)^{y_{i,t}} \\ & \cdot \prod_{t=1}^T \prod_{i=1}^N \left(1 - g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\ldots,t-s},\, \{\mathbf{y}_\tau\}_{\tau=t-1,\ldots,t-q}; \theta\right)\right)^{1-y_{i,t}},\end{aligned} \tag{23}$$

where $\mathbf{p}_t$ is recursively defined by

$$\mathbf{p}_t = \mathbf{g}\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\ldots,t-s},\, \{\mathbf{y}_\tau\}_{\tau=t-1,\ldots,t-q}; \theta\right) \quad \text{for all } t,$$

$$
\begin{aligned}
\mathcal{Q}_T(\theta) := \log \mathcal{L}_T(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{N}\Big[& y_{i,t}\log g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q};\theta\right) \\
& + (1-y_{i,t})\log\left(1 - g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q};\theta\right)\right)\Big], \\
\mathbf{p}_t = \mathbf{g}\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q};\theta\right) & \quad \text{for all } t.
\end{aligned} \tag{24}
$$

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 $\mathbf{y}_t$. Lipschitz-type contraction in Assumptions I, II ensures that asymptotically this initialization bias disappears.[2]

Assumption III.

(i) *$\theta_0 \in \Theta$ for some compact set $\Theta \in \mathbb{R}^m$;*
(ii) *$\mathbf{g}\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q};\theta\right)$ 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) *If*

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

then $\theta = \theta_0$.

Continuity of $\mathbf{g}$ in θ and its mapping into $[0,1]^N$ imply that the parameter space Θ is closed. Boundedness, and hence compactness, is trivially satisfied in all linear GAB examples. For general nonlinear $\mathbf{g}$, compact support can be relaxed if, with positive probability, $\mathbf{g}\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q};\theta\right)$ goes to 0 or 1 as $\theta \to \infty$. This ensures that the log-likelihood is never maximized at unbounded values of θ.

Assumption III(iii) is required to ensure that the log-likelihood is well-defined. 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.

[2]Let $\mathbf{p}_t$ and $\mathbf{p}'_t$ be two trajectories with different initial probability values (but with the same sequence of outcomes $\mathbf{y}$). Let $\ell_{i,t}(p) := y_{i,t}\log p + (1-y_{i,t})\log(1-p)$. Since the trajectory of $\mathbf{y}_t$ is the same, both Assumptions I and II imply that for all i,t, $|p_{i,t} - p'_{i,t}| \le C\rho^t$ for some $C > 0, \rho \in [0,1)$. Given that $g_i(\cdot;\theta) \in [\varepsilon, 1-\varepsilon]$, as in Assumption III(iii), $|\ell_{i,t}(p_{i,t}) - \ell_{i,t}(p'_{i,t})| \le \frac{1}{\varepsilon}|p_{i,t} - p'_{i,t}| \le \frac{C}{\varepsilon}\rho^t$ and $\frac{1}{T}\sum_{i,t}|\ell_{i,t}(p_{i,t}) - \ell_{i,t}(p'_{i,t})| \le \frac{CN}{T\varepsilon}\frac{1}{1-\rho} \xrightarrow[T\to\infty]{} 0$ uniformly in θ belonging to a compact.

Assumption III(iv) requires the existence of a unique parameter value θ that is compatible with the data, i.e., identification of θ_0. This condition is again satisfied for linear models, since the stationary distribution of $\mathbf{p}_t$ is continuous (when it is not degenerate).

The maximum likelihood estimator solves

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

Theorem 10. *Suppose that the multivariate binary time series model* (1)–(2) *satisfies Assumption III and either Assumption I or II. Then* $\hat{\theta}^{MLE} \xrightarrow[T\to\infty]{p} \theta_0$.

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) $\mathbf{g}\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q};\theta\right)$ *is twice continuously differentiable on a neighborhood of* θ_0*;*
(iii) $H_0 = \mathbb{E}\sum_{i=1}^{N} \frac{1}{g_{i,t,\theta_0}(1-g_{i,t,\theta_0})}\left(\frac{\partial}{\partial\theta}g_{i,t,\theta_0}\right)\left(\frac{\partial}{\partial\theta}g_{i,t,\theta_0}\right)^{\mathsf{T}}$ *is nonsingular, where*

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

Theorem 11. *Suppose that the multivariate binary time series model* (1)–(2) *satisfies Assumptions III, IV, and either I or II. Then* $\sqrt{T}(\hat{\theta}^{MLE}-\theta_0) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, H_0^{-1})$.

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 (24). First, in line with $\max_p p_{i,t} \xrightarrow[T\to\infty]{p} 0$, we expect

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

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

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

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

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

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

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 $\mathbf{z}_t \in \mathbb{R}^k$, so that

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

Provided $\mathbf{z}_t$ is strictly stationary and geometrically 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 10 and 11 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 $\frac{1}{N}\delta_i^{\mathsf{T}}\mathbf{z}_t$ to each $p_{i,t}$ equation in interactive and network GAB models. Letting $\bar{\delta} := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \delta_i$, this contributes an extra term $\bar{\delta}^{\mathsf{T}}\mathbf{z}_t$ 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 $\mathbf{z}_t = (\mathbf{z}_t^{com}, \mathbf{z}_t^{ind})$, where $\mathbf{z}_t^{com} \in \mathbb{R}^{r_c}$ is a common component of fixed dimension affecting all units, and $\mathbf{z}_t^{ind} = (\mathbf{z}_{1,t}^{ind}, \dots, \mathbf{z}_{N,t}^{ind}) \in \mathbb{R}^{Nr}$ stacks unit-specific components with $\mathbf{z}_{i,t}^{ind} \in \mathbb{R}^r$, 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 $\mathbf{z}_{i,t}^{ind}$ directly and on the cross-sectional average $\frac{1}{N}\sum_{j=1}^N \mathbf{z}_{j,t}^{ind}$ (or the weighted average $\sum_{j=1}^N W_{ij}\mathbf{z}_{j,t}^{ind}$ in the network GAB case, as in Theorem 9). The direct effect of $\mathbf{z}_{i,t}^{ind}$ 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 $\eta\frac{1}{N}\sum_{i=1}^N \mathbf{z}_{i,t}^{ind}$. 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 $\mathbf{z}_{i,t}^{ind}$ is not reduced by aggregation; the corresponding term in $p_{i,t}$ is therefore $\frac{\zeta_i}{N}\frac{1}{N}\sum_{j=1}^N \mathbf{z}_{j,t}^{ind}$ (or nonlinear bounded[3] function $\frac{\zeta_i}{N} f_z\left(\frac{1}{N}\sum_{j=1}^N \mathbf{z}_{j,t}^{ind}\right)$, so that f_z survives in the limit analogously to the f_γ term in Theorem 8). Thus, assuming that $\frac{1}{N}\sum_{i=1}^N \mathbf{z}_{i,t}^{ind} \xrightarrow[N\to\infty]{p} \bar{\mathbf{z}}_t$, $\mathbb{E}\frac{1}{N}\sum_{i=1}^N \mathbf{z}_{i,t}^{ind} \xrightarrow[N\to\infty]{} \mathbb{E}\bar{\mathbf{z}}_t < \infty$, and $\max_i \|\mathbf{z}_{i,t}^{ind}\|/N \xrightarrow[N\to\infty]{p} 0$ for each t, the aggregated covariate $\bar{\mathbf{z}}_t$ appears on the right-hand side of the limiting evolution equation for λ_t.

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 87 S&P 100 constituents available over the full period 01.01.2005–01.01.2025, yielding $5,032$ time series observations. The 87 stocks are those for which data are available throughout the entire span; we also retain only one of the

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

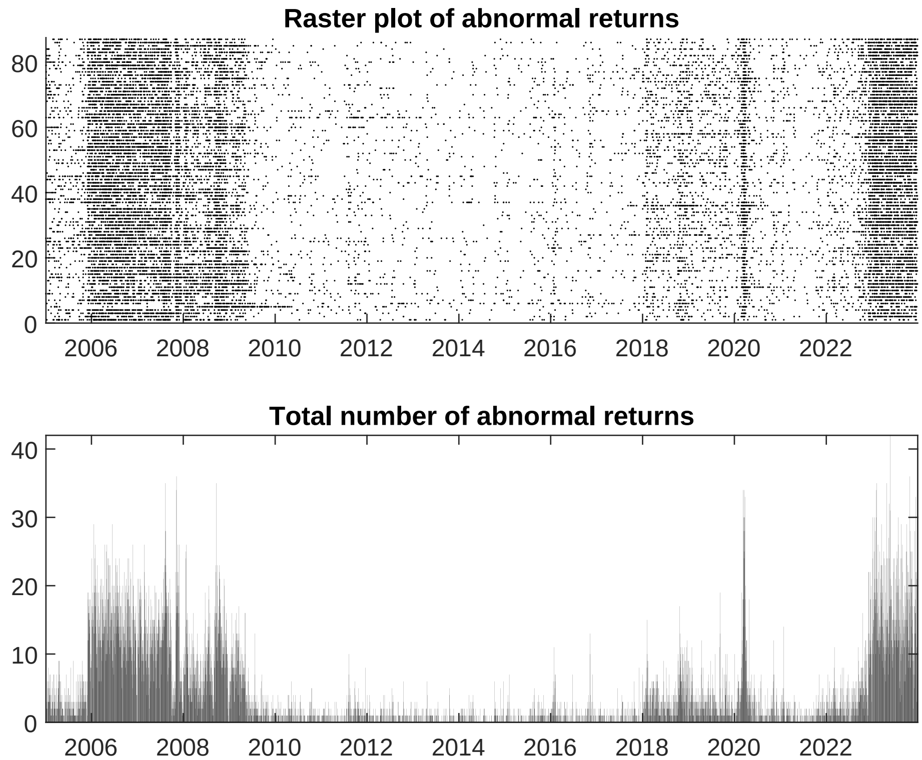


FIGURE 3. Abnormally low (below 5%) returns in S&P 100.

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 final year of data is reserved for forecast evaluation, so estimation uses observations up to 01.01.2024, giving $T = 4,780$. The 5th percentile threshold for each stock is 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 the crisis periods: financial crisis, COVID, and 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 small, with $\text{mean}(\omega_i) = 0.0006$, $\text{mean}(\alpha_i) = 0.027$. The averages of the remaining coefficients are $\text{mean}(\gamma_i) = 0.18$, $\text{mean}(\beta_i) = 0.78$.

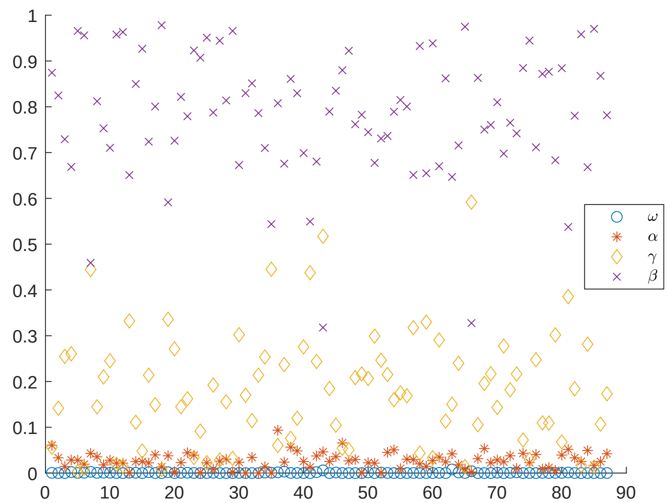


FIGURE 4. Interactive GAB estimates for 87 stocks.

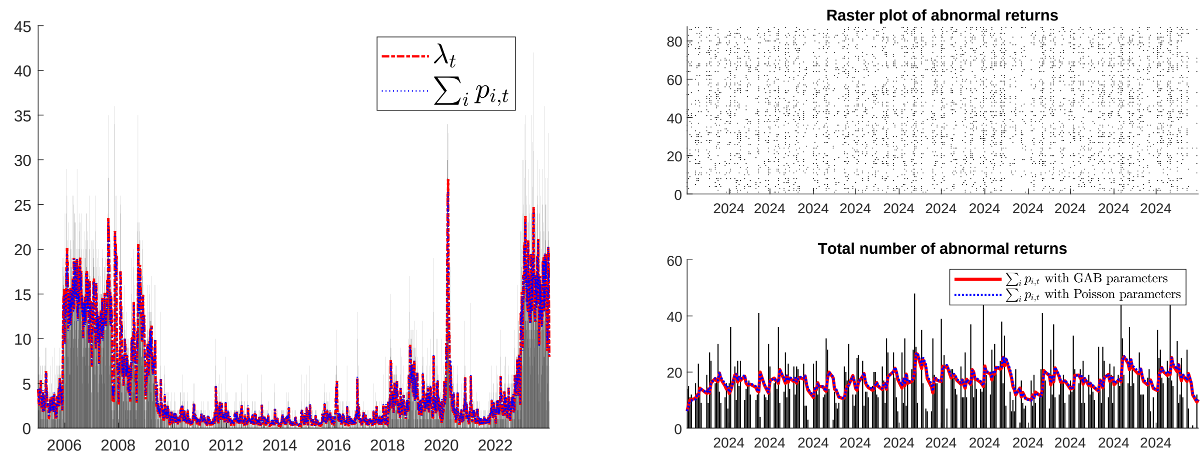
	$\sum_i \omega_i$	mean(ω_i)	mean(α_i)	mean(γ_i)	mean(β_i)
Interactive GAB	0.056	0.0006	0.027	0.18	0.78
Interactive GAB with $\alpha_i = 0$	0.098	0.001	–	0.25	0.72
	$\bar{c}$			$\bar{\gamma}$	β
Aggregated Poisson	0.049	–	–	0.24	0.75

TABLE 1. Estimation Results

We then 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. They are broadly consistent with the aggregation theorems, with the Poisson coefficients close to the averages of the GAB estimates. 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 overall similarity is further supported by the close agreement between the filtered Poisson intensity λ_t and the sum of GAB probabilities $\sum_i p_{i,t}$, as shown in Figure 5a.

Finally, we compare the forecasting performance of the GAB model using the one-year sample that was excluded from the estimation stage. The out-of-sample binary data are constructed by comparing returns to the same 5th percentile threshold used previously. We generate recursive 1-step-ahead forecasts of $y_{i,t}$ over the out-of-sample period $(T+1, \ldots, T+252$, one year of trading days) based on four models.

The first model is the interactive GAB specification without α_i (heterogeneous parameters). We impose $\alpha_i = 0$ to align the GAB specification with the Poisson limit structure. The second is an interactive GAB model with homogeneous parameters calibrated from the



(A) In-sample fit.

(B) Out-of-sample fit.

FIGURE 5. Interactive GAB probabilities $\sum_i p_{i,t}$ and Poisson intensity λ_t.

Poisson maximum likelihood estimates, i.e., $\omega_i = \bar{c}/N$, $\gamma_i = \bar{\gamma}$, 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 interactive GAB with heterogeneous parameters, achieves the best forecasting performance. Although the GAB and Poisson estimation results in Table 1 are close, the second (homogeneous) model performs slightly worse. Both naive benchmarks are substantially outperformed.

Figure 5b presents the out-of-sample data, both aggregated and disaggregated, together with the path 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 out-of-sample period features a large number of abnormally low returns (21.3% of the data points[4]), indicating elevated uncertainty, likely reflecting elevated macroeconomic uncertainty in 2024. This highlights the importance of the interaction term $\frac{1}{N}\sum_i y_{i,t}$ in capturing overall market conditions; the term which is absent in alternative third and fourth models.

[4]This is consistent with MSE results. For the fourth model, the squared forecast error is equal to 0 when two consecutive realizations coincide and 1 when they differ. If the observations were i.i.d. Bernoulli, the latter would occur with probability $2p(1-p) \approx 0.32$ for $p \approx 0.2$. For the third model, the squared forecast error takes values 0.05^2 and 0.95^2, and $MSE \approx 0.2 \cdot 0.05^2 + 0.8 \cdot 0.95^2 \approx 0.18$. For the first two models, $MSE \approx 0.16$ is close to the optimal value, which corresponds to the forecast $\hat{y}_{i,t} \approx 0.2$ yielding $MSE = p(1-p)^2 + (1-p)p^2 \approx 0.16$.

Model 1	Model 2	Model 3	Model 4
0.1618	0.1624	0.1844	0.3180

Table 2. Out-of-sample MSE (averaged across stocks and the out-of-sample period). Model 1: Interactive (heterogeneous) GAB with $\alpha_i = 0$. Model 2: Interactive (homogeneous) GAB with Poisson parameters. Model 3: 5% probability of success. Model 4: Tomorrow is the same as today.

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 geometric 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 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 are going to 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 $f: X\to\mathbb{R}$. The transition kernel maps

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

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). □

Proof of Theorem 3. Consider two stationary solutions coupled on the same probability space (i.e., use the same errors $u_{it}\sim i.i.d.\ U[0,1]$ for both trajectories). The two initial conditions are $\mathbf{p}:=\{\mathbf{p}_0,\dots,\mathbf{p}_{-s+1}\}$, $\mathbf{y}:=\{\mathbf{y}_0,\dots,\mathbf{y}_{-q+1}\}$ and $\mathbf{p}':=\{\mathbf{p}'_0,\dots,\mathbf{p}'_{-s+1}\}$, $\mathbf{y}':=\{\mathbf{y}'_0,\dots,\mathbf{y}'_{-q+1}\}$ (can be assumed independent in the coupling). First, observe that

$$
\begin{aligned}
\mathbb{E}_{t-1}|y_{i,t}-y'_{i,t}| &= \mathbb{E}_{t-1}\left|I(u_{i,t}\le p_{i,t})-I(u_{i,t}\le p'_{i,t})\right|\\
&=\mathbb{E}_{t-1}I(p'_{i,t}<u_{i,t}\le p_{i,t}\text{ or }p_{i,t}<u_{i,t}\le p'_{i,t})\\
&=\mathbb{E}_{t-1}I(\min(p_{i,t},p'_{i,t})<u_{i,t}\le\max(p_{i,t},p'_{i,t}))\\
&=\max(p_{i,t},p'_{i,t})-\min(p_{i,t},p'_{i,t})=|p_{i,t}-p'_{i,t}|.
\end{aligned}\tag{25}
$$

Thus, assuming Assumption I holds,

(26)
$$
\begin{aligned}
\mathbb{E}|p_{i,t}-p'_{i,t}| &= \mathbb{E}\left|g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}\right)-g_i\left(\{\mathbf{p}'_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}'_\tau\}_{\tau=t-1,\dots,t-q}\right)\right| \\
&\le \mathbb{E}\left(\sum_{j=1}^{N}\sum_{\tau=1}^{s}|p_{j,t-\tau}-p'_{j,t-\tau}|\beta_{ij}^{\tau}+\sum_{j=1}^{N}\sum_{\tau=1}^{q}|y_{j,t-\tau}-y'_{j,t-\tau}|\alpha_{ij}^{\tau}\right) \\
&= \sum_{j=1}^{N}\sum_{\tau=1}^{s}\mathbb{E}|p_{j,t-\tau}-p'_{j,t-\tau}|\beta_{ij}^{\tau}+\sum_{j=1}^{N}\sum_{\tau=1}^{q}\mathbb{E}|p_{j,t-\tau}-p'_{j,t-\tau}|\alpha_{ij}^{\tau} \\
&= \sum_{j=1}^{N}\sum_{\tau=1}^{\max(s,q)}\mathbb{E}|p_{j,t-\tau}-p'_{j,t-\tau}|\left(\alpha_{ij}^{\tau}+\beta_{ij}^{\tau}\right),
\end{aligned}
$$

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

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

Then Eq. (26) implies

$$\eta_t\le\Phi\eta_{t-1}\le\Phi^t\eta_0. \tag{27}$$

By Assumption I, all eigenvalues of Φ lie inside the unit circle. Thus, $\Phi^t\to 0$ as $t\to\infty$, so that $\mathbb{E}|p_{i,t}-p'_{i,t}|\to 0$ for all i as $t\to\infty$. Thus, we must have $p_t-p'_t\xrightarrow[t\to\infty]{d}0$ and the distributional limits are the same. Geometric ergodicity is implied by the bound $\Phi^t\eta_0$ in (27).

Similarly, if instead we assume that Assumption II holds,

(28)
$$
\begin{aligned}
\mathbb{E}\sum_{i=1}^{N}|p_{i,t}-p'_{i,t}| &= \mathbb{E}\sum_{i=1}^{N}\left|g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}\right)-g_i\left(\{\mathbf{p}'_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}'_\tau\}_{\tau=t-1,\dots,t-q}\right)\right| \\
&\le K\mathbb{E}\left(\sum_{i=1}^{N}\sum_{\tau=1}^{s}|p_{i,t-\tau}-p'_{i,t-\tau}|+\sum_{i=1}^{N}\sum_{\tau=1}^{q}|y_{i,t-\tau}-y'_{i,t-\tau}|\right) \\
&= K\sum_{\tau=1}^{s}\mathbb{E}\sum_{i=1}^{N}|p_{i,t-\tau}-p'_{i,t-\tau}|+K\sum_{\tau=1}^{q}\mathbb{E}\sum_{i=1}^{N}|p_{i,t-\tau}-p'_{i,t-\tau}|.
\end{aligned}
$$

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

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

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,

$$\tilde{\eta}_t \le \tilde{\Phi}\tilde{\eta}_{t-1} \le \tilde{\Phi}^t \tilde{\eta}_0 \to 0 \quad \text{as } t \to \infty$$

and we obtain geometric ergodicity as well as $p_t - p'_t \xrightarrow[t\to\infty]{d} 0$. □

Appendix B. Proofs: Aggregation

Lemma 12. *Consider a sequence of random probabilities $q_i, i = 1,\dots,N$, such that*

$$\sum_{i=1}^N q_i \xrightarrow[N\to\infty]{d} \xi, \qquad \sum_{i=1}^N \mathbb{E}q_i \xrightarrow[N\to\infty]{} \mathbb{E}\xi, \qquad \max_{i=1,\dots,N} q_i \xrightarrow[N\to\infty]{p} 0,$$

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 $N \to \infty$, the sum $\sum_{i=1}^N y_i$ converges to a random variable Y which follows a Poisson distribution 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

$$G_N(z) := \mathbb{E}z^{S_N} = \prod_{i=1}^N (q_i z + (1-q_i)) = \prod_{i=1}^N (1 + q_i(z-1))$$

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

Proof of Theorem 4. At time $t = 0$, the process satisfies conditions of Lemma 12, thus,

$$\sum_{i=1}^N y_{i,0} \xrightarrow[N\to\infty]{d} Y_0, \text{ with } Y_0 \sim Pois\left(\lim_{N\to\infty} \sum_{i=1}^N p_{i,0} \right) = Pois(\lambda_0).$$

Let us show that at time $t = 1$, the process also satisfies conditions of Lemma 12, so that

$$\sum_{i=1}^{N} y_{i,1} \xrightarrow[N\to\infty]{d} Y_1, \text{ with } Y_1 \sim Pois\left(\lim_{N\to\infty}\sum_{i=1}^{N} p_{i,1}\right) = Pois(\lambda_1).$$

By Eq. (8),

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

By Markov inequality for any $\varepsilon > 0$

$$P\left(\frac{1}{N}\sum_{j=1}^{N} y_{j,0} > \varepsilon\right) \leq \frac{1}{\varepsilon}\mathbb{E}\frac{1}{N}\sum_{j=1}^{N} y_{j,0} = \frac{1}{N\varepsilon}\sum_{j=1}^{N}\mathbb{E}p_{j,0} \xrightarrow[N\to\infty]{} 0.$$

Thus, since by assumptions, $\max_i p_{i,0} \to 0$ and $0 \leq \gamma_i \leq 1$, and, by $t = 0$ result, $\frac{1}{N}\sum_{j=1}^{N} y_{j,0} \xrightarrow[N\to\infty]{p} 0$,

$$0 \leq \max_i p_{i,1} \leq \frac{C}{N} + \frac{C}{N^\kappa} + \beta \max_i p_{i,0} + \frac{1}{N}\sum_{j=1}^{N} y_{j,0} \to 0.$$

Summing Eq. (8) with resect to i for $t = 1$,

$$\sum_{i=1}^{N} p_{i,1} = \frac{1}{N}\sum_{i=1}^{N} c_i + \frac{1}{N^\kappa}\sum_{i=1}^{N} a_i y_{i,0} + \beta\sum_{i=1}^{N} p_{i,0} + \frac{1}{N}\sum_{j=1}^{N} y_{j,0}\sum_{i=1}^{N}\gamma_i \xrightarrow[N\to\infty]{d} \bar{c} + \beta\lambda_0 + \bar{\gamma}Y_0 = \lambda_1$$

and

$$\sum_{i=1}^{N}\mathbb{E}p_{i,1} = \frac{1}{N}\sum_{i=1}^{N} c_i + \frac{1}{N^\kappa}\sum_{i=1}^{N} a_i\mathbb{E}y_{i,0} + \beta\sum_{i=1}^{N}\mathbb{E}p_{i,0} + \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}y_{j,0}\sum_{i=1}^{N}\gamma_i$$
$$\xrightarrow[N\to\infty]{d} \bar{c} + \beta\mathbb{E}\lambda_0 + \bar{\gamma}\mathbb{E}Y_0 = \mathbb{E}\lambda_1.$$

Thus, time $t = 1$ satisfies desired conditions. Repeating the same argument, by induction at any time t, $p_{i,t}$ satisfies conditions of Lemma 12, so that

$$\sum_{i=1}^{N} y_{i,t} \xrightarrow[N\to\infty]{d} Y_t,\ Y_t \sim Pois\left(\lim_{N\to\infty}\sum_{i=1}^{N} p_{i,t}\right) = Pois(\lambda_t).$$

Finally, summing (8) across i at time t, we get

$$\sum_{i=1}^{N} p_{i,t} = \frac{1}{N}\sum_{i=1}^{N} c_i + \left(\frac{1}{N}\sum_{i=1}^{N}\gamma_i\right)\sum_{i=1}^{N} y_{i,t-1} + \frac{1}{N^\kappa}\sum_{i=1}^{N} a_i y_{i,t-1} + \beta\sum_{i=1}^{N} p_{i,t-1}. \tag{29}$$

and taking the limit as $N \to \infty$ in (29), we get

$$\lambda_t = \bar{c} + \bar{\gamma}Y_{t-1} + \beta\lambda_{t-1}.$$

□

Proof of Proposition 7. Consider a stationary distribution and let us denote it as $\mathbf{p}^* = (p_1^*, \ldots, p_N^*)^\intercal$ with corresponding realizations $\mathbf{y}^* = (y_1^*, \ldots, y_N^*)^\intercal$. Define $v_{i,t} := y_{i,t}^* - \mathbb{E}_{t-1} y_{i,t}^* \equiv y_{i,t}^* - p_{i,t}^*$. Then $v_{i,t}$ is a zero-mean white noise process and $\mathbb{E} v_{i,t} v_{j,t} = 0$ for $i \neq j$ (since $u_{i,t}$ and $u_{j,t}$ are independent). Define two $N \times N$ matrices Φ and Π such that

$$\Pi = \begin{pmatrix} \alpha_1 + \frac{1}{N}\gamma_1 & \frac{1}{N}\gamma_1 & \ldots & \frac{1}{N}\gamma_1 \\ \frac{1}{N}\gamma_2 & \alpha_2 + \frac{1}{N}\gamma_2 & \ldots & \frac{1}{N}\gamma_2 \\ \ldots & \ldots & \ldots & \ldots \\ \frac{1}{N}\gamma_N & \frac{1}{N}\gamma_N & \ldots & \alpha_N + \frac{1}{N}\gamma_N \end{pmatrix},$$

$$\Phi = \begin{pmatrix} \alpha_1 + \beta + \frac{1}{N}\gamma_1 & \frac{1}{N}\gamma_1 & \ldots & \frac{1}{N}\gamma_1 \\ \frac{1}{N}\gamma_2 & \alpha_2 + \beta + \frac{1}{N}\gamma_2 & \ldots & \frac{1}{N}\gamma_2 \\ \ldots & \ldots & \ldots & \ldots \\ \frac{1}{N}\gamma_N & \frac{1}{N}\gamma_N & \ldots & \alpha_N + \beta + \frac{1}{N}\gamma_N \end{pmatrix} = \Pi + \beta I_N,$$

where I_N is an $N \times N$ identity matrix. Then, letting $\boldsymbol{\omega} := (\omega_1, \ldots, \omega_N)^\intercal$ and $\mathbf{v}_t = (v_{1,t}, \ldots, v_{N,t})^\intercal$, the stationary solution satisfies

$$\mathbf{p}_t^* = \boldsymbol{\omega} + \Phi \mathbf{p}_{t-1}^* + \Pi \mathbf{v}_{t-1} = \boldsymbol{\mu} + \sum_{k=0}^{\infty} \Phi^k \Pi \mathbf{v}_{t-1-k} \tag{30}$$

and

$$\sum_{i=1}^{N} p_{i,t}^* = \sum_{i=1}^{N} \mu_i + (1, \ldots, 1) \sum_{k=0}^{\infty} \Phi^k \Pi \mathbf{v}_{t-1-k}.$$

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

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

$$c_i (1 - \beta - a_i/N^\kappa)^{-1} \approx c_i (1-\beta)^{-1} \left(1 + \frac{a_i}{(1-\beta)N^\kappa}\right)$$

and

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

Next, let us analyze the variance of $\sum_{i=1}^{N} p^*_{i,t}$. First, note that $v_{i,t} \in [-1,1]$, so it has finite variance and

$$\mathbb{E}v_{i,t}^2 = \mathbb{E}(y^*_{i,t} - p^*_{i,t})^2 = \mathbb{E}y^*_{i,t} - \mathbb{E}(p^*_{i,t})^2 = \mu_i - \mathbb{E}(p^*_{i,t})^2 = \mu_i(1-\mu_i) - \mathbb{V}p^*_{i,t}.$$

Let $\Omega := \mathbb{V}\mathbf{p}^*_t$, so that $\Omega_{ij} := Cov(p^*_{i,t}, p^*_{j,t})$. Because $\mathbf{v}_t$ and $\mathbf{p}_t$ are uncorrelated, Eq. (30) implies

$$\Omega = \Phi\Omega\Phi^\intercal + \Pi diag(\mu_1(1-\mu_1) - \Omega_{11}, \dots, \mu_N(1-\mu_N) - \Omega_{NN})\Pi^\intercal. \tag{31}$$

Note that $\mathbb{V}\sum_{i=1}^{N} p^*_{i,t} = 1_N^\intercal \Omega 1_N$, where $1_N = (1,\dots,1)^\intercal$, so in order to analyze the variance one can study at $1_N^\intercal\Phi$ and $1_N^\intercal\Pi$:

(32)

$$1_N^\intercal\Pi = \left(\alpha_1 + \frac{1}{N}\sum_{i=1}^{N}\gamma_i, \dots, \alpha_N + \frac{1}{N}\sum_{i=1}^{N}\gamma_i\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\gamma_i + O(N^{-\kappa})\right)1_N^\intercal,$$

$$1_N^\intercal\Phi = \left(\alpha_1 + \beta + \frac{1}{N}\sum_{i=1}^{N}\gamma_i, \dots, \alpha_N + \beta + \frac{1}{N}\sum_{i=1}^{N}\gamma_i\right) = \left(\beta + \frac{1}{N}\sum_{i=1}^{N}\gamma_i + O(N^{-\kappa})\right)1_N^\intercal.$$

Plugging (32) into (31) premultiplied by $1_N^\intercal$ from the left and 1_N from the right, we get

$$\begin{aligned}&\left(1 - \left(\beta + \frac{1}{N}\sum_{i=1}^{N}\gamma_i + O(N^{-\kappa})\right)^2\right)1_N^\intercal\Omega 1_N\\ &= \left(\frac{1}{N}\sum_{i=1}^{N}\gamma_i + O(N^{-\kappa})\right)^2 \sum_{i=1}^{N}\left(\mu_i(1-\mu_i) - \Omega_{ii}\right).\end{aligned} \tag{33}$$

Let us show that Eq. (33) implies $1_N^\intercal\Omega 1_N \xrightarrow[N\to\infty]{} \frac{\bar\gamma^2}{1-(\beta+\bar\gamma)^2}\lim_{N\to\infty}\sum_{i=1}^{N}\mu_i = \frac{\bar c\bar\gamma^2}{(1-(\beta+\bar\gamma)^2)(1-\beta-\bar\gamma)}$, where the denominator is positive, since $\beta + \gamma_i \le \bar C < 1$ for all i. Eq. (8) implies that each μ_i is (uniformly) of order $1/N$, so that each μ_i^2 is of order $1/N^2$ and $\sum_i \mu_i^2 \xrightarrow[N\to\infty]{} 0$. We are left with analyzing $\sum_i \Omega_{ii}$. Let us apply the variance operator to Eq. (8), noting that $\mathbb{E}y_{i,t}^2 \equiv \mathbb{E}y_{i,t} = \mu_i$, $\mathbb{E}y_{i,t}p_{j,t} = \mathbb{E}p_{i,t}p_{j,t}$, $\mathbb{E}y_{i,t}p_{k,t} = \mathbb{E}p_{i,t}p_{k,t}$ for all i,j and $k \neq i$:

$$\begin{aligned}\Omega_{ii} = {}&\mu_i(1-\mu_i)\left(\alpha_i^2 + \frac{2}{N}\alpha_i\gamma_i\right) + \frac{1}{N^2}\gamma_i^2\sum_{j=1}^{N}\mu_j(1-\mu_j) + \Omega_{ii}(\beta^2 + 2\alpha_i\beta)\\ &+ \frac{1}{N^2}\gamma_i^2\sum_{k\neq j}\Omega_{kj} + \frac{2}{N}\gamma_i\beta\sum_{j=1}^{N}\Omega_{ij} + \frac{2}{N}\gamma_i\alpha_i\left(\sum_{j=1}^{N}\Omega_{ij} - \Omega_{ii}\right).\end{aligned}$$

Summing the above with respect to i we obtain

$$\begin{aligned}\sum_{i=1}^N \Omega_{ii} = &\sum_{i=1}^N \mu_i(1-\mu_i)\left(\alpha_i^2 + \frac{2}{N}\alpha_i\gamma_i\right) + \left(\sum_{i=1}^N \frac{1}{N^2}\gamma_i^2\right)\sum_{j=1}^N \mu_j(1-\mu_j) \\ &+ \left(\beta^2 + O(N^{-\kappa})\right)\sum_{i=1}^N \Omega_{ii} + \left(\sum_{i=1}^N \frac{1}{N^2}\gamma_i^2\right)\sum_{i,j=1}^N \Omega_{ij} - \left(\sum_{i=1}^N \frac{1}{N^2}\gamma_i^2\right)\sum_{i=1}^N \Omega_{ii} \\ &+ \frac{2}{N}\sum_{i,j=1}^N \gamma_i(\beta+\alpha_i)\Omega_{ij} - \frac{2}{N}\sum_{i=1}^N \gamma_i\alpha_i\Omega_{ii}.\end{aligned}$$

Since $\alpha_i = O(N^{-\kappa})$ and $0 \leq \gamma_i \leq 1$, $\sum_{i=1}^N \mu_i(1-\mu_i)\left(\alpha_i^2 + \frac{2}{N}\alpha_i\gamma_i\right) = o(1)$ and $\left(\sum_{i=1}^N \frac{1}{N^2}\gamma_i^2\right)\sum_{j=1}^N \mu_j(1-\mu_j) = o(1)$, so that we are left with

$$\left(1-\beta^2 + O(N^{-\kappa}) + O(N^{-1})\right)\sum_{i=1}^N \Omega_{ii} = o(1) + O(N^{-1})\sum_{i,j=1}^N \Omega_{ij}, \tag{34}$$

implying that either $\sum_{i=1}^N \Omega_{ii} = o(1)$ or $\sum_{i=1}^N \Omega_{ii} = O\left(\frac{1}{N}\sum_{i,j=1}^N \Omega_{ij}\right) = o\left(\sum_{i,j=1}^N \Omega_{ij}\right)$, depending on whichever is larger. Each case, if plugged into Eq. (33), guarantees that the term $\sum_{i=1}^N \Omega_{ii}$ is negligible, leading to the desired conclusion

$$\mathbb{V}\sum_{i=1}^N p_{i,t}^* = \sum_{i,j=1}^N \Omega_{ij} \xrightarrow[N\to\infty]{} \frac{\bar{c}\bar{\gamma}^2}{(1-(\beta+\bar{\gamma})^2)(1-\beta-\bar{\gamma})}.$$

Thus, the limit of

$$\sum_{i=1}^N p_{i,t}^* = \sum_{i=1}^N \mu_i + 1_N^\intercal \sum_{k=0}^\infty \Phi^k \Pi \mathbf{v}_{t-1-k}$$

is finite and corresponds to $\frac{\bar{c}}{1-\beta-\bar{\gamma}} + \bar{\gamma}\sum_{k=0}^\infty (\beta+\bar{\gamma})^k \bar{v}_{t-1-k}$, where $\bar{v}_t$ is a zero mean white noise process with variance $\frac{\bar{c}}{1-\beta-\bar{\gamma}}$. (Here we used (32) to simplify $1_N^\intercal \Phi^k \Pi$.)

We are left checking that $\max_{i=1,\dots,N} p_{i,t}^* \xrightarrow[N\to\infty]{p} 0$. To do this, note that, for any $\varepsilon > 0$,

$$\begin{aligned}Prob\left(\max_{i=1,\dots,N} p_{i,t}^* > \varepsilon\right) &\leq \sum_{i=1}^N Prob(p_{i,t}^* > \varepsilon) = \sum_{i=1}^N Prob\left((p_{i,t}^*)^2 > \varepsilon^2\right) \\ &\leq \frac{1}{\varepsilon^2}\sum_{i=1}^N \mathbb{E}(p_{i,t}^*)^2 = \frac{1}{\varepsilon^2}\sum_{i=1}^N (\Omega_{ii} + \mu_i^2) \xrightarrow[N\to\infty]{} 0,\end{aligned}$$

since μ_i^2 is (uniformly) $O(1/N^2)$ and $\sum_{i=1}^{N} \Omega_{ii}$ is $o(1)$, as was shown in (and in the discussion after) (34). □

Proof of Theorem 8. 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 $t = 0, \dots, -s+1$, we are exactly in the setting of Lemma 12, so the desired convergence is satisfied. Let us check that the conditions of Lemma 12 are satisfied at $t = 1$.

At $t = 1$, we know that $\frac{1}{N^m} \sum_{j=1}^{N} y_{j,\tau} \xrightarrow[N\to\infty]{p} 0$ and $\frac{1}{N^m} \sum_{j=1}^{N} p_{j,\tau} \xrightarrow[N\to\infty]{p} 0$ for $\tau = 0, \dots, -s+1$ and any $m > 0$. Since $f_{\gamma,i}(0, \dots, 0) = 0$ and $\nabla f_{\gamma,i}$ is uniformly bounded,

(35)
$$\left| f_{\gamma,i}\left(\frac{1}{N}\sum_{j=1}^{N} y_{j,t-1}, \dots, \frac{1}{N}\sum_{j=1}^{N} y_{j,t-s}, \frac{1}{N}\sum_{j=1}^{N} p_{j,t-1}, \dots, \frac{1}{N}\sum_{j=1}^{N} p_{j,t-s} \right) \right|$$
$$\le M \left(\frac{1}{N}\sum_{j=1}^{N} y_{j,t-1} + \dots + \frac{1}{N}\sum_{j=1}^{N} y_{j,t-s} + \frac{1}{N}\sum_{j=1}^{N} p_{j,t-1} + \dots + \frac{1}{N}\sum_{j=1}^{N} p_{j,t-s} \right) \xrightarrow[N\to\infty]{p} 0.$$

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

(36)
$$\frac{1}{N^\kappa} f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) \xrightarrow[N\to\infty]{p} 0 \quad \text{uniformly in } i, \qquad \frac{1}{N^\kappa}\sum_{i=1}^{N} f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) \xrightarrow[N\to\infty]{p} 0.$$

Finally, because f_γ is continuous,

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

so that combining it with $\|\boldsymbol{\gamma}_i\| < C$, we get

(37)
$$\left| \frac{1}{N}\boldsymbol{\gamma}_i^\mathsf{T} f_\gamma\left(\sum_{j=1}^{N} y_{j,0}, \dots, \sum_{j=1}^{N} y_{j,-s+1}, \sum_{j=1}^{N} p_{j,0}, \dots, \sum_{j=1}^{N} p_{j,-s+1} \right) \right|$$
$$\le \frac{C}{N}\left\| f_\gamma\left(\sum_{j=1}^{N} y_{j,0}, \dots, \sum_{j=1}^{N} y_{j,-s+1}, \sum_{j=1}^{N} p_{j,0}, \dots, \sum_{j=1}^{N} p_{j,-s+1} \right) \right\| \xrightarrow[N\to\infty]{p} 0 \quad \text{uniformly in } i.$$

Combining Eq. (35), (36), (37), and the assumption on the initial conditions $\max_i p_{i,\tau} \to 0$, $\tau = 0, \dots, -s+1$, we deduce

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

Summing Eq. (19) with resect to i for $t=1$, and using Eq. 36 and the fact that

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

we obtain

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

To get the convergence of expectations we need to show that

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

If f_γ is bounded, then uniform integrability holds, and we get the convergence of expectations. Now consider the case of continuously differentiable with bounded Jacobian f_γ. Let $Y_N := \left(\sum_{j=1}^{N} y_{j,0},\ldots,\sum_{j=1}^{N} y_{j,-s+1},\sum_{j=1}^{N} p_{j,0},\ldots,\sum_{j=1}^{N} p_{j,-s+1}\right)$ and $Y := (X_0,\ldots,X_{-s+1},\lambda_0,\ldots\lambda_{-s+1})$. Then, since the Jacobian of f_γ, J_{f_γ} is bounded

$$|f_\gamma(Y_N) - f_\gamma(Y)| \leq \|J_{f_\gamma}(\cdot)\|\|Y_N - Y\| \leq M\|Y_N - Y\|$$

and

$$|\mathbb{E}f_\gamma(Y_N) - \mathbb{E}f_\gamma(Y)| \leq \mathbb{E}|f_\gamma(Y_N) - f_\gamma(Y)| \leq M\mathbb{E}\|Y_N - Y\|.$$

To show that $\mathbb{E}\|Y_N - Y\| \xrightarrow[N\to\infty]{} 0$, note that all coordinates of Y_N are non-negative (sums of binary outcomes and nonegative probabilities). Then, by Fatou's lemma,

$$\mathbb{E}Y \le \lim_{N\to\infty} \inf \mathbb{E}Y_N.$$

Thus, for each of $2s$ coordinates of Y_N, $Y_N(\tau)$,

$$\mathbb{E}|Y_N(\tau) - Y(\tau)| = \mathbb{E}Y_N(\tau) + \mathbb{E}Y(\tau) - 2\mathbb{E}\min(Y_N(\tau), Y(\tau)) \xrightarrow[N\to\infty]{} 0,$$

where by the dominated convergence theorem $\min(Y_N(\tau), Y(\tau)) \le Y(\tau)$ and $\mathbb{E}\min(Y_N(\tau), Y(\tau)) \to \mathbb{E}Y(\tau)$. Therefore, $\sum_{\tau=1}^{2s} \mathbb{E}|Y_N(\tau) - Y(\tau)| \xrightarrow[N\to\infty]{} 0$ and by the equivalence of norms in $\mathbb{R}^{2s}$, $\mathbb{E}\|Y_N - Y\| \xrightarrow[N\to\infty]{} 0$. This gives us the desired convergence of expectations.

The convergence $\frac{1}{N^\kappa}\sum_{i=1}^{N} \mathbb{E}f_{\alpha,i}(y_{i,0}, \dots, y_{i,-s+1}) \to 0$ follows from Eq. (36). The convergence

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

follows from the same Taylor expansion as before.

To sum up,

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

By induction, repeating the same steps for $t > 1$, we obtain desired result. □

Proof of Theorem 9. 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.

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

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,

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

Therefore, repeating the steps of the proof of Theorem 4, we obtain desired result. □

Appendix C. Proofs: Estimation

Proof of Theorem 10. The population log-likelihood is

$$Q(\theta) = \sum_{i=1}^{N} Q_i(\theta),$$

(38)
$$Q_i(\theta) = \mathbb{E}\Big[y_{i,t} \log g_i \left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}; \theta \right) + (1 - y_{i,t}) \log \left(1 - g_i \left(\{\mathbf{p}_\tau\}_{\tau=t-1,\dots,t-s},\ \{\mathbf{y}_\tau\}_{\tau=t-1,\dots,t-q}; \theta \right) \right) \Big],$$

where the expectation is taken under stationary distribution with true parameter θ_0.

To shorten notation, let us write

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

Consider

(39)
$$Q_i(\theta) - Q_i(\theta_0) = \mathbb{E} \log \left(g_{i,t,\theta}^{y_{i,t}} (1 - g_{i,t,\theta})^{1-y_{i,t}} \right) - \mathbb{E} \log \left(g_{i,t,\theta_0}^{y_{i,t}} (1 - g_{i,t,\theta_0})^{1-y_{i,t}} \right)$$
$$= \mathbb{E} \log \frac{g_{i,t,\theta}^{y_{i,t}} (1 - g_{i,t,\theta})^{1-y_{i,t}}}{g_{i,t,\theta_0}^{y_{i,t}} (1 - g_{i,t,\theta_0})^{1-y_{i,t}}} \leq \mathbb{E} \frac{g_{i,t,\theta}^{y_{i,t}} (1 - g_{i,t,\theta})^{1-y_{i,t}}}{g_{i,t,\theta_0}^{y_{i,t}} (1 - g_{i,t,\theta_0})^{1-y_{i,t}}} - 1 = \mathbb{E}\mathbb{E}_{t-1} \frac{g_{i,t,\theta}^{y_{i,t}} (1 - g_{i,t,\theta})^{1-y_{i,t}}}{g_{i,t,\theta_0}^{y_{i,t}} (1 - g_{i,t,\theta_0})^{1-y_{i,t}}} - 1$$
$$= \mathbb{E} \left(\frac{g_{i,t,\theta}}{g_{i,t,\theta_0}} g_{i,t,\theta_0} + \frac{1 - g_{i,t,\theta}}{1 - g_{i,t,\theta_0}} (1 - g_{i,t,\theta_0}) \right) - 1 \equiv 1 - 1 = 0,$$

where we used the inequality $\log(a) \leq a - 1$ for any $a \geq 0$. Thus, θ_0 is a maximum of $Q(\theta)$. The value of $Q(\theta_0)$ is finite, since

$$Q_i(\theta_0) = \mathbb{E}\left[y_{i,t}\log g_{i,t,\theta_0} + (1-y_{i,t})\log\left(1-g_{i,t,\theta_0}\right)\right] = \mathbb{E}\mathbb{E}_{t-1}\left[y_{i,t}\log g_{i,t,\theta_0} + (1-y_{i,t})\log\left(1-g_{i,t,\theta_0}\right)\right]$$
$$= \mathbb{E}\left[g_{i,t,\theta_0}\log g_{i,t,\theta_0} + (1-g_{i,t,\theta_0})\log\left(1-g_{i,t,\theta_0}\right)\right]$$

and $\lim_{a\to 0+} a\log a = 0$. Moreover, θ_0 is the unique maximum of $Q(\theta)$, because equality in (39) is only possible if $\frac{g_{i,t,\theta}^{y_{i,t}}(1-g_{i,t,\theta})^{1-y_{i,t}}}{g_{i,t,\theta_0}^{y_{i,t}}(1-g_{i,t,\theta_0})^{1-y_{i,t}}} = 1$ with probability 1. So equality requires $g_{i,t,\theta} = g_{i,t,\theta_0}$ for all i (because $y_{i,t}$ takes both values 0 and 1 with positive probability under Assumption III(iii)). This cannot hold by Assumption III(iv) if $\theta \neq \theta_0$.

Next, we need to verify that the sample log-likelihood converges uniformly in probability to the population one. Note that

$$y_{i,t}\log g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1}^{t-s},\, \{\mathbf{y}_\tau\}_{\tau=t-1}^{t-q};\theta\right) + (1-y_{i,t})\log\left(1-g_i\left(\{\mathbf{p}_\tau\}_{\tau=t-1}^{t-s},\, \{\mathbf{y}_\tau\}_{\tau=t-1}^{t-q};\theta\right)\right)$$

is a continuous function of θ (by Assumption III(iii) both $|\log g_i|$ and $|\log(1-g_i)|$ are bounded by $|\log\varepsilon|$). Combining this with the fact that Θ is compact, we infer that

$$\mathbb{E}\sup_{\theta\in\Theta}\left|y_{i,t}\log g_{i,t,\theta} + (1-y_{i,t})\log(1-g_{i,t,\theta})\right| < \infty.$$

Therefore, by Hayashi (2000, Theorem 7.3), $\hat{\theta}^{MLE} \xrightarrow[T\to\infty]{p} \theta_0$. □

Proof of Theorem 11. Define the score function and the Hessian

$$s_t(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^N\left[y_{i,t}\log g_{i,t,\theta} + (1-y_{i,t})\log\left(1-g_{i,t,\theta}\right)\right] = \sum_{i=1}^N\left[\frac{y_{i,t}}{g_{i,t,\theta}} - \frac{1-y_{i,t}}{1-g_{i,t,\theta}}\right]\frac{\partial}{\partial\theta}g_{i,t,\theta} \tag{40}$$

(41)
$$H_t(\theta) = \frac{\partial^2}{\partial\theta\partial\theta^\mathsf{T}}\sum_{i=1}^N\left[y_{i,t}\log g_{i,t,\theta} + (1-y_{i,t})\log\left(1-g_{i,t,\theta}\right)\right]$$
$$= \sum_{i=1}^N\left[\left(\frac{y_{i,t}}{g_{i,t,\theta}} - \frac{1-y_{i,t}}{1-g_{i,t,\theta}}\right)\frac{\partial^2}{\partial\theta\partial\theta^\mathsf{T}}g_{i,t,\theta} - \left(\frac{y_{i,t}}{g_{i,t,\theta}^2} + \frac{1-y_{i,t}}{(1-g_{i,t,\theta})^2}\right)\left(\frac{\partial}{\partial\theta}g_{i,t,\theta}\right)\left(\frac{\partial}{\partial\theta}g_{i,t,\theta}\right)^\mathsf{T}\right]$$

Because $g_{i,t,\theta}$ is twice continuously differentiable on a neighborhood of θ_0, $g_{i,t,\theta} \in [\varepsilon, 1-\varepsilon]$ and all arguments belong to compact spaces, both the score and the Hessian are uniformly bounded. Moreover, for $\theta = \theta_0$, $\mathbb{E}_{t-1}y_{i,t} = g_{i,t,\theta_0}$, and

$$\mathbb{E}_{t-1}s_t(\theta_0) = \sum_{i=1}^N\frac{\partial}{\partial\theta}g_{i,t,\theta_0}\mathbb{E}_{t-1}\left(\frac{y_{i,t}}{g_{i,t,\theta_0}} - \frac{1-y_{i,t}}{1-g_{i,t,\theta_0}}\right) = 0,$$

meaning that $s_t(\theta_0)$ is a martingale difference sequence. Therefore, $\mathbb{E}s_t(\theta_0) = 0$ and

$$\forall \tau > 0 \qquad \mathbb{E}s_t(\theta_0)s_{t-\tau}(\theta_0)^\intercal = 0,$$

$$\begin{aligned}\mathbb{E}s_t(\theta_0)s_t(\theta_0)^\intercal &= \sum_{i=1}^{N} \mathbb{E}\frac{\partial}{\partial\theta}g_{i,t,\theta_0}\frac{\partial}{\partial\theta^\intercal}g_{i,t,\theta_0}\mathbb{E}_{t-1}\left(\frac{y_{i,t}}{g_{i,t,\theta_0}} - \frac{1-y_{i,t}}{1-g_{i,t,\theta_0}}\right)^2 \\ &+ \sum_{i\neq j} \mathbb{E}\frac{\partial}{\partial\theta}g_{i,t,\theta_0}\frac{\partial}{\partial\theta^\intercal}g_{j,t,\theta_0}\mathbb{E}_{t-1}\left(\frac{y_{i,t}}{g_{i,t,\theta_0}} - \frac{1-y_{i,t}}{1-g_{i,t,\theta_0}}\right)\left(\frac{y_{j,t}}{g_{j,t,\theta_0}} - \frac{1-y_{j,t}}{1-g_{j,t,\theta_0}}\right) \\ &= \sum_{i=1}^{N} \mathbb{E}\frac{\partial}{\partial\theta}g_{i,t,\theta_0}\frac{\partial}{\partial\theta^\intercal}g_{i,t,\theta_0}\left(\frac{1}{g_{i,t,\theta_0}} + \frac{1}{1-g_{i,t,\theta_0}}\right) = H_0,\end{aligned}$$

where we used $y_{i,t}^2 = y_{i,t}$, $(1-y_{i,t})^2 = 1-y_{i,t}$, $y_{i,t}(1-y_{i,t}) \equiv 0$, and the fact that, conditional on $\mathcal{I}_{t-1}$, $y_{i,t}$ and $y_{j,t}$ are independent. Thus, the score satisfies CLT

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

Similarly, for $\theta = \theta_0$,

$$\mathbb{E}H_t(\theta_0) = -\sum_{i=1}^{N} \mathbb{E}\frac{\partial}{\partial\theta}g_{i,t,\theta_0}\frac{\partial}{\partial\theta^\intercal}g_{i,t,\theta_0}\left(\frac{1}{g_{i,t,\theta_0}} + \frac{1}{1-g_{i,t,\theta_0}}\right) = -H_0.$$

Thus, applying Hayashi (2000, Theorem 7.8),

$$\sqrt{T}(\hat{\theta}^{MLE} - \theta_0) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, H_0^{-1}H_0H_0^{-1}) = \mathcal{N}(0, H_0^{-1}). \quad \square$$

Appendix D. Data

As of December 27, 2025, the S&P100 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.

Thirteen of the above companies are not available for the entire period under consideration, 01.01.2005–01.01.2024. The tickers of those companies are: ABBV, AVGO, GM, MA, META, NOW, PM, PLTR, PYPL, TMUS, TSLA, UBER, and V.

References

Aknouche, A. and Francq, 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.

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*.

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

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, 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