Econometric Theory, **0**, 2025, 1–44.
doi:10.1017/S0266466625100182

SWITCHING REGIME INTEGER AUTOREGRESSIONS

LEOPOLDO CATANIA
Aarhus University and CoRE

EDUARDO ROSSI
University of Pavia

PAOLO SANTUCCI DE MAGISTRIS
Luiss University

Time series of counts often display complex dynamic and distributional characteristics. For this reason, we develop a flexible framework combining the integer-valued autoregressive (INAR) model with a latent Markov structure, leading to the hidden Markov model-INAR (HMM-INAR). First, we illustrate conditions for the existence of an ergodic and stationary solution and derive closed-form expressions for the autocorrelation function and its components. Second, we show consistency and asymptotic normality of the conditional maximum likelihood estimator. Third, we derive an efficient expectation–maximization algorithm with steps available in closed form which allows for fast computation of the estimator. Fourth, we provide an empirical illustration and estimate the HMM-INAR on the number of trades of the Standard & Poor's Depositary Receipts S&P 500 Exchange-Traded Fund Trust. The combination of the latent HMM structure with a simple INAR(1) formulation not only provides better fit compared to alternative specifications for count data, but it also preserves the economic interpretation of the results.

1. INTRODUCTION

Traditionally, most time-series models have been developed for continuous data. However, many of the recorded time series in finance, economics, climatology, and biology are counts. Examples are the number of people infected by rare

We are grateful to the Editor, Prof. Yixiao Sun, and two anonymous reviewers for their extremely useful remarks that greatly improved the quality of the current version of the manuscript. We are also grateful to Kim Christensen for providing us with the high-frequency data on the number of trades and trading volume used in the empirical analysis. We also thank the participants at the HFFE workshop in Pisa (September 2018), the HAX conference in Xiamen (October 2018), the ICEEE conference in Lecce (January 2019), the NBER-NSF Time series conference in Hong Kong (August 2019), the Econometrics in the Arena workshop in Verona (September 2019), the IWEEE Time Series conference in Venezia (January 2020), the SNDE Symposium in Padua (March 2024), the SofiE conference in Rio de Janeiro (June 2024), and the workshop in honor of Carlo Giannini in Pavia (September 2024) for useful comments and suggestions. L.C. and P.S.d.M. acknowledge the research support of the Project 2 grants of the Danish Council for Independent Research (IRFD), Social Sciences, number 8019-00015A and 3099-00089B, and of the Aarhus Universitets Forskningsfond (AUFF). P.S.d.M and E.R. acknowledge the research support of the European Union's - Next Generation EU program through the Italian PRIN 2022 - M4C2, Investment 1.1 - "Monitoring Risks in Financial Markets" – Codice Cineca: 2022NEL482 - CUP: I53D23003410008. Address correspondence to Leopoldo Catania, Department of Economics and Business Economics and CoRE, Aarhus University, Aarhus, Denmark, e-mail: leopoldo.catania@econ.au.dk.

diseases, the number of transactions recorded per unit interval of a given stock on the financial markets, records of weapon offenses for different cities in the same district/area, and the annual number of hurricanes in a U.S. state. The availability of count data challenges the adequacy of the standard specifications, such as the well-known ARMA class. In particular, the need to provide coherent forecasts of integer-valued variables argues against approaches based on continuous transformations of the original data, and instead motivates the use of statistical models designed for discrete random variables (see Jung and Tremayne, 2006b). As a result, recent years have seen the emergence of various linear and nonlinear models and methods for the statistical analysis of integer-valued time series, providing alternatives to the ARMA specification. For an overview, see the latest surveys in Fokianos (2012), Davis et al. (2016, 2021), and the discussion in Aknouche and Franq (2021). Overall, the statistical analysis of discrete-valued stochastic processes in $\mathbb{N}$ (or $\mathbb{Z}$) poses substantial difficulties from a methodological viewpoint, greatly complicating the underlying theory and model interpretation.

We contribute to this strand of literature by providing a new statistical framework for the analysis of time series of counts that are typically characterized by high over-dispersion, often associated with the switch from a low-counts regime to a high-counts regime. In this article, we develop a flexible univariate model specification that belongs to the class of integer-valued autoregressive (INAR) models, originally proposed by Al-Osh and Alzaid (1987) and McKenzie (1985), which can be viewed as an alternative to the Poisson autoregression specifications outlined in Rydberg and Shephard (2000) and Fokianos, Rahbek, and Tjøstheim (2009) and generalized more recently in Roy and Karmakar (2021) and Armillotta, Luati, and Lupparelli (2022). Empirical applications of INAR models span several fields (see, among others, Thyregod et al., 1999; Rudholm, 2001; Gouriéroux and Jasiak, 2004; Pavlopoulos and Karlis, 2008). The main feature of the INAR is the use of the binomial thinning operator of Steutel and Van Harn (1979) to model the autocorrelation of the observations (see Weiß, 2008, 2015). Unfortunately, the autocorrelation function of the standard INAR(1) model mimics that of an AR(1) process, making it overly restrictive for practical applications. Furthermore, the generalization of the INAR(1) model to the p-th order autoregression, i.e., INAR(p), or to ARFIMA-type dynamics (e.g., the INARFIMA), comes at the cost of losing exact distributional properties, thus resorting to complicated inference techniques (see Brännäs and Hellström, 2001 and Quoreshi, 2014). For this reason, our modeling framework builds upon the simple INAR(1) setup, which we combine with a latent Markov structure that allows for a very flexible characterization of the dynamics of the series at hand and its complex distributional features. Further extensions aimed at improving flexibility include the random-coefficient INAR(1) of Zheng, Basawa, and Datta (2007), the random-environment INAR(1) of Nastić, Laketa, and Ristić (2016), and the Markov-switching INAR models of Alerini, Olteanu, and Ridgway (2017) and Lu and Wang (2022).

Our model belongs to the class of hidden Markov models (HMMs) (see Vermunt, Langeheine, and Böckenholt, 1999; Bartolucci and Farcomeni, 2009; Bartolucci, Farcomeni, and Pennoni, 2012; Zucchini, MacDonald, and Langrock

2017), which we refer to as HMM-INAR. The hidden Markov structure is made up of two independent Markov chains, one on the thinning operator and one on the innovations. Conditional on the latent hidden Markov processes, the design of the model is analogous to that of a simple INAR(1) Poisson process. We study the conditions for a unique stationary solution and compute the analytical expressions for the first two moments and the autocovariance function. Relying on the results of Douc, Moulines, and Rydén (2004), this allows us to derive consistency and asymptotic normality of the conditional maximum likelihood estimator (MLE). Estimation is carried out by resorting to an efficient expectation–maximization (EM) algorithm, whose steps are available in closed form, thereby enabling fast computation of the MLE. We also examine the finite-sample properties of the MLE through Monte Carlo simulations. In addition, with the hidden Markov representation of the HMM-INAR, we derive the predictive, filtered, and smoothed distributions of the latent variables as well as the joint predictive distribution of the variable at hand. We provide an empirical illustration, estimating the HMM-INAR using high-frequency data on the number of trades in the SPDR S&P 500 Exchange-Traded Fund (SPY). The time series exhibits complex dynamics and extreme distributional characteristics, making it a suitable framework for evaluating the adaptability of the HMM-INAR to these features.

The article is organized as follows. Section 2 presents the structure of the model. Section 3 sets out the assumptions and discusses the existence of an ergodic and stationary solution as well as the derivation of the first two moments. In Section 4, we establish the consistency and asymptotic normality of the conditional MLE. Section 5.1 presents an EM algorithm for computing the conditional MLE. Section 6 provides the empirical illustration. Finally, Section 7 concludes the article. The Appendix reports the proofs, and a supplementary material contains additional results on model selection and on the empirical application.[1]

2. THE HMM-INAR

To fix the notation, we first introduce the baseline INAR(1) specification (see Al-Osh and Alzaid, 1987). Let $Y_t \in \mathbb{N}$, $t \in \mathbb{Z}$, be a nonnegative integer-valued random variable following an INAR(1) process as

$$Y_t = A_t + \eta_t, \tag{1}$$

where η_t is the innovation term distributed as a Poisson random variable with intensity λ and the term $A_t = \alpha \circ Y_{t-1}$ is defined in terms of the binomial thinning operator, such as

[1] All results presented in this article are fully reproducible using the code available at: https://github.com/leopoldocatania/hmminar/.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

$$\alpha \circ Y_{t-1} = \sum_{n=1}^{Y_{t-1}} X_{t,n}, \tag{2}$$

where $X_{t,n}$ are *i.i.d.* Bernoulli random variables with success rate $\alpha \in [0,1)$. In other words, the observable process $\{Y_t\}$ in (1) is generated by two latent random components: A_t, representing the survivors from time $t-1$, and the innovation term η_t, representing the new arrivals at time t.

If the goal is to model persistent time series of counts characterized by extreme over-dispersion, the simple dynamic structure of the INAR(1) model is not sufficiently flexible. In theory, the INAR(1) specification could be extended by including $p > 1$ lags of A_t, i.e., the INAR(p) (see Alzaid and Al-Osh, 1990; Du and Li, 1991). Unfortunately, establishing theoretical properties for the general INAR(p) model is challenging, as discussed by Brännäs and Hellström (2001). Moreover, maximum likelihood estimation is typically complex and computationally demanding, as shown by Bu, McCabe, and Hadri (2008) and Pedeli, Davison, and Fokianos (2015).[2] This greatly limits the range of applicability of the INAR(p) model and makes it not particularly suitable to be combined with an HMM structure, although Bu and McCabe (2008) show that an INAR(p) model can be treated as a Markov chain with benefits for the computation of forecasts and their confidence intervals. Furthermore, to account for over-dispersion or under-dispersion of the data at hand, one may adopt a flexible distribution for the innovations, η_t. For instance, several alternatives to the Poisson distribution have been proposed; see the Negative Binomial of Al-Osh and Aly (1992) and Gouriéroux and Lu (2019), the compound Poisson of Schweer and Weiß (2014), the Geometric-INAR of Bourguignon and Weiß (2017), the mixture INAR (Mix-INAR) of Pavlopoulos and Karlis (2008) and Roick, Karlis, and McNicholas (2021), the zero-modified geometric INAR(1) of Kang et al. (2024), as well as the flexible specifications of Qian and Zhu (2025), Aknouche and Scotto (2024), and Weiß and Zhu (2024) within the class of INGARCH models.

We build the HMM-INAR upon the considerations outlined above. In particular, we develop a flexible yet interpretable INAR specification by employing a mixture of Poisson distributions with dynamic mixture probabilities for the innovations, η_t, and pairing it with a time-varying probability of survivorship, α. Specifically, we extend the baseline INAR(1) by allowing both α and η_t to depend upon two latent Markov chains, ($\{S_t^\alpha\}$ and $\{S_t^\eta\}$), and an additional unobserved process, ($\{Z_t\}$). These unobserved processes significantly contribute to increasing the flexibility of the model, while still maintaining the simple (and interpretable) INAR(1) structure outlined in (1). We let $Y_t \in \mathbb{N}$ be a nonnegative integer-valued random variable

[2] Furthermore, the interpretation of the features of the INAR(p) model is not as straightforward as for the INAR(1). The INAR(p) allows for multiple parametrizations when $p > 1$ (see the discussion in Jung and Tremayne (2006b) and the review in Jung and Tremayne (2006a)). In contrast, the INAR(1) model characterizes the random variable of interest as the sum of survivors and new arrivals.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

following the HMM-INAR process:

$$Y_t = A_{t;S_t^\alpha} + \eta_{t;Z_t(S_t^\eta)}, \tag{3}$$

where $\{S_t^\alpha\}$ is an ergodic first-order Markov chain with state space $[1,\ldots,J]$ and transition probability matrix $\boldsymbol{\Gamma}^\alpha = [\gamma_{i,j}^\alpha]_{i,j=1}^J$, such that $P(S_t^\alpha = i|S_{t-1}^\alpha = j, S_{t-s}^\alpha, s > 1) = P(S_t^\alpha = i|S_{t-1}^\alpha = j) = \gamma_{i,j}^\alpha$. The random variable $A_{t;S_t^\alpha}$ is also integer-valued and defined as

$$A_{t;S_t^\alpha} = \alpha_{S_t^\alpha} \circ Y_{t-1}.$$

In this case, the binomial thinning operator is such that $\alpha_{S_t^\alpha} \circ Y_{t-1} = \sum_{n=1}^{Y_{t-1}} X_{t,n}$, with $X_{t,n} = \sum_{j=1}^J \mathbb{1}(S_t^\alpha = j)X_{t,n,j}$, where $\mathbb{1}(\cdot)$ denotes the indicator function, and $X_{t,n,j}$ are Bernoulli random variables, independent over n, t, and j, with success rate $\alpha_j \in [0,1]$, i.e., $P(X_{t,n,j} = 1) = \alpha_j$, for $j = 1,\ldots,J$, with $\min_j \alpha_j < 1$. In other words, the realization of S_t^α determines the success probability of $X_{t,n}$. Hence, conditional on Y_{t-1} and S_t^α, the random variable $\alpha_{S_t^\alpha} \circ Y_{t-1}$ is Binomial with size Y_{t-1} and success rate $\alpha_{S_t^\alpha}$.

The innovation term, $\eta_{t;Z_t(S_t^\eta)}$, depends upon the realization of the unobserved variable Z_t, which in turn is affected by an additional ergodic unobserved first-order Markov chain, $\{S_t^\eta\}$, according to the hierarchical structure $Z_t = \sum_{l=1}{}^L I(S_t^\eta = l)Z_{t,l}$. The random variables $Z_{t,l}$, $l = 1,\ldots,L$, are assumed to be independent of S_t^α, S_t^η, and among themselves. They follow a categorical distribution on $[1,\ldots,K]$, such that $P(Z_{t,l} = k) = \omega_{l,k}$, for $l = 1,\ldots,L$. To simplify the exposition, the notation η_t is used in place of $\eta_{t;Z_t(S_t^\eta)}$, and the stochastic representation $\eta_t = \sum_{k=1}^K \mathbb{1}(Z_t = k)\eta_{t,k}$ is exploited. Here, the variables $\eta_{t,k}$ for $k = 1,\ldots,K$ are assumed to be *i.i.d.* Poisson distributed with intensity $0 < \lambda_k < \infty$, meaning $P(\eta_{t,k} = u) = \frac{\lambda_k^u e^{-\lambda_k}}{u!}$. All $\eta_{t,k}$ are assumed to be independent of S_t^η, S_t^α, and Z_t. The chain $\{S_t^\eta\}$ serves the purpose of representing the innovation term η_t as a mixture of K Poisson variables, where the composition evolves over time among L configurations, based on the realizations of $\{S_t^\eta\}$.

In the next section, we study the properties of the model and derive closed-form expressions of the mean, variance, and autocovariance function of $\{Y_t\}$ based on the HMM-INAR specification.

3. PROPERTIES OF THE HMM-INAR

Before presenting the probabilistic properties of the HMM-INAR, we summarize all the assumptions made so far in Section 2.

Assumption 1.

a) $\{S_t^\alpha\}$ is an ergodic unobserved first-order Markov chain with state space $[1,\ldots,J]$, with J fixed and finite, and transition probability matrix $\boldsymbol{\Gamma}^\alpha = [\gamma_{i,j}^\alpha]_{i,j=1}^J$, with $\gamma_{i,j}^\alpha \in (0,1)$ for all i,j, $\sum_{m=1}^J \gamma_{i,m}^\alpha = 1$ for all i, and stationary distribution $\boldsymbol{\pi}^\alpha$, where $\boldsymbol{\pi}^\alpha = (\pi_1^\alpha,\ldots,\pi_J^\alpha)'$ and $\boldsymbol{\pi}^\alpha = \boldsymbol{\Gamma}^{\alpha\prime}\boldsymbol{\pi}^\alpha$.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

b) $\{S_t^\eta\}$ is an ergodic unobserved first-order Markov chain, independent from $\{S_t^\alpha\}$, with state space $[1,\ldots,L]$ with L fixed and finite, and transition probability matrix $\mathbf{\Gamma}^\eta = [\gamma_{i,j}^\eta]_{i,j=1}^L$, with $\gamma_{i,j}^\eta \in (0,1)$ for all i,j, $\sum_{m=1}^L \gamma_{i,m}^\eta = 1$ for all i, and stationary distribution $\boldsymbol{\pi}^\eta$, where $\boldsymbol{\pi}^\eta = (\pi_1^\eta,\ldots,\pi_L^\eta)'$ and $\boldsymbol{\pi}^\eta = \mathbf{\Gamma}^{\eta\prime}\boldsymbol{\pi}^\eta$.
c) $\{Z_{t,l}\}$, for all $l = 1,\ldots,L$, are *i.i.d.* categorical random variables, independent among themselves, and independent from $\{S_t^\alpha\}$ and $\{S_t^\eta\}$, with $K \times 1$ vector of probabilities $\boldsymbol{\omega}_l = \left(\omega_{l,1},\ldots,\omega_{l,K}\right)'$ with K fixed and finite, $\omega_{l,k} \in (0,1)$, $\sum_{k=1}^K \omega_{l,k} = 1$ for $l = 1,\ldots,L$.
d) $\{X_{t,n,j}\}$, for all $n = 1,2,\ldots$, are *i.i.d.* Bernoulli random variables with success rate $\alpha_j \in [0,1]$, for all $j = 1,\ldots,J$, with $\min_{j=1,\ldots,J} \alpha_j < 1$, independent among themselves, and independent from $\{S_t^\alpha, S_t^\eta, Z_{t,l}, l = 1,\ldots,L\}$.
e) $\{\eta_{t,k}\}$, for $k = 1,\ldots,K$, are *i.i.d.* Poisson random variables with intensities $\lambda_k \in (0,\infty)$, independent among themselves and from $\{S_t^\alpha, S_t^\eta, X_{t,n,j}, Z_{t,l}, l = 1,\ldots,L, j = 1,\ldots,J, n = 1,2,\ldots\}$.

Assumption 1 specifies the stochastic characteristics of the components in the HMM-INAR model. In particular, Assumption 1(d) permits some, though not all, α_j to be equal to one. The presence of such "unit-root regimes" is a typical feature in regime-switching and mixture autoregressive models (see Hamilton, 1989; Wong and Li, 2000).

To study the properties of the HMM-INAR model, it is convenient to resort to an alternative (yet equivalent) stochastic representation of the system depicted in Figure 1. Specifically, we define a process $\{\mathcal{S}_t\}$ constructed by combining $\{S_t^\alpha\}$, $\{S_t^\eta\}$, and $\{Z_t\}$. $\{\mathcal{S}_t\}$ is still a first-order ergodic Markov chain with state space $\mathbb{H} = [1,\ldots,H]$, where $H = JKL$, with $H \times H$ transition probability matrix $\mathbf{\Gamma}^{\mathcal{S}} = [\gamma_{i,j}^{\mathcal{S}}]_{i,j=1}^H$, and with stationary distribution denoted by the $(H \times 1)$ vector of probabilities $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_H)'$ (see Lemma 1 below). The probabilities $\gamma_{i,j}$ and $\boldsymbol{\pi}_h$ are recovered from the definition of $\{\mathcal{S}_t\}$. Specifically, let $(h_1,h_2,h_3) \to h$ be the unique mapping between the indexes of the triplet $\{(S_t^\alpha, Z_t, S_t^\eta)\}$ and those of $\{\mathcal{S}_t\}$, i.e., $(S_t^\alpha = h_1, Z_t = h_2, S_t^\eta = h_3)$ for some $h_1 = 1,\ldots,J$, $h_2 = 1,\ldots,K$, $h_3 = 1,\ldots,L$, denotes $\mathcal{S}_t = h$, where $h \in \mathbb{H}$. Then, $\gamma_{m,h}^{\mathcal{S}} := P(\mathcal{S}_t = h|\mathcal{S}_{t-1} = m) = P(S_t^\alpha = h_1, Z_t = h_2, S_t^\eta = h_3|S_{t-1}^\alpha = m_1, Z_{t-1} = m_2, S_{t-1}^\eta = m_3)$ and $\pi_h^{\mathcal{S}} := P(\mathcal{S}_t = h) = P(S_t^\alpha = h_1, Z_t = h_2, S_t^\eta = h_3)$ with

$$
\begin{aligned}
\gamma_{m,h}^{\mathcal{S}} &= P(Z_t = h_2|S_t^\eta = h_3)P(S_t^\eta = h_3|S_{t-1}^\eta = m_3)P(S_t^\alpha = h_1|S_{t-1}^\alpha = m_1) \\
&= \omega_{h_3,h_2}\gamma_{m_3,h_3}^\eta \gamma_{m_1,h_1}^\alpha, \\
\pi_h^{\mathcal{S}} &= P(Z_t = h_2|S_t^\eta = h_3)P(S_t^\eta = h_3)P(S_t^\alpha = h_1) = \omega_{h_3,h_2}\pi_{h_3}^\eta \pi_{h_1}^\alpha,
\end{aligned}
$$

for all $h,m = 1,\ldots H$. It should be noted that the mapping $(S_t^\alpha, Z_t, S_t^\eta) \to \mathcal{S}_t$ is injective, allowing us to express the HMM-INAR model in (3) as

$$
Y_t = A_{t,\mathcal{S}_t} + \eta_{t,\mathcal{S}_t}, \tag{4}
$$

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

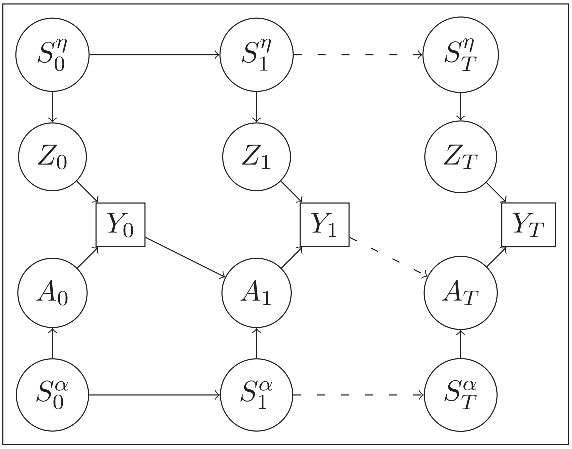


Figure 1. The HMM-INAR path diagram.

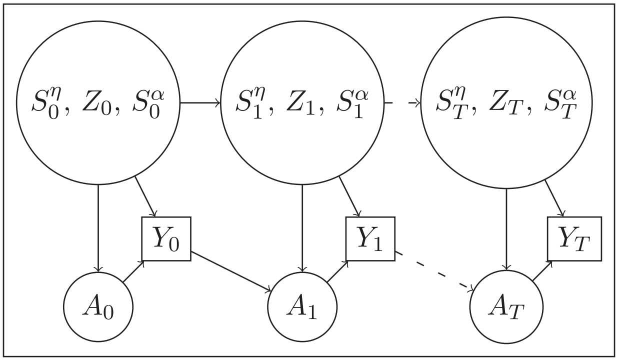


Figure 2. The HMM-INAR path diagram under equivalent parametrization.

where the path diagram of this representation is reported in Figure 2. The conditional probability mass function $P(Y_t = y_t | Y_{t-1} = y_{t-1}, \mathcal{S}_t = h)$ is

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}, \mathcal{S}_t = h) = \sum_{q=0}^{y_t \wedge y_{t-1}} e^{-\lambda_{h_2}} \frac{\lambda_{h_2}^q}{q!} \binom{y_{t-1}}{y_t - q} \alpha_{h_1}^{y_t - q} (1 - \alpha_{h_1})^{y_{t-1} - y_t + q}, \tag{5}$$

where $y_t \wedge y_{t-1} = \min(y_t, y_{t-1})$, and it does not depend on $h_3 = 1, \ldots, L$.

Examining the conditional distribution reveals how the components of the Markov chains contribute to its structure. Specifically, when the chain S_t^η is turned off (i.e., $L = 1$), the conditional distribution of η_t simplifies to a static Poisson mixture with K terms. Further setting $K = 1$ reduces the distribution of η_t to a

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

standard Poisson random variable. A similar interpretation applies to the chain associated with the Binomial thinning operator. Disabling the chain S_t^α (i.e., setting $J=1$) results in the Binomial thinning operator producing a binomial distribution with size Y_{t-1}, while it is a mixture of binomial distributions with dynamic weights when $J>1$. In the next section, we establish the ergodicity and stationarity of $\{(Y_t,\mathcal{S}_t)\}$, and in Theorem 2, we derive the moments of Y_t, showing that although Y_t is conditionally independent of S_t^η given (S_t^α, Z_t), the inclusion of the S_t^η chain, with $L>1$ states, provides additional flexibility in the autocorrelation structure of Y_t, through the serial dependence of η_t.

3.1. Ergodicity and Stationarity

To show the existence and uniqueness of a strictly stationary ergodic solution of HMM-INAR, we employ the $\mathcal{S}_t$-representation of the model in (4). We indicate with φ a measure on $(\mathbb{H},\mathcal{H})$, which satisfies $\varphi\{i\}>0$ for all $i\in\mathbb{H}$, and with μ a Lebesgue measure on $(\mathbb{N},\mathcal{B})$, such that $\mu(A)$ implies $(\mu\times\varphi)(A\times B)>0$ for $A\in\mathcal{B}$ and $B\in\mathcal{H}$. Let $(\Omega,\mathcal{F},\mathbb{P})$ denote a generic probability space, and let $\mathcal{H}$ and $\mathcal{B}$ be the σ-algebras generated by all subsets of $\mathbb{H}$ and $\mathbb{N}$, respectively. The following lemma establishes the ergodicity and stationarity of $\{\mathcal{S}_t\}$ as a consequence of ergodicity and stationarity of $\{S_t^\alpha\}$, $\{S_t^\eta\}$, and $\{Z_t\}$.

LEMMA 1. *Suppose Assumption 1.a)–c) holds. Then, $\{\mathcal{S}_t\}$ is a time-homogeneous first-order ergodic and stationary Markov chain.*

The proof is provided in Appendix A.1. According to Lemma 1, the chain $\{\mathcal{S}_t\}$ defined on $(\Omega,\mathcal{F},\mathbb{P})$ is stationary and ergodic, and it takes values in $\mathbb{H}$. Furthermore, $\{\mathcal{S}_t\}$ is φ-irreducible. In the following lemma, we establish two results on $\{(Y_t,\mathcal{S}_t)\}$ that are later used in Theorem 1 to derive the geometric ergodicity of $\{(Y_t,\mathcal{S}_t)\}$.

LEMMA 2. *Suppose Assumption 1 holds. Then*

a) *$\{(Y_t,\mathcal{S}_t)\}$ is a time-homogeneous Markov chain defined on $(\Omega,\mathcal{F},\mathbb{P})$ with state space $(\mathbb{N}\times\mathbb{H},\mathcal{B}\times\mathcal{H})$.*
b) *The Markov chain $\{(Y_t,\mathcal{S}_t)\}$ is $(\mu\times\varphi)$-irreducible and aperiodic.*

The proof is provided in Appendix A.2. In the following theorem, we establish a sufficient condition for $\{(Y_t,\mathcal{S}_t)\}$ to be geometrically ergodic. It is noteworthy that Theorem 1 also establishes that the geometric ergodicity of $\{(Y_t,\mathcal{S}_t)\}$ implies that the observable process $\{Y_t\}$ is geometrically ergodic.

THEOREM 1. *Suppose Assumption 1 holds. Then, the Markov chain $\{(Y_t,\mathcal{S}_t)\}$ is geometrically ergodic. Furthermore, there exist a unique probability distribution*

π^* and a positive number $\beta < 1$, such that, for any initial value $x \in \mathbb{N}$ with $Y_0 = x$,

$$\lim_{t\to\infty} \beta^{-t} \|P(Y_t \in \cdot | Y_0 = x) - \pi^*(\cdot)\|_\tau = 0, \tag{6}$$

where $\|\cdot\|_\tau$ is the total variation norm.

The proof is provided in Appendix A.3.

3.2. Moments

We now derive the first two moments of the HMM-INAR and its autocovariance function. For the derivation of the moments, we consider the single chain representation outlined in Figure 2. We denote by $\underline{\boldsymbol{\alpha}} = \boldsymbol{\iota}_{KL} \otimes \boldsymbol{\alpha}$ and $\underline{\boldsymbol{\lambda}} = \boldsymbol{\iota}_{JL} \otimes \boldsymbol{\lambda}$ the two vectors of length H containing the coefficients α_j and λ_k for all the states of $\{\mathcal{S}_t\}$, with $\boldsymbol{\iota}_N$ denoting the vector of ones of dimension N. Let also $\mathbf{A} = \mathrm{diag}(\underline{\boldsymbol{\alpha}})$, $\boldsymbol{\Lambda} = \mathrm{diag}(\underline{\boldsymbol{\lambda}})$, and $\boldsymbol{\Pi} = \mathrm{diag}(\boldsymbol{\pi})$ be three $H \times H$ diagonal matrices, and define $\mathbf{G} = \boldsymbol{\Pi}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Pi}$, and, for a square matrix $\mathbf{X}$, $\mathbf{X}^{(n)} = \prod_{i=1}^{n} \mathbf{X}$. In the following proposition, we derive the first and second moments of Y_t, $A_{t,\mathcal{S}_t}$, and $\eta_{t,\mathcal{S}_t}$, as well as their cross moments. For ease of notation, we write A_t and η_t in place of $A_{t,\mathcal{S}_t}$ and $\eta_{t,\mathcal{S}_t}$, respectively. In the following, we indicate with $\mathbf{I}_H$ the $H \times H$ identity matrix.

THEOREM 2. *Consider the HMM-INAR in (4), and let $\mathbf{Y}_{(1)}$ and $\mathbf{Y}_{(2)}$ denote $(\mathbb{E}[Y_t|\mathcal{S}_t = h], h = 1,\dots,H)'$ and $(\mathbb{E}[Y_t^2|\mathcal{S}_t = h], h = 1,\dots,H)'$, respectively. By Assumption 1, it follows that $\mathbf{Y}_{(1)} = (\mathbf{I}_H - \mathbf{A}\mathbf{G})^{-1}\underline{\boldsymbol{\lambda}}$, $\mathbb{E}[Y_t] = \boldsymbol{\pi}'\mathbf{Y}_{(1)}$, $\mathbb{E}[A_t] = \boldsymbol{\pi}'\mathbf{A}\mathbf{G}\mathbf{Y}_{(1)}$, and $\mathbb{E}[\eta_t] = \boldsymbol{\pi}'\underline{\boldsymbol{\lambda}}$. Furthermore,*

$$\mathbf{Y}_{(2)} = (\mathbf{I}_H - \mathbf{A}\mathbf{A}\mathbf{G})^{-1}\left\{(\mathbf{I}_H + \boldsymbol{\Lambda})\underline{\boldsymbol{\lambda}} + [\mathbf{A}(\mathbf{I}_H - \mathbf{A}) + 2\boldsymbol{\Lambda}\mathbf{A}]\mathbf{G}\mathbf{Y}_{(1)}\right\},$$

with $\mathbb{E}[Y_t^2] = \boldsymbol{\pi}'\mathbf{Y}_{(2)}$, $\mathbb{E}[A_t^2] = \boldsymbol{\pi}'\mathbf{A}\left[(\mathbf{I}_H - \mathbf{A})\mathbf{G}\mathbf{Y}_{(1)} + \mathbf{A}\mathbf{G}\mathbf{Y}_{(2)}\right]$, $\mathbb{E}[\eta_t^2] = \boldsymbol{\pi}'(\mathbf{I}_H + \boldsymbol{\Lambda})\underline{\boldsymbol{\lambda}}$, and $\mathbb{E}[A_t\eta_t] = \boldsymbol{\pi}'\boldsymbol{\Lambda}\mathbf{A}\mathbf{G}\mathbf{Y}_{(1)}$. Finally, for $k > 0$,

$$\mathbb{E}[A_tA_{t-k}] = \underline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}'\mathbf{M}_k\boldsymbol{\Pi}\mathbf{A}\mathbf{G}\mathbf{Y}_{(1)} + \underline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}'\left(\mathbf{A}\boldsymbol{\Gamma}'\right)^{(k-1)}\boldsymbol{\Pi}\left[\mathbf{Y}_{(2)} - \left(\boldsymbol{\Lambda}\mathbf{Y}_{(1)} + \underline{\boldsymbol{\lambda}}\right)\right]$$
$$\mathbb{E}[A_t\eta_{t-k}] = \underline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}'\mathbf{M}_k\boldsymbol{\Pi}\underline{\boldsymbol{\lambda}} + \underline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}'\left(\mathbf{A}\boldsymbol{\Gamma}'\right)^{(k-1)}\boldsymbol{\Pi}\left(\boldsymbol{\Lambda}\mathbf{Y}_{(1)} + \underline{\boldsymbol{\lambda}}\right)$$
$$\mathbb{E}[\eta_tA_{t-k}] = \underline{\boldsymbol{\lambda}}'\boldsymbol{\Gamma}^{(k)'}\boldsymbol{\Pi}\mathbf{A}\mathbf{G}\mathbf{Y}_{(1)}, \qquad \mathbb{E}[\eta_t\eta_{t-k}] = \underline{\boldsymbol{\lambda}}'\boldsymbol{\Gamma}^{(k)'}\boldsymbol{\Pi}\underline{\boldsymbol{\lambda}},$$

and

$$\mathbb{E}[Y_tY_{t-k}] = \underline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}'\mathbf{M}_k\boldsymbol{\Pi}\mathbf{Y}_{(1)} + \underline{\boldsymbol{\lambda}}'\boldsymbol{\Gamma}^{(k)'}\boldsymbol{\Pi}\mathbf{Y}_{(1)} + \underline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}'\left(\mathbf{A}\boldsymbol{\Gamma}'\right)^{(k-1)}\boldsymbol{\Pi}\mathbf{Y}_{(2)},$$

where $\mathbf{M}_1 = \mathbf{0}$, and $\mathbf{M}_k = \sum_{i=0}^{k-2}\left(\mathbf{A}\boldsymbol{\Gamma}'\right)^{(i)}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{(k-i-1)'}$ for $k > 1$.

The proof is provided in Appendix A.4. Given the results in Theorem 3, we can analyze and compare the amount of over-dispersion and autocorrelation generated by various HMM-INAR models. To clarify the state-space dimensions of S_t^α, Z_t, and S_t^η for each HMM-INAR specification, we introduce the notation HMM(J,K,L)-INAR. The models considered here range from the simple INAR(1)

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

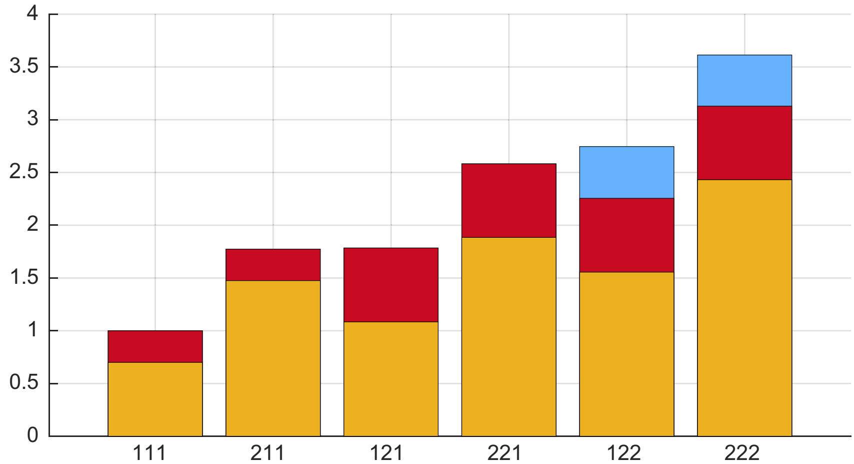


Figure 3. Over-dispersion decomposition. The HMM(1,1,1)-INAR is denoted as 111, the HMM(2,1,1)-INAR is denoted as 211, the HMM(1,2,1)-INAR is denoted as 121, the HMM(2,2,1)-INAR is denoted as 221, the HMM(1,2,2)-INAR is denoted as 122, and the HMM(2,2,2)-INAR is denoted as 222. The yellow area is the contribution of the variance of A_t to ID, the red area represents the contribution of the variance of (η_t), and the blue area represents the contribution of $2\mathbb{C}ov[A_t, \eta_t]$.

model (denoted as HMM(1,1,1)-INAR) to the HMM(2,2,2)-INAR model. Across all models, parameter configurations are chosen to ensure a fixed mean of $\mathbb{E}[Y_t] = 10$. For the INAR(1) model, the parameters are set to $\alpha = 0.7$ and $\lambda = 3$. In contrast, for the more sophisticated HMM(2,2,2)-INAR model, the parameters are specified as $\alpha_1 = 0.80$, $\alpha_2 = 0.565$, $\lambda_1 = 1$, $\lambda_2 = 5$, and

$$\mathbf{\Gamma}^{\alpha} = \begin{pmatrix} 0.85 & 0.15 \\ 0.15 & 0.85 \end{pmatrix}, \quad \boldsymbol{\omega}_1 = \begin{pmatrix} 0.2 \\ 0.8 \end{pmatrix}, \quad \boldsymbol{\omega}_2 = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}, \quad \mathbf{\Gamma}^{\eta} = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}.$$

Other specifications are derived by "switching off" certain components, restoring those of the baseline INAR(1). In Figure 3, the height of each bar represents the degree of over-dispersion generated by each model, as represented by the index of dispersion (ID), defined as $\text{ID} = \mathbb{V}ar[Y_t]/\mathbb{E}[Y_t]$. Furthermore, following Theorem 2, ID can be decomposed as

$$\text{ID} = \frac{\mathbb{V}ar[A_t] + \mathbb{V}ar[\eta_t] + 2\mathbb{C}ov(A_t, \eta_t)}{\mathbb{E}[Y_t]}.$$

In Figure 3, the yellow area represents the contribution of $\mathbb{V}ar[A_t]$ to the over-dispersion, the red area corresponds to the contribution of $\mathbb{V}ar[\eta_t]$, and the blue area reflects the contribution of $2\mathbb{C}ov[A_t, \eta_t]$. As expected, the baseline INAR(1) model does not produce over-dispersion (i.e., $\text{ID} = 1$), with most of its variability driven by $\mathbb{V}ar[A_t]$. All other HMM-INAR configurations generate over-dispersion,

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

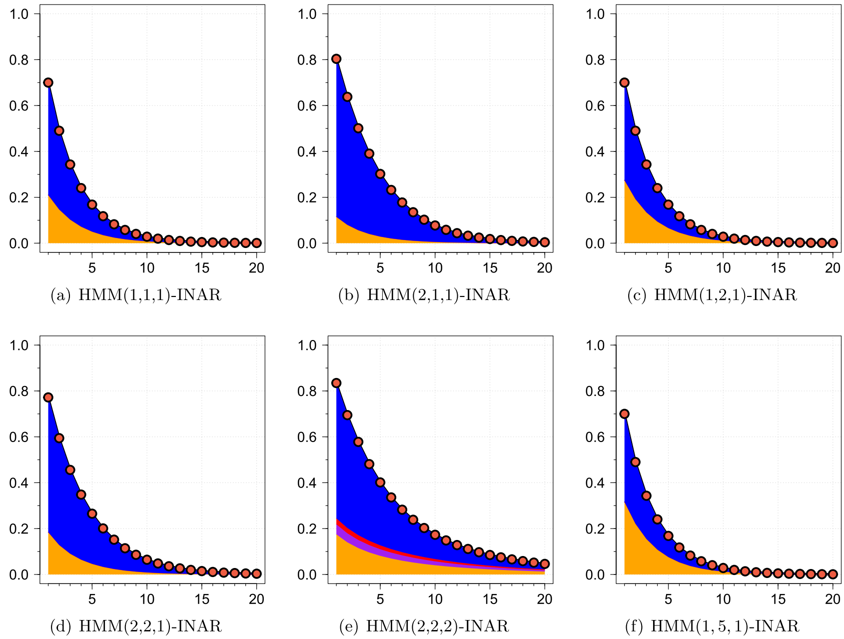


FIGURE 4. The ACF of the HMM-INAR model and its decomposition. The figure reports the ACF of the HMM-INAR (orange dots) and its decomposition into four components. The yellow area is $\mathbb{C}ov[A_t, \eta_{t-k}]/\mathbb{V}ar[Y_t]$, the purple area is $\mathbb{C}ov[\eta_t, A_{t-k}]/\mathbb{V}ar[Y_t]$, the red area is $\mathbb{C}ov[\eta_t, \eta_{t-k}]/\mathbb{V}ar[Y_t]$, and the blue area is $\mathbb{C}ov[A_t, A_{t-k}]/\mathbb{V}ar[Y_t]$. Panel a) reports the ACF of the baseline HMM(1,1,1)-INAR, Panel b) the HMM(2,1,1)-INAR, Panel c) the HMM(1,2,1)-INAR, Panel d) the HMM(2,2,1)-INAR, Panel e) the HMM(2,2,2)-INAR, and Panel f) the HMM(1, 5, 1)-INAR.

with the magnitude increasing as the number of states grows. Interestingly, models HMM(2,1,1)-INAR and HMM(1,2,1)-INAR exhibit similar levels of over-dispersion (approximately 1.8), but their sources differ. In the HMM(2,1,1)-INAR, the majority of variability (about 83%) is due to $\mathbb{V}ar[A_t]$, driven by the increased persistence of the α chain. In contrast, the HMM(1,2,1)-INAR model attributes a substantial portion of its over-dispersion (around 40%) to $\mathbb{V}ar[\eta_t]$, arising from the mixture of two Poisson random variables. When combining the α chain and the Poisson mixture, as in the HMM(2,2,1)-INAR model, over-dispersion increases further, with $\mathbb{V}ar[A_t]$ remaining the dominant contributor. In all these cases, $L = 1$, and $\mathbb{C}ov[A_t, \eta_t] = 0$. Finally, over-dispersion is amplified in models with $L > 1$, such as the HMM(1,2,2)-INAR model and the HMM(2,2,2)-INAR model. In these cases, a significant portion of over-dispersion is attributed to $\mathbb{C}ov[A_t, \eta_t]$, which arises from the time-varying composition of the Poisson mixture.

The persistence generated by the HMM-INAR process can be seen in Figure 4, which illustrates the autocorrelation function of Y_t for lags from 1 to 20 of alternative parameterizations, as decomposed into four components. Some insights

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

can be drawn. First, Panel (a) displays the ACF of the HMM(1,1,1)-INAR model, which is the plain vanilla INAR(1). Second, in Panel (b), the chain on α with $J = 2$ plays a dominant role in contributing to the autocorrelation, significantly more important than the one with $K = 2$ in Panel (c), which is an INAR(1) with a mixture of two Poisson random variables for η_t. Notably, in the model shown in Panel (d), which incorporates both a chain on α and a Poisson mixture for η_t, the contribution from the autocovariance of A_t (the blue area) is far more important than the cross-autocovariance between A_t and η_{t-k} (the yellow area). Third, the ACF of the model with $L > 1$, shown in Panel (e), exhibits a more complex and diverse structure compared to models with $L = 1$. This is attributed to the contributions from the autocovariance of η_t (red area) and the cross-autocovariance between η_t and A_{t-k} (purple area). Fourth, Panel (f) reports the ACF of a Mix-INAR model with five components ($J = 1$, $K = 5$, $L = 1$), calibrated to match the number of parameters ($M = 10$), mean, and over-dispersion of the HMM(2,2,2)-INAR model. The parameters of the HMM$(1,5,1)$-INAR model are set to $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$, $\lambda_4 = 15$, and $\lambda_5 = 26.5$, with $\boldsymbol{\omega}_1 = (0.18, 0.22, 0.57, 0.01, 0.02)'$. In this case, the overall autocorrelation of the series is lower than that of the HMM(2,2,2)-INAR model. In summary, by leveraging the double chain structure and the categorical variable, the HMM-INAR process can generate a rich variety of dynamic behaviors with relatively few parameters, avoiding the need for extreme parameter configurations.

4. MAXIMUM LIKELIHOOD ESTIMATION

The estimation of the HMM-INAR parameters can be carried out by conditional maximum likelihood (ML). In this section, we discuss the properties of the estimator, while in Section 5.1 below, we present an efficient EM algorithm with steps available in closed form that allow for fast computation of the MLE. For our asymptotic analysis, we rely on the results from Douc et al. (2004) for general Markov switching autoregressive models, exploiting the $\mathcal{S}_t$ representation of the HMM-INAR model.

The HMM-INAR parameters are $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)' \in [0,1]^J$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)' \in (0, \infty)^K$, $\boldsymbol{\omega} = (\boldsymbol{\omega}_1', \ldots, \boldsymbol{\omega}_L')' \in \mathbb{S}_K^L$ (where $\mathbb{S}_K$ is the standard K-th simplex), $\mathrm{vec}(\boldsymbol{\Gamma}^\alpha) \in \mathbb{M}_J$ and $\mathrm{vec}(\boldsymbol{\Gamma}^\eta) \in \mathbb{M}_L$, where $\mathbb{M}_A$ denotes the space of all vectorized $A \times A$ stochastic matrices with positive elements. Model parameters are collected in the $(M \times 1)$ vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\lambda}', \boldsymbol{\omega}', \mathrm{vec}(\boldsymbol{\Gamma}^\alpha)', \mathrm{vec}(\boldsymbol{\Gamma}^\eta)')'$, where $M = J + K + (K-1)L + J(J-1) + L(L-1)$. Furthermore, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta} = [0,1]^J \times (0, \infty)^K \times \mathbb{S}_K^L \times \mathbb{M}_J \times \mathbb{M}_L \subset \mathbb{R}^M$. Following Proposition 1 in Zucchini et al. (2017), the conditional log-likelihood function for a sample of length T, $\mathbf{y}_{1:T}$, denoted as $\ell_T(\boldsymbol{\theta}, s_0) = \log P(\mathbf{Y}_{1:T} = \mathbf{y}_{1:T} | \mathcal{S}_0 = s_0; \boldsymbol{\theta})$, takes the following form:

$$\ell_T(\boldsymbol{\theta}, s_0) = \log \left[\boldsymbol{e}_0' \mathbf{P}_1 \left(\prod_{t=2}^{T} \boldsymbol{\Gamma} \mathbf{P}_t \right) \boldsymbol{\iota}_H \right], \tag{7}$$

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

where $\boldsymbol{\iota}_H$ is a vector of ones of length H, $\boldsymbol{e}_0$ is the s_0-th column of the H-dimensional identity matrix, $\mathbf{P}_t$ is an $H \times H$ diagonal matrix with generic element $p_{t;h,h} = P(Y_t = y_t|\mathcal{S}_t = h, Y_{t-1} = y_{t-1})$ defined in Equation (5) for $h = 1, \dots, H$, and $\boldsymbol{\Gamma}^{\mathcal{S}}$ is the transition probability matrix of $\{\mathcal{S}_t\}$. The log-likelihood function (7) is defined conditionally on an arbitrary initial state $s_0 \in \mathbb{H}$ (which implies an initial state for S_t^{α}, S_t^{η}, and Z_t, through the map h) whose effect is proved to be asymptotically negligible by Douc et al. (2004), under our set of assumptions. The same applies to the initial Y_0, which is set to zero, i.e., $Y_0 = 0$. The MLE is

$$\widehat{\boldsymbol{\theta}}_{T,s_0} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell_T(\boldsymbol{\theta}, s_0), \tag{8}$$

for an arbitrary $s_0 \in \mathbb{H}$. The following assumption is required to establish the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_{T,s_0}$. Let $\boldsymbol{\Omega}$ be a $(K \times L)$ matrix with generic element $\omega_{k,l}$ for $k = 1, \dots, K$ and $l = 1, \dots, L$.

Assumption 2. The matrix $\boldsymbol{\Omega}$ has full rank with $K \geq L$. Also, if $K > 1$, $\lambda_i > \lambda_j$ for $i > j$, $i,j = 1, \dots, K$. If $K = 1$, then $\alpha_i > \alpha_j$ for $i > j$, $i,j = 1, \dots, J$.

Assumption 3. The parameter space $\boldsymbol{\Theta}$ is compact.

Assumption 2 is a sufficient condition for identification and implies that, once conditioning on S_t^{η} and S_t^{α}, the conditional probability mass functions are linearly independent. The assumption entails that the parameters λ and α can be different among the states, while imposing an ordering among them is not restrictive and it only prevents label swapping. Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ be the true vector of parameters. The following theorem establishes the consistency and the asymptotic normality of the conditional MLE in (8).

THEOREM 3. *Suppose Assumptions 1–3 hold. Then,* $\widehat{\boldsymbol{\theta}}_{T,s_0} \to \boldsymbol{\theta}_0$ *a.s. for any* $s_0 \in \mathbb{H}$. *If also* $\boldsymbol{\theta}_0 \in \dot{\boldsymbol{\Theta}}$, *where* $\dot{\boldsymbol{\Theta}} = int(\boldsymbol{\Theta})$, *then* $\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,s_0} - \boldsymbol{\theta}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1})$, *where* $\mathbf{I}(\boldsymbol{\theta}_0)$ *is the Fisher information matrix at* $\boldsymbol{\theta}_0$. *Furthermore,* $-\frac{1}{T}\nabla^2_{\boldsymbol{\theta}}\ell_T(\widehat{\boldsymbol{\theta}}_{T,s_0}, s_0) \to \mathbf{I}(\boldsymbol{\theta}_0)$ *a.s. for any* $s_0 \in \mathbb{H}$, *with* $\mathbf{I}(\boldsymbol{\theta}_0)$ *positive definite.*

The proof of Theorem 3, based on the findings of Douc et al. (2004), is presented in Appendix A.5. For consistency, it is sufficient that one α_j is less than 1, while asymptotic normality requires $\alpha_j \in (0,1)$ for all j. Additionally, the Hessian matrix, $-\frac{1}{T}\nabla^2_{\boldsymbol{\theta}}\ell_T(\widehat{\boldsymbol{\theta}}_{T,s_0}, s_0)$, almost surely converges to the Fisher information, irrespective of the initial choice of s_0.

The asymptotic distribution in Theorem 3 can be used to construct standard statistical tests, such as the asymptotic normal test based on a t-statistic, and the asymptotic chi-squared test based on the LR, LM, or Wald statistic, provided all coefficients are identifiable under the null hypothesis, and they are not on the boundary of the parameter space. In Section 5.2, we will show, through Monte Carlo simulations, the quality of the Gaussian approximation of the Z test in finite samples. From a computational perspective, the estimation of $\widehat{\boldsymbol{\theta}}_{T,s_0}$ presents practical difficulties, as it requires numerical constrained optimization

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

of a complex, highly nonlinear function with potentially numerous parameters. To overcome these challenges, Section 5.1 presents an EM algorithm featuring a closed-form M-step, ensuring convergence to $\widehat{\boldsymbol{\theta}}_{T,s_0}$.

5. EM ALGORITHM AND SIMULATION STUDY

5.1. EM Algorithm for the HMM-INAR Model

The EM algorithm presented below is an extension of the procedure proposed in Catania and Di Mari (2021) and Catania, Di Mari, and Santucci de Magistris (2022) for ML estimation of a hierarchical Markov-switching model for multivariate count data, and that of Pavlopoulos and Karlis (2008) for the Mix-INAR model. Differently from the previous section, we do not work with the initial condition s_0 but define generic initial distributions for $\{S_t^\alpha\}$ and $\{S_t^\eta\}$ at time $t=1$ by $\boldsymbol{\delta}^\alpha$ and $\boldsymbol{\delta}^\eta$, respectively, and include them in $\boldsymbol{\theta}$.[3] This choice is made out of convenience and has no asymptotic effect under Assumption 1.

We derive the joint distribution of a series of T observed variables $(\mathbf{Y}_{1:T})$ and unobserved variables $(\mathbf{S}_{1:T}^\eta, \mathbf{S}_{1:T}^\alpha, \mathbf{Z}_{1:T}, \boldsymbol{\eta}_{1:T})$, where $\boldsymbol{\eta}_{1:T} = (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_T')'$ with $\boldsymbol{\eta}_t' = (\eta_{t,1}, \ldots, \eta_{t,K})'$ and similarly for $\mathbf{Z}_{1:T}$. Conditional on Y_0, the joint distribution denoted by $\mathcal{P} := P(\mathbf{Y}_{1:T}, \mathbf{S}_{1:T}^\eta, \mathbf{S}_{1:T}^\alpha, \mathbf{Z}_{1:T}, \boldsymbol{\eta}_{1:T}|Y_0)$[4] is

$$
\begin{aligned}
\mathcal{P} &= P(S_1^\alpha)P(S_1^\eta)\prod_{t=2}^{T} P(S_t^\alpha|S_{t-1}^\alpha)\prod_{t=2}^{T} P(S_t^\eta|S_{t-1}^\eta)\prod_{t=1}^{T} P(Z_t|S_t^\eta) \\
&\quad \times \prod_{t=1}^{T} P(Y_t|Y_{t-1}, S_t^\alpha, Z_t, \eta_t)\prod_{t=1}^{T} P(\eta_t|Z_t) \\
&= P(S_1^\alpha)P(S_1^\eta)\prod_{t=2}^{T} \gamma^\alpha_{S_{t-1}^\alpha, S_t^\alpha}\prod_{t=2}^{T} \gamma^\eta_{S_{t-1}^\eta, S_t^\eta}\prod_{t=1}^{T} \omega_{S_t^\eta, Z_t} \\
&\quad \times \prod_{t=1}^{T} \alpha_{S_t^\alpha}^{Y_t-\eta_t}(1-\alpha_{S_t^\alpha})^{Y_{t-1}-Y_t+\eta_t}\prod_{t=1}^{T} \frac{e^{-\lambda_{Z_t}}\lambda_{Z_t}^{\eta_t}}{\eta_t!},
\end{aligned}
$$

where $P(S_1^\alpha)$ and $P(S_1^\eta)$ represent an element of the initial distribution of S_t^α and S_t^η, which are denoted as $\boldsymbol{\delta}^\alpha = (\delta_1^\alpha, \ldots, \delta_J^\alpha)'$ and $\boldsymbol{\delta}^\eta = (\delta_1^\eta, \ldots, \delta_L^\eta)'$, respectively. The complete data log-likelihood (CDLL) is obtained by introducing the following augmenting variables: $u_{1,l}^\eta = 1$ if $S_1^\eta = l$, $u_{t,j}^\alpha = 1$ if $S_t^\alpha = j$, $v_{t,i,l}^\eta = 1$ if $S_{t-1}^\eta = i$ and $S_t^\eta = l$, $v_{t,i,j}^\alpha = 1$ if $S_{t-1}^\alpha = i$ and $S_t^\alpha = j$, and $z_{t,l,k} = 1$ if $Z_t = k$ and $S_t^\eta = l$. The CDLL is

[3] The MLE of $\boldsymbol{\delta}^\alpha$ and $\boldsymbol{\delta}^\eta$ is a unit vector, and it is known to be inconsistent (see Levinson et al., 1983, p. 1055). This will have no asymptotic effect on the estimation of the other parameters, which are shown to be consistent and asymptotically normal for any choice of the initial distribution of the two Markov chains in Theorem 3.

[4] We adopt the notation $P(X)$ to indicate $P(X=x)$ for a realization x of the random variable X.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

$$\begin{aligned}\ell_{T,c}(\boldsymbol{\theta}) \propto &\sum_{j=1}^{J} u_{1,j}^{\alpha} \log(\delta_j^{\alpha}) + \sum_{l=1}^{L} u_{1,l}^{\eta} \log(\delta_l^{\eta}) + \sum_{t=2}^{T}\sum_{j=1}^{J}\sum_{i=1}^{J} v_{t,i,j}^{\alpha} \log(\gamma_{i,j}^{\alpha}) \\ &+ \sum_{t=2}^{T}\sum_{l=1}^{L}\sum_{i=1}^{L} v_{t,i,l}^{\eta} \log(\gamma_{i,l}^{\eta}) + \sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{k=1}^{K} z_{t,l,k} \log(\omega_{l,k}) \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K} z_{t,k}\left[-\lambda_k + \eta_{t,k}\log(\lambda_k)\right] + \\ &+ \sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K} u_{t,j}^{\alpha} z_{t,k}\left[(Y_t - \eta_{t,k})\log(\alpha_j) + (Y_{t-1} - Y_t + \eta_{t,k})\log(1-\alpha_j)\right],\end{aligned} \tag{9}$$

where $z_{t,k} = \sum_{l=1}^{L} z_{t,l,k}$.

Maximization of (9) is unfeasible due to the presence of latent quantities. The EM algorithm treats these unobserved terms as missing values and proceeds with the iterative maximization of the expected value of the CDLL. Specifically, let $\boldsymbol{\theta}^{(m)}$ be the value of the parameters at the m-th iteration. The EM maximizes $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) = \mathbb{E}^{\boldsymbol{\theta}^{(m)}}\left[\ell_{T,c}(\boldsymbol{\theta})\right]$ (M-step), where the expectation (E-step) is taken with respect to the joint distribution of the missing variables conditional on the observed variables, i.e., $\boldsymbol{\theta}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)})$. The E-step amounts to computing the expectation of the augmenting variables $u_{t,l}^{\eta}, u_{t,j}^{\alpha}, v_{t,j,l}^{\eta}, v_{t,j,l}^{\alpha}, z_{t,l,k}$, and $\eta_{t,k}$ conditionally on $\mathbf{Y}_{0:T}$. These are

$$\widehat{u}_{1,l}^{\eta} := \mathbb{E}[u_{1,l}^{\eta}|\mathbf{Y}_{0:T}] = \sum_{j=1}^{J}\sum_{k=1}^{K} P(S_1^{\eta} = l, S_1^{\alpha} = j, Z_1 = k|\mathbf{Y}_{0:T}),$$

$$\widehat{u}_{t,j}^{\alpha} := \mathbb{E}[u_{t,j}^{\alpha}|\mathbf{Y}_{0:T}] = \sum_{l=1}^{L}\sum_{k=1}^{K} P(S_t^{\eta} = l, S_t^{\alpha} = j, Z_t = k|\mathbf{Y}_{0:T}),$$

$$\widehat{z}_{t,l,k} := \mathbb{E}[z_{t,l,k}|\mathbf{Y}_{0:T}] = \sum_{j=1}^{J} P(S_t^{\alpha} = j, S_t^{\eta} = l, Z_t = k|\mathbf{Y}_{0:T}),$$

$$\widehat{\eta}_{t,k} := \mathbb{E}[\eta_{t,k}|\mathbf{Y}_{0:T}] = \sum_{j=1}^{J} P(S_t^{\alpha} = j|\mathbf{Y}_{0:T}) \frac{\lambda_k P(Y_t = y_t - 1|Y_{t-1}, Z_t = k, S_t^{\alpha} = j)}{P(Y_t = y_t|Y_{t-1}, Z_t = k, S_t^{\alpha} = j)},$$

where $P(Y_t = y|Y_{t-1}, Z_t = k, S_t^{\alpha} = j)$ is reported in (5), and

$$\begin{aligned}\widehat{v}_{t,m,l}^{\eta} &:= \mathbb{E}[v_{t,m,l}^{\eta}|\mathbf{Y}_{0:T}] \\ &= \sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{k=1}^{K}\sum_{g=1}^{K} P(S_t^{\eta} = l, S_{t-1}^{\eta} = m, S_t^{\alpha} = j, S_{t-1}^{\alpha} = i, Z_t = k, Z_{t-1} = g|\mathbf{Y}_{1:T}),\end{aligned}$$

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

$$\widehat{v}^{\alpha}_{t,i,j} := \mathbb{E}[v^{\alpha}_{t,i,j}|\mathbf{Y}_{0:T}]$$
$$= \sum_{l=1}^{L}\sum_{m=1}^{L}\sum_{k=1}^{K}\sum_{g=1}^{K} P(S^{\eta}_t = m, S^{\eta}_{t-1} = l, S^{\alpha}_t = j, S^{\alpha}_{t-1} = i, Z_t = k, Z_{t-1} = g|\mathbf{Y}_{1:T}).$$

The computation of the conditional probabilities required in the E-step of the algorithm is achieved via a run of the forward filtering backward smoothing algorithm, exploiting the single chain representation of the model discussed in Section 2. It follows that $\mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{(m)})$ is given by

$$\mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{(m)}) \propto \sum_{l=1}^{L}\widehat{u}^{\eta}_{1,l}\log(\delta^{\eta}_l) + \sum_{t=2}^{T}\sum_{j=1}^{J}\sum_{i=1}^{J}\widehat{v}^{\alpha}_{t,i,j}\log(\gamma^{\alpha}_{i,j}) + \sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{k=1}^{K}\widehat{z}_{t,l,k}\log(\omega_{l,k})$$
$$+\sum_{j=1}^{J}\widehat{u}^{\alpha}_{1,j}\log(\delta^{\alpha}_j) + \sum_{t=2}^{T}\sum_{l=1}^{L}\sum_{i=1}^{L}\widehat{v}^{\eta}_{t,i,l}\log(\gamma^{\eta}_{i,l})$$
$$+\sum_{t=1}^{T}\sum_{k=1}^{K}\widehat{z}_{t,k}\left[-\lambda_k + \widehat{\eta}_{t,k}\log(\lambda_k)\right]$$
$$+\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K}\widehat{u}^{\alpha}_{t,j}\widehat{z}_{t,k}\left[(Y_t - \widehat{\eta}_{t,k})\log(\alpha_j) + (Y_{t-1} - Y_t + \widehat{\eta}_{t,k})\log(1-\alpha_j)\right],$$

whose maximum is available in closed form with $\gamma^{\alpha\,(m+1)}_{i,j} = \frac{\sum_{t=2}^{T}\widehat{v}^{\alpha}_{t,i,j}}{\sum_{t=2}^{T}\sum_{h=1}^{J}\widehat{v}^{\alpha}_{t,h,j}}$, $\gamma^{\eta\,(m+1)}_{l,i} = \frac{\sum_{t=2}^{T}\widehat{v}^{\eta}_{t,l,i}}{\sum_{h=1}^{L}\sum_{t=2}^{T}\widehat{v}^{\eta}_{t,l,h}}$, $\omega^{(m+1)}_{l,k} = \frac{\sum_{t=1}^{T}\widehat{z}_{t,l,k}}{\sum_{t=1}^{T}\sum_{h=1}^{K}\widehat{z}_{t,l,h}}$, $\alpha^{(m+1)}_j = \frac{\sum_{t=1}^{T}\sum_{k=1}^{K}\widehat{u}^{\alpha}_{t,j}\widehat{z}_{t,k}(Y_t-\widehat{\eta}_{t,k})}{\sum_{t=1}^{T}\sum_{k=1}^{K}\widehat{u}^{\alpha}_{t,j}\widehat{z}_{t,k}Y_{t-1}}$, $\lambda^{(m+1)}_k = \frac{\sum_{t=1}^{T}\widehat{z}_{t,k}\widehat{\eta}_{t,k}}{\sum_{t=1}^{T}\widehat{z}_{t,k}}$, and $\delta^{\eta\,(m+1)}_l = \widehat{u}^{\eta}_{1,l}$, $\delta^{\alpha\,(m+1)}_j = \widehat{u}^{\alpha}_{1,j}$.

Given an initial guess $\boldsymbol{\theta}^{(0)}$, the algorithm iterates between the E-step and the M-step until convergence. Convergence to a local optimum is guaranteed since the M-step increases the likelihood value at each iteration. As for standard HMMs, the likelihood function can present multiple local optima, and there is no guarantee that convergence to the global optimum is achieved. To this end, running the algorithm several times with different starting values is standard practice to better explore the likelihood surface.

5.2. Monte Carlo Analysis

We now investigate the finite sample properties of the MLE computed with the EM algorithm. We consider an HMM(2,2,2)-INAR model with parameters $\lambda_1 = 1$, $\lambda_2 = 7$, $\gamma^{\alpha}_{1,1} = \gamma^{\alpha}_{2,2} = \gamma^{\eta}_{1,1} = \gamma^{\eta}_{2,2} = 0.9$, $\alpha_1 = 0.4$, $\alpha_2 = 0.9$, $\boldsymbol{\omega}_1 = (0.7, 0.3)'$, and $\boldsymbol{\omega}_2 = (0.3, 0.7)'$. The experiment proceeds as follows: first, we simulate a sequence of T observations from the HMM(2,2,2)-INAR model, and second, we estimate the model using the EM algorithm outlined above. We iterate this procedure

Table 1. This table reports the bias, the RMSE, and the frequency of the rejection of the null hypothesis ($\theta_i = 0$) of the Z-test statistic ($\hat{\theta}_i / S.E.(\hat{\theta}_i)$) at 5% significance level based on the ML estimates of an HMM-INAR model with $J = 2$, $K = 2$, and $L = 2$

	Bias				RMSE				Rejection frequency			
T	250	500	1,000	5,000	250	500	1,000	5,000	250	500	1,000	5,000
γ_{11}^{η}	0.020	0.017	0.013	0.002	0.114	0.102	0.081	0.031	0.170	0.181	0.148	0.069
γ_{22}^{η}	−0.017	−0.014	−0.009	−0.002	0.112	0.099	0.079	0.029	0.182	0.183	0.153	0.054
γ_{11}^{α}	0.008	0.004	0.002	0.000	0.045	0.030	0.020	0.009	0.062	0.057	0.052	0.047
γ_{22}^{α}	−0.005	−0.002	−0.001	0.000	0.036	0.024	0.017	0.007	0.060	0.059	0.054	0.052
ω_{11}	−0.049	−0.038	−0.025	−0.003	0.160	0.133	0.102	0.038	0.083	0.104	0.099	0.049
ω_{21}	0.048	0.036	0.022	0.004	0.172	0.143	0.109	0.042	0.096	0.109	0.100	0.038
α_1	0.003	0.001	0.001	0.000	0.038	0.026	0.018	0.008	0.061	0.063	0.059	0.059
α_2	0.001	0.000	0.000	0.000	0.016	0.010	0.007	0.003	0.065	0.056	0.055	0.051
λ_1	−0.013	−0.006	−0.001	0.001	0.330	0.217	0.149	0.063	0.087	0.072	0.064	0.050
λ_2	−0.029	−0.011	−0.007	0.000	0.465	0.320	0.224	0.097	0.054	0.053	0.053	0.044

Note: The results are based on 10,000 replications.

$B =$10,000 times. We consider four sample sizes: small ($T = 250$), medium-small ($T = 500$), medium-large ($T = 1,000$), and large ($T = 5,000$).

Table 1 reports the bias, root mean squared error (RMSE), and the rejection frequency of the Z-test at 5% significance level based on the ML estimates and their asymptotic distribution in Theorem 3. Note that we do not report the results for $\gamma_{1,2}^{\eta}$, $\gamma_{2,1}^{\eta}$, $\gamma_{1,2}^{\alpha}$, $\gamma_{2,1}^{\alpha}$, $\omega_{1,2}$, and $\omega_{2,2}$ because these are obtained as a deterministic transformation of the other parameters (for instance, $\gamma_{1,2}^{\eta} = 1 - \gamma_{1,1}^{\eta}$). Results indicate that the bias and RMSE of the MLE behave as expected. The bias is generally small even for $T = 500$, and the RMSE decreases as expected when the sample size increases. Empirical frequency of rejection of the null hypothesis against a two-sided alternative by the Z-test indicates that the Gaussian approximation is adequate even for moderate sample sizes. Finally, the adoption of the EM algorithm, which does not require numerical optimization, rules out numerical instabilities, thus contributing to the quality of the estimates.

5.3. Model Selection

The asymptotic results presented in Section 4 are valid for a fixed and known number of states (K, J, and L). However, it is well-known that the parameters of an HMM are not identifiable if the number of hidden states is over-specified (see, among others, Rydén, Teräsvirta, and Aasbrink, 1998). Therefore, selecting the correct number of hidden states, or *order selection*, is crucial for consistent

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

parameter inference in HMMs. We consider BIC for order selection, which is given by

$$\mathrm{BIC} = -2\ell_T(\hat{\boldsymbol{\theta}}) + M\log(T).$$

In particular, BIC is known to be strongly consistent in i.i.d. settings as well as in certain non-i.i.d. settings and in finite mixture models, which commonly use penalized likelihood approaches (see Claeskens and Hjort (2008) for a review). In the context of HMMs, Csiszár and Shields (2000) and Gassiat and Boucheron (2003) show strong consistency of BIC for observations that take a finite set of values, even without imposing an upper bound on the order of the HMM. Monte Carlo simulations support the reliability of BIC for the HMM-INAR model, with correct model selection achieved in over 99% of cases when $T = 1,000$, as reported in Table S3 in the Supplementary Material. Importantly, the percentage of correctly identified models increases with T, and the performance is notably strong even for smaller sample sizes ($T = 250$). In contrast, the Akaike Information Criterion (AIC) displays non-monotonic behavior as T increases, reflecting challenges with consistency.

6. EMPIRICAL ILLUSTRATION

Trading volume is a crucial financial metric that reflects various factors, such as the arrival of new information, trader disagreements, microstructural frictions, and liquidity issues, as surveyed by Karpoff (1987). Beyond systematic diurnal patterns, such as those linked to the opening and closing of the trading day, high-frequency trading volume trajectories likely indicate the presence of informed traders, resulting in distinct trading regimes. For instance, Barardehi, Bernhardt, and Davies (2019) build a refined Amihud (2002) measurement of illiquidity based on the theoretical framework developed in Easley and O'Hara (1987), which highlights how multiple regimes may correspond to periods of intense news arrival and a predominant presence of informed traders, while periods of low trading activity may reflect information staleness and prolonged inactivity. Furthermore, liquidity concerns associated with the price impact of large trades induce traders to dynamically split orders into a sequence of smaller trades, the so-called trade splitting effect, where large transactions are broken up into smaller ones executed sequentially. Therefore, the number of trades for a specific instrument over short intervals provides valuable insights into the trading activity and liquidity in financial markets, as discussed in Tauchen and Pitts (1983) and, more recently, in Ranaldo and Santucci de Magistris (2022), among others.

In this study, we analyze the time series of the number of trades sampled at one-minute intervals for the Standard & Poor's 500 ETF (SPY) over the period from 2 January 2001 to 31 August 2001, using data from the TAQ database. Each trading day comprises 390 observations, resulting in 146 days and a total of 59,940 data points. The time series displays both extreme distributional traits, shifting from low to high trading regimes, associated with *low counts* and *high counts*, including

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

Table 2. The table reports the mean, median, mode, max, min, standard deviations (SD), excess kurtosis (Ex. Kurt), skewness coefficient (Skew), and ID

Mean	Median	Mode	Max	Min	SD	Ex. Kurt	Skew	ID	Period
8.60	7.00	5.00	209.00	0.00	6.76	74.44	4.89	5.31	Full
14.75	12.00	11.00	209.00	0.00	12.32	55.58	5.65	10.29	Opening
5.81	5.00	4.00	96.00	0.00	4.14	24.57	2.63	2.95	Midday
11.56	11.00	8.00	64.00	0.00	6.02	4.12	1.33	3.14	Closing

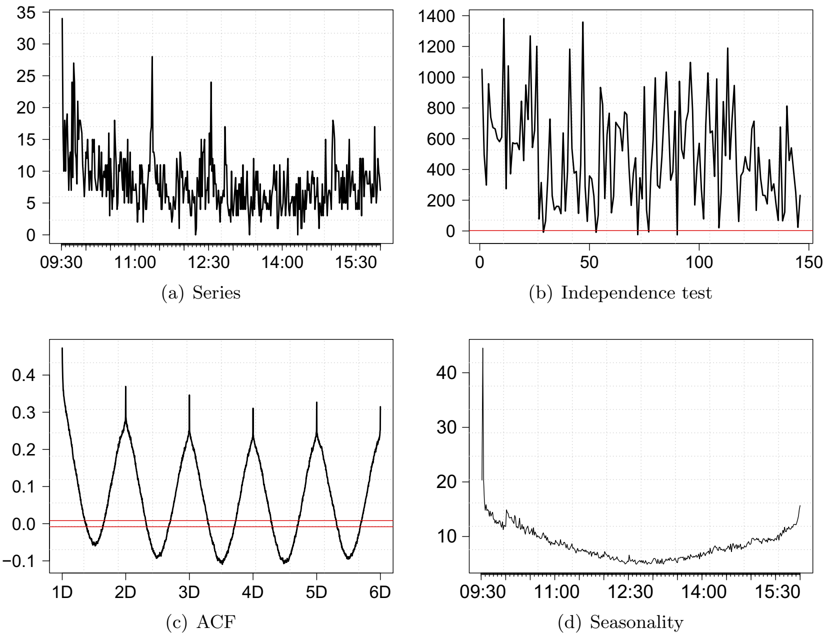
Note: Period indicates the intra-daily interval used to compute the statistics. Results are reported for the entire day (Day), the first 30 minutes of trading activity (Opening), the period between 12:00 and 14:00 (Midday), and the last 30 minutes of trading activity (Closing).

periods with no trades. This makes the sample an ideal series to assess how well Markov chains can adapt to these distinctive attributes. Furthermore, adopting the INAR specification for the number of trades allows for an intuitive interpretation of the components of the trading activity in financial markets. Indeed, the total number of transactions at time t can be disentangled into a component responsible for the arrival of new trades (η_{t,S_t}) and another (A_{t,S_t}) possibly associated with trade splitting.

Table 2 presents the summary statistics for the sample under analysis. Looking at the entire day, the number of trades exhibits substantial excess kurtosis and overdispersion, as indicated by the sample ID, which is 5.31. Additionally, we analyze the sample statistics across different times of the trading day. As anticipated, the average number of trades is higher at the beginning and end of the trading session. Notably, the standard deviation and kurtosis are also larger during these periods, indicating greater dispersion compared to Midday. The ID reaches its maximum at the opening, with a value of 10.29.

In Figure 5, Panel a) reports the number of trades of SPY at the one-minute frequency recorded on 9 May 2001. Panel b) reports the value of the test statistics of Harris and McCabe (2019) for the null hypothesis of independence computed over daily sub-samples (146 days with 390 observations per day). The red solid line denotes the critical value at the 1% significance level based on the asymptotic distribution. The test statistic rejects the null hypothesis in almost all cases (141 out of 146), thus supporting the evidence that the intra-daily number of trades is not an independent sequence.[5] The strong seasonality in the number of trades is evident when looking at the empirical autocorrelation function and the average number of trades computed over different minutes of the day, as reported in Panels c) and d) in Figure 5, respectively. Due to the presence of seasonality and overnight periods (when trading does not take place), the HMM-INAR model is therefore extended to account for the peculiar features of the data at hand. Notably, the EM

[5]When computing the test of Harris and McCabe (2019) on the whole sample, the test reports a value of 3,073,730, which is far above the 1% critical value.



(a) Series (b) Independence test (c) ACF (d) Seasonality

FIGURE 5. Panel a) reports the one-minute frequency number of SPY trades recorded on 9 May 2001. Panel b) reports the value of the test statistic of Harris and McCabe (2019) computed for all daily subsamples. The red solid line denotes the critical value at 1% significance level based on the asymptotic distribution. Panel c) reports the empirical ACF computed over the full sample. Panel d) presents the average number of trades per one-minute intra-daily period, computed over the 146-day sample.

algorithm illustrated in Section 5.1 for the estimation of the new model parameters requires a straightforward adaptation to the new specification, incurring only a minor additional computational cost.

6.1. The Seasonal HMM-INAR Model

Financial data sampled at high frequencies typically exhibit strong intra-daily seasonal patterns (see, among others, the recent contribution by Andersen, Thyrsgaard, and Todorov, 2019. The intra-daily periodicity reflects different trading patterns associated with the traders' working hours and/or market operating rules. These features cannot be fully accounted for by the baseline HMM-INAR model introduced in Section 2. For instance, at lunch, trading activity is low, while at the beginning and end of the trading day, activity is more pronounced since traders tend to balance their positions after (before) the market opens (closes). Furthermore, the institutional settings of the U.S. financial markets operating in the interval between 9 AM until 4 PM create a long interval, called the *overnight period*, between consecutive trading days. In terms of the INAR specification, the non-trading hours

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

during the overnight period entail that the term $A_{t;S_t^\alpha}$, which is a function of the number of trades executed in the previous interval, has to be low at the opening of the market hours.

The strong intra-daily periodicity calls for a proper treatment within the HMM-INAR model, which highlights the flexibility of this model specification to the inclusion of additional features. In particular, seasonality can be introduced by assuming that $\lambda_{t,k} = \lambda_k \sum_{p=1}^{P} f_t(p)\beta_p$, $k = 1,\ldots,K$, where P is the number of periods or seasons. To avoid an identifiability problem, we impose $\beta_1 = 1$. Incorporating intra-daily seasonal patterns using dummy variables is especially beneficial in the context of the HMM-INAR model, as it enables obtaining a closed-form M-step for the seasonal parameters $\beta_p, p = 1,\ldots,P$. The parameter λ_k is the baseline intensity for the k-th component of the mixture and β_p is the intra-daily period-specific multiplicative term, that is, $f_t(p) = 1$ if t is in season p, and zero otherwise. The presence of an overnight break in the transactions can also be accommodated by introducing a specific dummy at the opening of the trading day. Hence, the term $A_{S_t^\alpha}$ is modeled as

$$A_{t,S_t^\alpha} = \left[\alpha_{S_t^\alpha}(1-O_t) + \varpi O_t\right] \circ N_{t-1}, \tag{10}$$

where $O_t = 1$, if time t coincides with the opening and 0 otherwise, and ϖ determines the thinning operator at the opening of the trading day. The parameter ϖ is expected to be very low reflecting the reduction in persistence due to the overnight period.

The $\mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{(m)})$ function is easily extended to account for the intra-daily seasonality as

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{(m)}) &\propto \sum_{j=1}^{J} \widehat{u}_{1,j}^{\alpha} \log(\delta_j^\alpha) + \sum_{l=1}^{L} \widehat{u}_{1,l}^{\eta} \log(\delta_l^\eta) + \sum_{j=1}^{J}\sum_{i=1}^{J}\sum_{t=2}^{T} \widehat{v}_{t,j,i}^{\alpha} \log(\gamma_{j,i}^{\alpha}) \\ &+ \sum_{l=1}^{L}\sum_{i=1}^{L}\sum_{t=2}^{T} \widehat{v}_{t,l,i}^{\eta} \log(\gamma_{l,i}^{\eta}) + \sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{k=1}^{K} \widehat{z}_{t,l,k} \log(\omega_{l,k}) \\ &+ \sum_{t=1}^{T}\sum_{k=1}^{K} \widehat{z}_{t,k} \left[-\lambda_k \sum_{p=1}^{P} f_t(p)\beta_p + \widehat{\eta}_{t,k} \left(\log(\lambda_k) + \log\left(\sum_{p=1}^{P} f_t(p)\beta_p \right) \right) \right] \\ &+ \sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K} \widehat{u}_{t,j}^{\alpha} \widehat{z}_{t,k} \left[(Y_t - \widehat{\eta}_{t,k}) \log(\alpha_j) + (Y_{t-1} - Y_t + \widehat{\eta}_{t,k}) \log(1-\alpha_j) \right]. \end{aligned} \tag{11}$$

The E-step remains unchanged apart from replacing λ_k with $\lambda_{t,k}$, and α_j with $\alpha_j(1-O_t) + \varpi O_t$. The M-steps for λ_k and α_j are modified as $\lambda_k^{(m+1)} = \frac{\sum_{t=1}^{T} \widehat{z}_{t,k}\widehat{\eta}_{t,k}}{\sum_{t=1}^{T} \widehat{z}_{t,k} \sum_{p=1}^{P} f_t(p)\beta_p^{(m)}}$, $\alpha_j^{(m+1)} = \frac{\sum_{t=1}^{T}\sum_{k=1}^{K}(1-O_t)\widehat{u}_{t,j}^{\alpha}\widehat{z}_{t,k}(Y_t - \widehat{\eta}_{t,k})}{\sum_{t=1}^{T}\sum_{k=1}^{K}(1-O_t)\widehat{u}_{t,j}^{\alpha}\widehat{z}_{t,k}Y_{t-1}}$, while the update for β_p is $\beta_p^{(m+1)} = \frac{\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} \widehat{u}_{t,j,k,l} f_t(p)\widehat{\ell}_{t,j,k}}{\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} \widehat{u}_{t,j,k,l} f_t(p)\lambda_k^{(m+1)}}$, where $\widehat{u}_{t,j,k,l} = P(S_t^\alpha =$

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

$j, Z_t = k, S_t^\eta = l | \mathbf{Y}_{0:T})$ and $\widehat{\ell}_{t,j,k} = \frac{\lambda_{t,k}^{(m)} P(Y_t = y_t - 1 | Y_{t-1}, Z_t = k, S_t^\alpha = j)}{P(Y_t = y_t | Y_{t-1}, Z_t = k, S_t^\alpha = j)}$. Finally, $\varpi^{(m+1)} = \frac{\sum_{t=1}^T \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L O_t \widehat{u}_{t,j,k,l} (Y_t - \widehat{\ell}_{t,j,k})}{\sum_{t=1}^T O_t Y_{t-1}}$ is the update of ϖ. Since $\lambda_k^{(m+1)}$ is computed using the estimate of β_p obtained at the previous iteration, the resulting algorithm is effectively an expectation conditional maximization (ECM) (see Meng and Rubin, 1993).

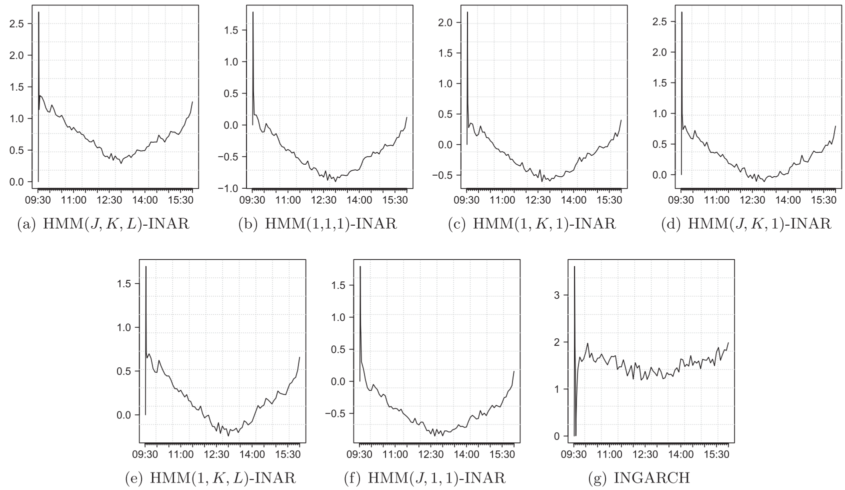
6.2. Estimation

We estimate the seasonal HMM-INAR model using the time series of SPY trades sampled from 2 January 2001 to 16 May 2001 (in-sample period). The in-sample period consists of 73 days and 390 intervals per day for a total of 28,470 observations. The second half of the sample, from 17 May 2001 to 31 August 2021 is used for the out-of-sample analysis, as described in Section 6.3 below. Using BIC for the unrestricted HMM(J, K, L)-INAR model, the selected number of states is $J = 4$, $K = 8$, and $L = 3$. Furthermore, we define $P = 81$ seasons, specified as follows: three seasons for each of the three minutes after market opening, one season for minutes 4 and 5, and 77 seasons for the intervals 6–10 minutes, 11–15 minutes, and so on. The decision to include 81 seasons is inherently arbitrary and reflects the characteristics of the series under consideration. Moreover, due to the extensive length of the series, the seasonal coefficients are estimated with remarkable precision, as outlined by the standard errors provided in Section 1.1 of the Supplementary Material. Thus, with the inclusion of the periodic terms, the total number of parameters to be estimated equals $J(J-1) + L(L-1) + (K-1)L + J + K + P = 132$, with a corresponding BIC value of 153885.3.

Beyond the unrestricted HMM(J, K, L)-INAR model, we consider several restricted HMM-INAR specifications that include the same seasonal components presented in Section 6.1. The restricted specifications, whose optimal number of states is again selected via the BIC, are: the INAR(1), that is, the HMM(1,1,1)-INAR model ($M = 83$, BIC $= 184387.7$); the Mix-INAR, that is, the HMM$(1, K, 1)$-INAR model with $K = 9$ ($M = 99$, BIC$= 160966.4$); one chain HMM-INAR model with $\{S_t^\alpha\}$ and mixture innovations, that is, HMM(J, K, L)-INAR model with $J = 5$ and $K = 6$ ($M = 117$, BIC $= 155379.7$); one chain HMM-INAR model with $\{S_t^\eta\}$, that is, the HMM$(1, K, L)$-INAR model with $K = 7$ and $L = 5$ ($M = 139$, BIC $= 154313.2$); one chain HMM-INAR model with $\{S_t^\alpha\}$ and Poisson innovations, that is, the HMM$(J, 1, 1)$-INAR model with $J = 5$ ($M = 107$, BIC $= 165750.7$). We also consider a periodic version of the Negative Binomial Softplus INGARCH(1,1) model proposed recently by Weiß, Zhu, and Hoshiyar (2022) (labeled INGARCH; see Section 2 of the Supplementary Material for details) with the same seasonal specification as in the HMM-INAR model. The INGARCH model has a total of $M = 85$ parameters, and the BIC is 158305.3.[6]

[6]In Section 3 of the Supplementary Material, we analyze versions of the HMM-INAR and INGARCH models where periodicity is introduced via a parsimonious harmonic term, following Rossi and Fantazzini (2014). For the HMM-

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press



(a) HMM(J,K,L)-INAR (b) HMM(1,1,1)-INAR (c) HMM($1,K,1$)-INAR (d) HMM($J,K,1$)-INAR

(e) HMM($1,K,L$)-INAR (f) HMM($J,1,1$)-INAR (g) INGARCH

FIGURE 6. Estimates of β_p in logarithmic values in the unrestricted HMM(J,K,L)-INAR model, its sub-models, and the INGARCH of Weiß et al. (2022). The trading day from 9:30 AM to 16:00 PM is divided into 81 intervals of 5 minutes each. The first coefficient is $\log(\beta_1) = 0$ by construction.

Figure 6 reports the estimates of the seasonal parameters (β_p) in logarithmic scale for a number of alternative specifications. The parameter estimates and the standard errors are reported in Section 1.1 of the Supplementary Material. In particular, as shown in Figure 6, the estimates of β_p are similar in all HMM-INAR specifications and they reflect the intra-daily pattern of the trading intensity, which is higher on average, particularly at the beginning of the trading day.

With regard to the autoregressive parameter, the states of $\{S_t^\alpha\}$ are associated with estimates of α ranging from 0.067 to 0.777, thus indicating that part of the dynamics of the series can be attributed to the seasonal dummies associated with the parameters β_p. Furthermore, the parameter governing the opening of the trading session in the thinning operator, ϖ, is estimated at zero in all specifications, signaling the absence of persistence of the number of trades at the opening as a consequence of the long overnight break in the transactions. The transition matrix $\hat{\mathbf{\Gamma}}^\alpha$ does not display a clear pattern, meaning that it is quite likely that $\{S_t^\alpha\}$ switches state at each point in time. Concerning the estimates of the parameters associated with $\{Z_t\}$ and $\{S_t^\eta\}$, we note that the BIC selects a mixture with a large number of components, whose weights are governed by the $(K \times L)$ matrix $\hat{\mathbf{\Omega}}$, which allows us to identify three clear patterns (low, intermediate, and high mean/volatility) in most cases. The $(L \times L)$ transition matrix $\mathbf{\Gamma}^\eta$ is much more polarized than $\mathbf{\Gamma}^\alpha$ in

INAR model, the results are qualitatively similar to those presented here, particularly for the in-sample analysis (Figures S2 and S3 in the Supplementary Material), although the out-of-sample fit slightly worsens (Table S2 in the Supplementary Material). The INGARCH model shows slightly improved results. The estimated intra-daily seasonal patterns align well with the empirical pattern shown in Panel d) of Figure 5.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

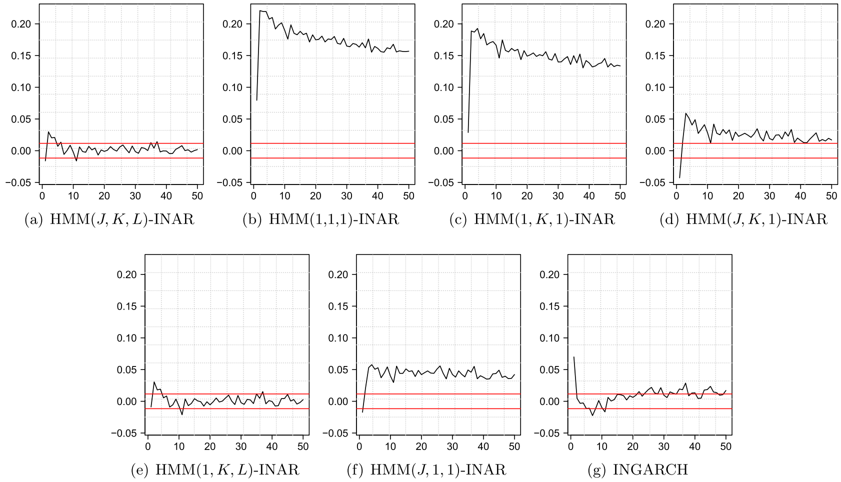


FIGURE 7. The ACF of the standardized residuals for the unrestricted HMM(J,K,L)-INAR model, its sub-models, and the INGARCH of Weiß et al. (2022). The red lines indicate the 95% confidence intervals under the null hypothesis of zero autocorrelation. Standardized residuals are computed as $\hat{v}_t = (Y_t - \mathbb{E}[Y_t|\mathbf{Y}_{1:t-1};\hat{\boldsymbol{\theta}}])/\sqrt{\mathbb{V}ar[Y_t|\mathbf{Y}_{1:t-1};\hat{\boldsymbol{\theta}}]}$.

all cases. This indicates the presence of persistent states with a low probability of switching from one state to another.

The standardized residual ACFs shown in Figure 7 indicate that the unrestricted HMM(J,K,L)-INAR model effectively captures the dynamic patterns of the series, whereas both the simple INAR(1) and the Mix-INAR fail to capture the dynamics of trade counts. Some residual autocorrelation remains in both HMM(J,K,L)-INAR and HMM($J,1,1$)-INAR models, whereas the models with $L>1$, such as the HMM($1,K,L$)-INAR model, as well as the INGARCH model exhibit no residual autocorrelation.

Finally, we evaluate the accuracy of the probability predictions of the HMM-INAR model and compare the results. Figure 8 presents the in-sample randomized probability integral transform (PIT) as proposed by Brockwell (2007).[7] The unrestricted HMM(J,K,L)-INAR model, along with the HMM($1,K,1$)-INAR, HMM($J,K,1$)-INAR, and HMM($1,K,L$)-INAR, exhibit a similar fit to the empirical distribution, whereas other model specifications show PITs that deviate significantly from uniformity. When combined with the residual ACF analysis, these findings underscore the strong performance of the HMM-INAR model with only the chain $\{S_t^{\eta}\}$ (i.e., the HMM($1,K,L$)-INAR model). This suggests that

[7] An alternative goodness-of-fit test for count process models, introduced by Fokianos and Neumann (2013), is based on smoothed versions of the empirical process of Pearson residuals.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

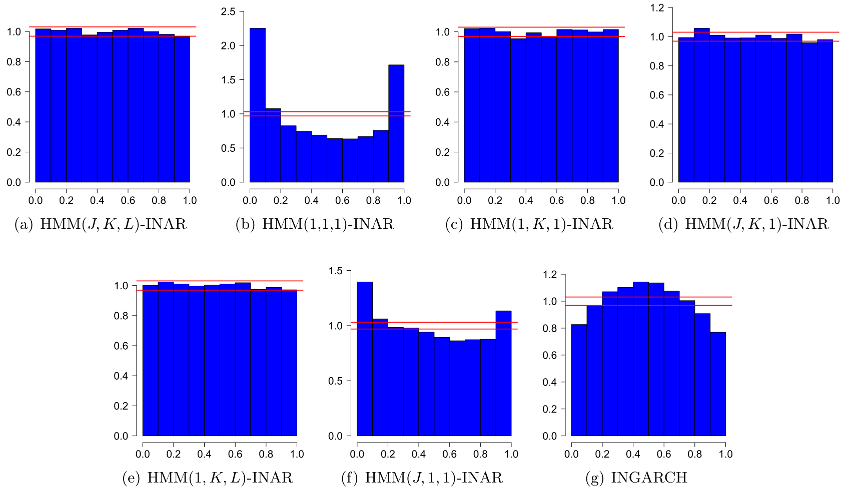


FIGURE 8. In-sample randomized PIT for the unrestricted HMM(J, K, L)-INAR model, its sub-models, and the INGARCH. The red lines indicate the 95% confidence interval under the null hypothesis of being $\mathcal{U}(0, 1)$, as in Diebold, Gunther, and Tay (1998).

incorporating the overnight/opening effect on $A_{t;S_t^\alpha}$, as defined in (10), may be sufficient to achieve an adequate empirical fit.

6.3. Forecast

We now extend the in-sample analysis and evaluate the quality of the forecasts provided by the HMM-INAR model both in terms of point and probability forecasts in the out-of-sample period. The out-of-sample period consists of the number of trades recorded from 17 May 2001 to 31 August 2001, for a total of 73 trading days (and 28,470 observations) after the in-sample period. The parameters of the model are kept constant at their estimated values throughout the entire out-of-sample period. This rather extreme setting for the out-of-sample period allows us to highlight the flexibility of the HMM-INAR model, namely, its ability to adapt to possibly mutated market conditions.

We first look at the quality of the point forecasts and compare the out-of-sample predictions of the HMM-INAR model with those of the alternative specifications used in the in-sample analysis. We follow Freeland and McCabe (2004b) and produce coherent forecasts by computing the one-step ahead predictive median of the number of trades, that is, Median$[N_{t+1}|\mathbf{N}_{1:t}]$. We add two naive predictors: the random walk (RW), that is, Median$[N_{t+1}|\mathbf{N}_{1:t}] = N_t$, and the one-day-lag random walk (RW390), that is, Median$[N_{t+1}|\mathbf{N}_{1:t}] = N_{t+1-p}$. Table 3 reports the median absolute forecast error (MAFE) of all model specifications relative to those of the unrestricted HMM(J, K, L)-INAR model. A value of the ratio larger

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

TABLE 3. MAFE for the unrestricted HMM(J,K,L)-INAR model, the INAR(1), i.e., the HMM(1,1,1)-INAR, the Mix-INAR, i.e., the HMM($1,K,1$)-INAR, the HMM($J,K,1$)-INAR, the HMM($1,K,L$)-INAR, and the HMM($J,1,1$)-INAR

(J,K,L)	(1,1,1)	($1,K,1$)	($J,K,1$)	($1,K,L$)	($J,1,1$)	RW	RW78	INGARCH	Periods
1.00	1.13	1.10	1.03	1.02	1.04	1.56	2.60	1.05	Full
1.00	1.20	1.20	1.25	0.88	1.12	1.53	1.28	0.97	Opening
1.00	1.10	1.07	1.07	1.01	1.05	1.61	2.52	1.07	Midday
1.00	1.19	1.13	1.03	1.02	1.03	1.60	2.61	1.04	Closing

Note: We also consider the random walk (RW), the random walk with lag length equal to 390 periods (RW390), and the Periodic Negative Binomial Softplus INGARCH(1, 1) (INGARCH). The results are reported relative to the HMM(J,K,L)-INAR model. Green (red) cells indicate significant superior (inferior) predictive ability at the 5% confidence level of each model with respect to the HMM-INAR model according to the Diebold and Mariano (1995) test statistics computed with HAC standard errors. The results are reported for the entire day (Day), the first 15 minutes of trading activity (Opening), the period between 11:45 and 12:00 (Midday), and the last 15 minutes of trading activity (Closing).

(smaller) than one indicates under-performance (over-performance) of the j-th model with respect to the HMM(J,K,L)-INAR model. The table also indicates whether the differences in the quality of the forecasts are statistically significant (at 5% significance level) by means of the Diebold and Mariano (1995) test. In almost all cases, the HMM(J,K,L)-INAR model provides significantly superior point forecasts compared to those obtained by the other specification. In only one case, the performance of the unrestricted HMM(J,K,L)-INAR model is significantly worse than that of the competing models. This is the case of the HMM($1,K,L$)-INAR model and of the INGARCH at the opening of the trading day. Overall, this evidence suggests that the unrestricted HMM(J,K,L)-INAR specification is superior to the other specifications being the only one with the sufficient degree of flexibility required for a good fit to the data at hand.

We also assess the ability of each model specification to provide a good fit of the extreme over-dispersion in the data. For this reason, we compute the unconditional ID for each intra-daily interval, $\mathrm{ID}_p, p = 1, \ldots, 81$, where the latter is computed as

$$\mathrm{ID}_p = \frac{\mathbb{V}ar[Y_t|f_t(p) = 1]}{\mathbb{E}[Y_t|f_t(p) = 1]},$$

where $\mathbb{V}ar[Y_t|f_t(p) = 1]$ and $\mathbb{E}[Y_t|f_t(p) = 1]$ are computed by exploiting the results of Section 3.2. Let $\widehat{\mathrm{ID}}_p$ be the sample counterpart of ID_p and let $\widehat{\mathrm{ID}}_p^{\mathcal{M}_i}$ be the index computed according to model $\mathcal{M}_i$. We look at the ratio

$$R_p^{\mathcal{M}_i} = \frac{\widehat{\mathrm{ID}}_p}{\widehat{\mathrm{ID}}_p^{\mathcal{M}_i}},$$

and we report its average value over different intra-daily periods (see Table 4). A value of $R_p^{\mathcal{M}_i}$ lower (higher) than one indicates that a given model is overestimating

Table 4. Sample and model-based ID and excess of kurtosis ratios, R_p

ID						Ex. Kurt.						Period
(J,K,L)	(1,1,1)	$(1,K,1)$	$(J,K,1)$	$(1,K,L)$	$(J,1,1)$	(J,K,L)	(1,1,1)	$(1,K,1)$	$(J,K,1)$	$(1,K,L)$	$(J,1,1)$	
1.77	10.07	1.40	1.62	1.84	6.05	0.74	3.03	1.00	0.86	1.07	2.30	Full
1.01	2.89	1.21	1.06	1.04	1.49	1.00	5.36	1.78	1.51	1.80	3.93	Opening
0.68	2.95	0.80	0.74	0.71	1.11	0.34	2.21	0.57	0.51	0.62	1.58	Midday
1.06	4.08	1.18	1.09	1.10	2.05	0.79	4.42	1.36	1.17	1.41	3.21	Closing

Note: The models considered are: the unrestricted HMM(J,K,L)-INAR model, the INAR(1), i.e., the HMM(1,1,1)-INAR, the Mix-INAR, i.e., the HMM$(1,K,1)$-INAR, the HMM(J,K,L)-INAR, the HMM$(1,K,L)$-INAR, and the HMM$(J,1,1)$-INAR. The results are reported for the entire day (Day), the first 15 minutes of trading activity (Opening), the period between 11:45 and 12:00 (Midday), and the last 15 minutes of trading activity (Closing).

(underestimating) the empirical ID in a given intra-daily period. The ID for the unrestricted HMM(J,K,L)-INAR model is generally quite close to 1, with the exception of the Midday period, where the model tends to overestimate the empirical dispersion index. The baseline INAR(1) and the Mix-INAR do not generally provide a good fit of the ID and they tend to underestimate it, especially the baseline INAR. Only the HMM(J,K,L)-INAR and the HMM$(1,K,L)$-INAR models provide a fit to the empirical over-dispersion analogous to that of the unrestricted HMM(J,K,L)-INAR model. Instead, the restricted HMM$(J,1,1)$-INAR model tends to predict a low value of ID in most cases. A similar pattern arises if we look at the fit of the empirical kurtosis.

We also examine the autocorrelation of the forecast errors, which are defined here as $\hat{v}_t = Y_t - \mathbb{E}[Y_t|\mathbf{Y}_{1:t-1}]$. Figure 9 indicates that the forecast errors of the unrestricted HMM(J,K,L)-INAR model exhibit minimal autocorrelation, whereas the INAR(1) and HMM$(1,K,1)$-INAR models fail to adequately capture the dynamics of the number of trades even in the out-of-sample period. Again, the ACF of the forecast errors of the HMM$(1,K,L)$-INAR model is in line with that of the unrestricted model. Contrary to this, the other restricted models, namely, HMM(J,K,L)-INAR and HMM$(J,1,1)$-INAR, slightly underestimate the persistence of trade counts in the out-of-sample period. In contrast, the INGARCH model produces negative forecast error autocorrelation up to the tenth lag.

Finally, we evaluate the quality of the forecasts in terms of density predictions. Figure 10 presents the histograms of the out-of-sample randomized PITs for all the considered model specifications. In terms of fit to the empirical distribution, the results are largely consistent with those obtained for the in-sample period. Overall, combining this evidence with the PIT analysis, we conclude that the unrestricted HMM(J,K,L)-INAR model and the HMM$(1,K,L)$-INAR, that is, the most parametrized models in our application with $M = 132$ and $M = 139$ coefficients, exhibit similar performance both in-sample and out-of-sample.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

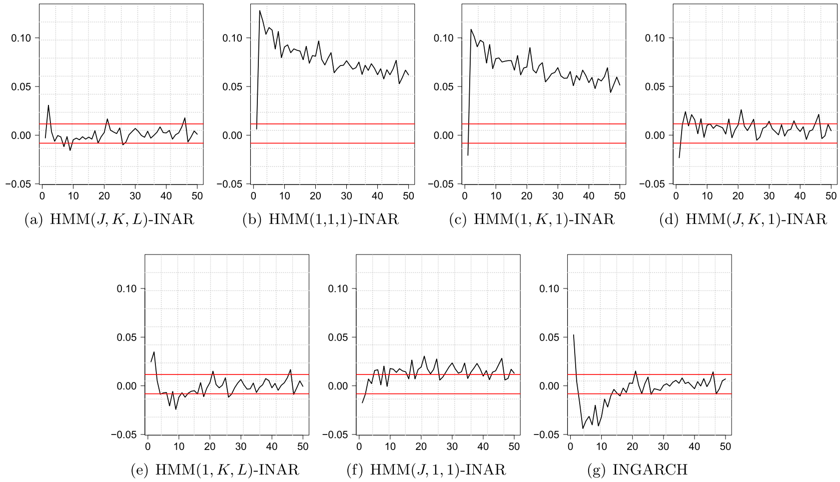


FIGURE 9. The ACF of the forecast errors for the unrestricted HMM(J,K,L)-INAR model, its submodels, and the INGARCH. The red lines indicate the 95% confidence intervals under the null hypothesis of zero autocorrelation.

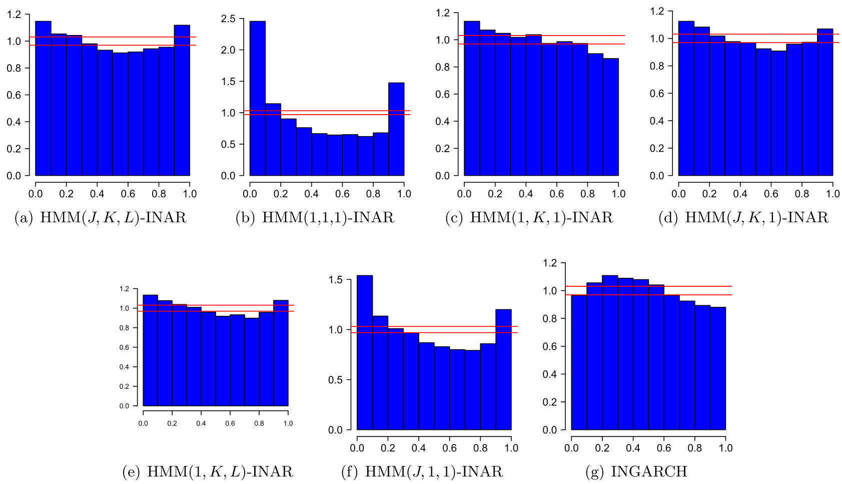


FIGURE 10. Out-of-sample randomized PIT for the unrestricted HMM(J,K,L)-INAR model, its submodels, and the INGARCH. The red lines indicate the 95% confidence intervals under the null hypothesis of being $\mathcal{U}(0,1)$ as in Diebold et al. (1998).

7. CONCLUSION

In this article, we propose and study a new modeling framework to deal with time series of counts that are characterized by high over-dispersion and persistence.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

In particular, we add a flexible hidden Markov chain structure to the baseline INAR(1). The Markov structure is made up of two independent chains: one on the binomial thinning operator and the other on the innovation term. Together, these determine both the autocorrelation structure and the conditional distribution of the process. The HMM-INAR model has appealing theoretical properties, which make inference possible and tractable by ML via the EM algorithm. The empirical analysis shows that the HMM-INAR model outperforms other specifications for count data previously proposed in the literature in fitting the data and in out-of-sample forecasting. The flexible structure of the HMM-INAR model also lends itself to extensions. For instance, the model can include seasonal terms or a compound structure. Alternatively, it can be adapted to the analysis of multivariate time series of counts along the lines of Livsey et al. (2018) and Fokianos et al. (2020), among others. We leave this to future research.

A. PROOFS

A.1. Proof of Lemma 1

Proof. Irreducibility and time homogeneity follow from

$$\begin{aligned}P(\mathcal{S}_1 = h|\mathcal{S}_0 = j) &= P(S_1^\alpha = h_1, Z_1 = h_2, S_1^\eta = h_3|S_0^\alpha = j_1, Z_0 = j_2, S_0^\eta = j_3)\\ &= \omega_{h_3,h_2}\gamma_{j_3,h_3}^\eta\gamma_{j_1,h_1}^\alpha > 0,\end{aligned}$$

by Assumption 1 and by using the unique mapping between the indexes of $\{\mathcal{S}_t\}$ and those of $\{(S_t^\alpha, Z_t, S_t^\eta)\}$, i.e., $(\mathcal{S}_t = h)$ denotes $(S_t^\alpha = h_1, Z_t = h_2, S_t^\eta = h_3)$, for $h = 1,\dots,H$. A similar argument shows that $\{\mathcal{S}_t\}$ is first-order Markov. Since the chain is irreducible and the state $j \in \mathbb{H}$ is aperiodic, i.e., $P(\mathcal{S}_1 = j|\mathcal{S}_0 = j) > 0$, the rest of the states are aperiodic. Finally, stationarity of $\{\mathcal{S}_t\}$ follows by construction, i.e., $P(\mathcal{S}_t = h) = P(S_t^\alpha = h_1, Z_t = h_2, S_t^\eta = h_3) = P(Z_t = h_2|S_t^\eta = h_3)P(S_t^\eta = h_3)P(S_t^\alpha = h_1) = \omega_{h_3,h_2}\pi_{h_3}^\eta\pi_{h_1}^\alpha$ is constant with respect to t. □

A.2. Proof of Lemma 2

Proof. The proof follows from Lemmas 1 and 2 in Tang and Wang (2014), as well as Tong (1990). The time homogeneity of the Markov chain $\{(Y_t, \mathcal{S}_t)\}$ follows from: *i*) the time homogeneity of $\{\mathcal{S}_t\}$ (Lemma 1), *ii*) the independence of $\{\mathcal{S}_t\}$ from $\{X_{t,n,j}\}$ and $\{\eta_{t,k}\}$, for all t,n,j,k, and *iii*) the stationarity of $\{X_{t,n,j}\}$ and $\{\eta_{t,k}\}$. Indeed, for a given $0 \le s < t$, it follows that

$$\begin{aligned}&P\left((Y_{t+1}, \mathcal{S}_{t+1}) = (y_{t+1}, h)|(Y_t, \mathcal{S}_t) = (y_t, m), (Y_s, \mathcal{S}_s) = (y_s, m_s)\right)\\ &= P\left(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1}, \mathcal{S}_{t+1} = h|(Y_t, \mathcal{S}_t) = (y_t, m), (Y_s, \mathcal{S}_s) = (y_s, m_s)\right)\\ &= P\left(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1}|(Y_t, \mathcal{S}_t) = (y_t, m)\right)P(S_{t+1} = h|Y_t = y_t, S_t = m)\\ &= P\left(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1}|Y_t = y_t, \mathcal{S}_t = m\right)P(S_{t+1} = h|S_t = m)\\ &= P(S_{t+1} = h|S_t = m)P\left(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1}\right),\end{aligned}$$

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

since $\{\eta_{t+1,h_2}\}$ and $\{\mathcal{S}_t\}$ are mutually independent by Assumption 1 and $P(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1})$ is given in (5). Furthermore, it holds that

$$P((Y_{t+1},\mathcal{S}_{t+1}) = (y_{t+1},h)|(Y_t,\mathcal{S}_t) = (y_t,m)) = P(\mathcal{S}_{t+1} = h|\mathcal{S}_t = m)P\left(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1}\right).$$

Hence,

$$\begin{aligned}&P\left((Y_{t+1},\mathcal{S}_{t+1}) = (y_{t+1},h)|(Y_t,\mathcal{S}_t) = (y_t,m),(Y_s,\mathcal{S}_s)\right)\\&\quad = P\left((Y_{t+1},\mathcal{S}_{t+1}) = (y_{t+1},h)|(Y_t,\mathcal{S}_t) = (y_t,m)\right),\end{aligned}$$

which means that $\{(Y_t,\mathcal{S}_t)\}$ is first-order Markov. Finally, time homogeneity follows from the time homogeneity of $\{\mathcal{S}_t\}$ (Lemma 1), and the stationarity of $\{X_{t,n,j}\}$ and $\{\eta_{t,k}\}$ for all n,j,k (Assumption 1).

As for the proof of part b) of Lemma 2, the $\mu \times \varphi$ irreducibility of $\{(Y_t,\mathcal{S}_t)\}$ follows from the irreducibility of $\{\mathcal{S}_t\}$ by Lemma 2 in Tang and Wang (2014). Indeed, given that $\gamma^{\mathcal{S}(t)}_{m,h} := P(\mathcal{S}_{t+s} = h|\mathcal{S}_s = m) > 0 \quad \forall t > 0$ by irreducibility, then considering t steps, we obtain $\gamma^{\mathcal{S}}_{m,m^{(1)}}\gamma^{\mathcal{S}}_{m^{(1)},m^{(2)}}\gamma^{\mathcal{S}}_{m^{(2)},m^{(3)}}\cdots\gamma^{\mathcal{S}}_{m^{(t-1)},h} > 0$. Hence, recalling that

$$\begin{aligned}&P\left((Y_{t+1},\mathcal{S}_{t+1}) = (y_{t+1},h)|(Y_t,\mathcal{S}_t) = (y_t,m)\right)\\&\quad = P(\mathcal{S}_{t+1} = h|\mathcal{S}_t = m)P\left(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1}\right),\end{aligned}$$

$\forall (y_t,m) \in (\mathbb{N} \times \mathbb{H})$, where $P\left(\alpha_{h_1} \circ y_t + \eta_{t+1,h_2} = y_{t+1}\right)$ is given in (5). Then,

$$\begin{aligned}&p^{(t)}\left((y_s,m),(y_{t+s},h)\right) := P\left((Y_{t+s},\mathcal{S}_{t+s}) = (y_{t+s},h)|(Y_s,\mathcal{S}_s) = (y_s,m)\right)\\&\quad = \sum_{(m^{(1)},m^{(2)},\cdots,m^{(t-1)})\in\mathbb{H}^{t-1}} \gamma^{\mathcal{S}}_{m,m^{(1)}}\gamma^{\mathcal{S}}_{m^{(1)},m^{(2)}}\gamma^{\mathcal{S}}_{m^{(2)},m^{(3)}}\cdots\gamma^{\mathcal{S}}_{m^{(t-1)},h}\\&\qquad \times \sum_{w_1,w_2,\ldots,w_{t-1}\in\mathbb{N}^{t-1}}\sum_{r_1=0}^{y_s\wedge w_1}\sum_{r_2=0}^{w_1\wedge w_2}\cdots\sum_{r_t=0}^{w_{t-1}\wedge y_t} P\left(\alpha_{m_1^{(1)}} \circ y_s + \eta_{s+1,m_2^{(1)}} = w_1\right)\\&\qquad \times P\left(\alpha_{m_1^{(2)}} \circ w_1 + \eta_{s+2,m_2^{(2)}} = w_2\right)\cdots P\left(\alpha_{m_1^{(t)}} \circ y_{t-1} + \eta_{s+t,m_2^{(t-1)}} = y_{t+s}\right) > 0,\end{aligned}$$

so that the Markov chain $\{(Y_t,\mathcal{S}_t)\}$ is $\mu \times \varphi$ irreducible.[8] Positivity follows from the fact that all transition probabilities are positive ($\gamma^{\mathcal{S}}_{i,j} > 0$ for all i,j) and from the fact that there exists at least one ergodic state of $\{S^{\alpha}_t\}$, say h^*_1, with $\alpha_{h^*_1} < 1$, which implies that $P\left(\alpha_{h^*_1} \circ y' + \eta_{t,k} = y\right) > 0$ for all $(y,y') \in \mathbb{N}^2$ and all k. Aperiodicity of $\{(Y_t,\mathcal{S}_t)\}$ follows directly from Proposition A1.2 of Tong (1990) and Chan (1990). □

A.3. Proof of Theorem 1

Proof. By an application of Lemma 2, it follows that $\{(Y_t,\mathcal{S}_t)\}$ is a $(\mu \times \varphi)$–irreducible, aperiodic Markov chain. To prove that $\{(Y_t,\mathcal{S}_t)\}$ is geometrically ergodic, we follow Theorem 1 in Tang and Wang (2014). They show that the conditions of Theorem 3.1 by

[8] Here, we use the notation $\mathbb{H}^{t-1} = \bigotimes_{q=1}^{t-1}\mathbb{H}$, and similarly for $\mathbb{N}^{t-1}$. Also, the indexes $m^{(1)}_u$ and $m^{(2)}_u$ are computed by mapping the state $m^{(u)}$ of the enlarged system to the original HMM-INAR specification, for $u = 1,\ldots,t-1$.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

Tweedie (1975) are satisfied for the nonnegative measurable function $g(y,i) = \sqrt{y^2+i^2}$, i.e., the Euclidean norm of $(y,i) \in (\mathbb{N} \times \mathbb{H})$. In our case, we have

$$\begin{aligned}\mathbb{E}[g(Y_1,\mathcal{S}_1)|Y_0 = y,\mathcal{S}_0 = i] &= \mathbb{E}[g(\alpha_{\mathcal{S}_1} \circ Y_0 + \eta_1,\mathcal{S}_1)|Y_0 = y,\mathcal{S}_0 = i] \\ &\leq \mathbb{E}[\alpha_{\mathcal{S}_1} \circ y|\mathcal{S}_0 = i] + \mathbb{E}[\eta_1|\mathcal{S}_0 = i] \\ &\leq ay + c + c_0,\end{aligned}$$

where the last inequality is satisfied by taking $c_0 = \max_{i\in\mathbb{H}} \mathbb{E}[\eta_1|\mathcal{S}_0 = i] + \mathbb{E}[\mathcal{S}_1|\mathcal{S}_0 = i]$ and noting that

$$\mathbb{E}[\alpha_{\mathcal{S}_1} \circ y|\mathcal{S}_0 = i] \leq ay + c, \quad \forall (y,i) \in (\mathbb{N} \times \mathbb{H}),$$

where $0 < a < 1$ and $c \geq 0$, is satisfied by taking

$$a = \max_{i\in\mathbb{H}} \sum_{m=1}^{H} \underline{\alpha}_m \gamma_{i,m}^{\mathcal{S}} < 1,$$

which follows from the fact that $\gamma_{i,m}^{\mathcal{S}} > 0$ for all i,m, $\sum_{m=1}^{H} \gamma_{i,m}^{\mathcal{S}} = 1$ for all i, $\underline{\alpha}_m \in (0,1]$ for all m, and $\min_{m\in\mathbb{H}} \underline{\alpha}_m < 1$ (Assumption 1). Here, $\underline{\alpha}_m$ represents the m-th element of $\underline{\boldsymbol{\alpha}}$, which is the vector of probabilities of success in the enlarged specification (see the discussion in Section 3.2).

The rest of the proof follows from Theorem 1 in Tang and Wang (2014). Since $\{(Y_t,\mathcal{S}_t)\}$ is geometrically ergodic, there exists a probability measure π on $(\mathbb{N} \times \mathbb{H}, \mathcal{B} \times \mathcal{H})$ and a constant $\beta \in (0,1)$, such that $\forall (y,m) \in (\mathbb{N} \times \mathbb{H})$

$$\lim_{t\to\infty} \beta^{-t} ||p^{(t)}((y,m),\cdot) - \pi(\cdot)||_\tau = 0,$$

where $||\cdot||_\tau$ denotes the total variation norm. Let π^* be a set function on $(\mathbb{N},\mathcal{B})$ such that $\pi^*(A) = \pi(A \times H), \forall A \in \mathcal{B}$, so that π^* is a probability measure on $(\mathbb{N},\mathcal{B})$. Then, for $Y_0 = y_0$, we get that

$$\begin{aligned}P(Y_t = y_t|Y_0 = y_0) &= \sum_{h\in\mathbb{H}} P(Y_t = y_t,\mathcal{S}_t = h|Y_0 = y_0) \\ &= \sum_{h\in\mathbb{H}} \sum_{m\in\mathbb{H}} P(Y_t = y_t,\mathcal{S}_t = h|Y_0 = y_0,\mathcal{S}_0 = m) \\ &\quad \times P(\mathcal{S}_0 = m|Y_0 = y_0), \quad \forall y_0 \in \mathbb{N},\end{aligned}$$

and, $\forall A \in \mathcal{B}$,

$$\pi^*(A) = \pi(A \times \mathbb{H}) = \sum_{h\in\mathbb{H}} \sum_{m\in\mathbb{H}} \pi(A \times \{h\}) P(\mathcal{S}_0 = m|Y_0 = y_0).$$

Since $\mathbb{H}$ is a finite set, it follows that

$$\lim_{t\to\infty} \beta^{-t} ||P(Y_t \in \cdot|Y_0 = y_0) - \pi^*(\cdot)||_\tau = 0,$$

which implies that π^* is the unique invariant probability measure of $\{Y_t\}$. □

A.4. Theorem 2

A.4.1. *Notation and a Useful Lemma.* Hereafter, for notational convenience, we omit the dependence of $A_{t,\mathcal{S}_t}$ and $\eta_{t,\mathcal{S}_t}$ on $\{\mathcal{S}_t\}$ and denote them as A_t and η_t, respectively.

A.4.2. *Proof of Theorem 2.* Let $Y_{t,m}$ be the process defined as $Y_{t,m} = y$ for $t \leq m$, where $y \in \mathbb{N}$ is arbitrary and $Y_{t,m} = A_{t,m} + \eta_{t,m}$, with $A_{t,m} = \alpha_{\mathcal{S}_t^\alpha} \circ Y_{t-1,m}$, and $\eta_{t,m}|\mathcal{S}_{t,m} = h \sim Pois(\lambda_h)$. Let $\mathbf{Y}_{(1),t,m}$ denote the vector $(\mathbb{E}[Y_{t,m}|\mathcal{S}_{t,m} = h], h = 1,\dots,H)'$, and set $P(\mathcal{S}_{t,m} = h) = \pi_h$, for $t \leq m$ and $P(\mathcal{S}_{t,m} = j|\mathcal{S}_{t-1,m} = i) = \gamma_{i,j}$. By construction, it follows that $P(\mathcal{S}_{t,m} = h) = \pi_h$, for $t > m$, because $\boldsymbol{\pi}$ is the stationary distribution associated with $\boldsymbol{\Gamma}$. We have $\mathbb{E}[Y_{t,m}|\mathcal{S}_{t,m} = h] = \mathbb{E}[A_{t,m}|\mathcal{S}_{t,m} = h] + \mathbb{E}[\eta_{t,m}|\mathcal{S}_{t,m} = h]$, where

$$
\begin{aligned}
&\mathbb{E}[A_{t,m}|\mathcal{S}_{t,m} = h] = \sum_{q=1}^{H} \mathbb{E}[A_{t,m}|\mathcal{S}_{t,m} = h, \mathcal{S}_{t-1,m} = q] P(\mathcal{S}_{t-1,m} = q|\mathcal{S}_{t,m} = h) \\
&\sum_{q=1}^{H} \mathbb{E}[\mathbb{E}[A_{t,m}|Y_{t-1,m}, \mathcal{S}_{t,m} = h, \mathcal{S}_{t-1,m} = q]|\mathcal{S}_{t,m} = h, \mathcal{S}_{t-1,m} = q]\gamma_{q,h}\frac{\pi_q}{\pi_h} \\
&\sum_{q=1}^{H} \alpha_h \mathbb{E}[Y_{t-1,m}|\mathcal{S}_{t-1,m} = q]\gamma_{q,h}\frac{\pi_q}{\pi_h}.
\end{aligned} \tag{A.1}
$$

Computing the expression above for all $h = 1,\dots,H$ and putting the result in a vector we obtain, for $t > m$, $\mathbf{Y}_{(1),t,m} = \underline{\boldsymbol{\lambda}} + \mathbf{AGY}_{(1),t-1,m}$, where $\mathbf{G} = \boldsymbol{\Pi}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Pi}$ is a stochastic matrix with positive entries, i.e., $\mathbf{G} = [g_{l,r}]$, with $g_{l,r} > 0$, and $g_{l,r} = \gamma_{r,l}\pi_r/\pi_l$. The fact that $g_{l,r} > 0$ for all $l, r = 1,\dots,H$ (which implies also that $g_{l,r} < 1$) follows by Assumption 1 because $\gamma_{r,l} > 0$ for all r, l, which also implies $\pi_h > 0$ for all h. According to the Banach fixed-point theorem, this recursion has a unique solution, independent of the initialization, for $m \to -\infty$, if there exists some $k \in \mathbb{N}$ such that $||(\mathbf{AG})^{(k)}|| < 1$ for some norm $||\cdot||$. Here, we set $||\cdot||$ to the matrix norm induced by the vector 1-norm, i.e., for a $D \times D$ matrix $\mathbf{L} = [l_{i,j}]_{i,j=1}^{D}$, we set $||\mathbf{L}|| = \max\limits_{j=1,\dots,D} \sum_{i=1}^{D} |l_{i,j}|$. Assume without loss of generality that $a_{1,1} = \tilde{\alpha}$, where $\tilde{\alpha} = \min_h \underline{\alpha}_h < 1$ (by Assumption 1) and consider the matrix

$$
\tilde{\mathbf{A}} = \begin{pmatrix} a & \mathbf{0}' \\ \mathbf{0} & \mathbf{I} \end{pmatrix},
$$

where $\mathbf{0}$ is a vector of zeros of length $H-1$ and $\mathbf{I}$ is the identity matrix of dimension $(H-1)\times(H-1)$. Then, $||(\mathbf{AG})^{(k)}|| \leq ||(\tilde{\mathbf{A}}\mathbf{G})^{(k)}||$ for all k. For $k = 1$, we have $||(\tilde{\mathbf{A}}\mathbf{G})|| = 1$. For $k = 2$, we have $||(\tilde{\mathbf{A}}\mathbf{G})^{(2)}|| = ||\tilde{\mathbf{A}}\mathbf{G}\tilde{\mathbf{A}}\mathbf{G}|| \leq ||\tilde{\mathbf{A}}\mathbf{G}\tilde{\mathbf{A}}||\,||\mathbf{G}|| = ||\tilde{\mathbf{A}}\mathbf{G}\tilde{\mathbf{A}}||$ because $||\mathbf{G}|| = 1$. Let $\mathbf{e}_i$ be the i-th unit basis vector in $\mathbb{R}^H$, and let $\boldsymbol{\iota} = \sum_{h=1}^{H} \mathbf{e}_h$. Then, the sum of the first row is given by $\mathbf{e}_1'\tilde{\mathbf{A}}\mathbf{G}\tilde{\mathbf{A}}\boldsymbol{\iota} = \tilde{\alpha}(1 - g_{1,1}(1-\tilde{\alpha})) \in (0,1)$ because $g_{1,1} \in (0,1)$ and $\tilde{\alpha} \in (0,1)$. Furthermore, $\mathbf{e}_h'\tilde{\mathbf{A}}\mathbf{G}\tilde{\mathbf{A}}\boldsymbol{\iota} = 1 - g_{h,1}(1-\tilde{\alpha}) \in (0,1)$, because $g_{1,1} \in (0,1)$ and $\tilde{\alpha} \in (0,1)$, for all $h > 1$. It then follows that $||(\mathbf{AG})^{(2)}|| \leq ||(\tilde{\mathbf{A}}\mathbf{G})^{(2)}|| < 1$, such that $\lim_{m\to-\infty} \mathbf{Y}_{(1),t,m} = (\mathbf{I}_H - \mathbf{AG})^{-1}\underline{\boldsymbol{\lambda}} = \mathbf{Y}_{(1)}$. To compute $\mathbb{E}[Y_t^2]$, we proceed in the same way and write $\mathbb{E}[Y_{t,m}^2] = \mathbb{E}[A_{t,m}^2] + \mathbb{E}[\eta_{t,m}^2] + 2\mathbb{E}[A_{t,m}\eta_{t,m}]$. We have $\mathbb{E}[\eta_{t,m}^2] = \sum_{h=1}^{H} \mathbb{E}[\eta_{t,m}^2|\mathcal{S}_{t,m} = h]\pi_h = \sum_{h=1}^{H} \lambda_h(1+\lambda_h)\pi_h$ and $\mathbb{E}[A_{t,m}\eta_{t,m}] = \sum_{h=1}^{H} \mathbb{E}[A_{t,m}|\mathcal{S}_{t,m} = h]\mathbb{E}[\eta_{t,m}|\mathcal{S}_{t,m} = h]\pi_h$, where $\mathbb{E}[A_{t,m}|\mathcal{S}_{t,m} = h]$ is given in (A.1) and $\mathbb{E}[\eta_{t,m}|\mathcal{S}_{t,m} = h] = \lambda_h$.

To compute $\mathbb{E}[A_{t,m}^2]$, we write $\mathbb{E}[A_{t,m}^2] = \sum_{h=1}^{H} \mathbb{E}[A_{t,m}^2|\mathcal{S}_{t,m} = h]\pi_h$, where

$$\mathbb{E}[A_{t,m}^2|\mathcal{S}_{t,m} = h] = \sum_{q=1}^{H} \mathbb{E}[A_{t,m}^2|\mathcal{S}_{t,m} = h, \mathcal{S}_{t-1,m} = h]P(\mathcal{S}_{t-1,m} = q|\mathcal{S}_{t,m} = h)$$

$$= \sum_{q=1}^{H} \mathbb{E}[\mathbb{E}[A_{t,m}^2|Y_{t-1,m}, \mathcal{S}_{t,m} = h, \mathcal{S}_{t-1,m} = h|\mathcal{S}_{t,m} = h, \mathcal{S}_{t-1,m} = h]g_{h,q}$$

$$= \sum_{q=1}^{H} (\alpha_h(1-\alpha_h)\mathbb{E}[Y_{t-1,m}|\mathcal{S}_{t-1,m} = q] + \alpha_h^2\mathbb{E}[Y_{t-1,m}^2|\mathcal{S}_{t-1,m} = q])g_{h,q}. \qquad \textbf{(A.2)}$$

Collecting the results for all $h = 1, \ldots, H$ in a vector, we obtain

$$\mathbf{Y}_{(2),t,m} = (\mathbf{I}_H + \mathbf{\Lambda})\underline{\boldsymbol{\lambda}} + [\mathbf{A}(\mathbf{I}_H - \mathbf{A}) + 2\mathbf{\Lambda}\mathbf{A}]\mathbf{G}\mathbf{Y}_{(1),t-1,m} + \mathbf{A}\mathbf{A}\mathbf{G}\mathbf{Y}_{(2),t-1,m},$$

and by noting that $||(\mathbf{AAG})^{(k)}|| \leq ||(\mathbf{AG})^{(k)}||$ for all k, and by letting $m \to -\infty$, we conclude that $\lim_{m\to-\infty} \mathbf{Y}_{(2),t,m} = \mathbf{Y}_{(2)}$, where the expression for $\mathbf{Y}_{(2)}$ is provided as in the statement of the theorem. The autocovariances are obtained via repetitive calculations, employing the same arguments used for deriving the first two moments, together with the Markov property of $\mathcal{S}_t$. To compute the first-order autocovariance, note that $Y_tY_{t-1} = A_tY_{t-1} + \eta_tA_{t-1} + \eta_t\eta_{t-1}$, and

$$\mathbb{E}[Y_tY_{t-1}] = \sum_{h=1}^{H}\sum_{q=1}^{H} \mathbb{E}[A_tY_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q]\gamma_{q,r}\pi_q$$
$$+ \sum_{h=1}^{H}\sum_{q=1}^{H} \mathbb{E}[\eta_tA_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q]\gamma_{q,r}\pi_q$$
$$+ \sum_{h=1}^{H}\sum_{q=1}^{H} \mathbb{E}[\eta_t\eta_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q]\gamma_{q,r}\pi_q,$$

where

$$\mathbb{E}[A_tY_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q] = \mathbb{E}[Y_{t-1}\mathbb{E}[A_t|Y_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q]|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q]$$
$$= \alpha_h\mathbb{E}[Y_{t-1}^2|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q]$$
$$= \alpha_h\mathbb{E}[Y_{t-1}^2|\mathcal{S}_{t-1} = q],$$

and $\mathbb{E}[\eta_tA_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q] = \lambda_h\mathbb{E}[A_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q] = \lambda_h\mathbb{E}[A_{t-1}|\mathcal{S}_{t-1} = q] = \lambda_h\alpha_q\mathbb{E}[Y_{t-2}|\mathcal{S}_{t-1} = q]$, and $\mathbb{E}[\eta_t\eta_{t-1}|\mathcal{S}_t = h, \mathcal{S}_{t-1} = q] = \lambda_h\lambda_q$. In matrix form, the expression for $\mathbb{E}[Y_tY_{t-1}]$ is

$$\mathbb{E}[Y_tY_{t-1}] = \boldsymbol{\lambda}'\mathbf{\Gamma}'\mathbf{\Pi}\mathbf{A}\tilde{\mathbf{Y}}_{(1)} + \boldsymbol{\lambda}'\mathbf{\Gamma}'\mathbf{\Pi}\boldsymbol{\lambda} + \boldsymbol{\alpha}'\Gamma'\mathbf{\Pi}\mathbf{Y}_{(2)}, \qquad \textbf{(A.3)}$$

where $\tilde{\mathbf{Y}}_{(1)}$ is a vector with generic element $\mathbb{E}[Y_{t-1}|\mathcal{S}_t = h]$, for $h = 1, \ldots, H$. After noting that $\mathbb{E}[Y_{t-1}|\mathcal{S}_t = h] = \sum_{q=1}^{H} \mathbb{E}[Y_{t-1}|\mathcal{S}_t = h]g_{h,q}$, we obtain $\tilde{\mathbf{Y}}_{(1)} = \mathbf{G}\mathbf{Y}_{(1)}$. Furthermore, since $\mathbb{E}[Y_t|\mathcal{S}_t = h] = \alpha_h\mathbb{E}[Y_{t-1}|\mathcal{S}_t = h] + \lambda_h$, and $\mathbb{E}[Y_t|\mathcal{S}_t = h] = \mathbb{E}[Y_{t-1}|\mathcal{S}_{t-1} = h]$ (which was proved previously), we obtain the following identity $\mathbf{A}\tilde{\mathbf{Y}}_{(1)} = \mathbf{Y}_{(1)} - \boldsymbol{\lambda}$. By replacing $\mathbf{A}\tilde{\mathbf{Y}}_{(1)}$ in (A.3), we obtain $\mathbb{E}[Y_tY_{t-1}] = \boldsymbol{\lambda}'\mathbf{\Gamma}'\mathbf{\Pi}\mathbf{Y}_{(1)} + \boldsymbol{\alpha}'\Gamma'\mathbf{\Pi}\mathbf{Y}_{(2)}$, which is the expression in

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

Theorem 2 when $k = 1$. The derivations for $k > 1$ and the remaining cross moments proceed in an analogous manner.

A.5. Theorem 3

A.5.1. *Notation and Useful Lemmas.* We introduce the following notation: $||\cdot||$ indicates the sup norm, i.e., for $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^n$, $||\mathbf{x}|| = \max_{i=1,\dots,n} |x_i|$, where x_i is the i-th element of $\mathbf{x}$, and $\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$ indicates the expectation evaluated at $\boldsymbol{\theta}$. Analogously, $\mathbb{V}ar_{\boldsymbol{\theta}}[\cdot]$ and $\mathbb{C}ov_{\boldsymbol{\theta}}[\cdot]$ indicate variance and covariance evaluated at $\boldsymbol{\theta}$. When required, the dependence of $P(Y_t|Y_{t-1},\mathcal{S}_t)$ on $\boldsymbol{\theta}$ is denoted as $P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t)$. The operators returning the gradient and Hessian matrix with respect to $\boldsymbol{\theta}$ are denoted as $\nabla_{\boldsymbol{\theta}}$ and $\nabla^2_{\boldsymbol{\theta}}$, respectively. $\dot{\boldsymbol{\Theta}}$ indicates the interior of $\boldsymbol{\Theta}$, i.e., $\dot{\boldsymbol{\Theta}} = int(\boldsymbol{\Theta})$. The following lemmas are useful for proving Theorem 3.

LEMMA 3. *Under the assumptions of Theorem 3,*

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\left|\left|\nabla_{\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)\right|\right|^2\right] < \infty, \tag{A.4}$$

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\left|\left|\nabla^2_{\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)\right|\right|\right] < \infty. \tag{A.5}$$

Proof. Let $\boldsymbol{\theta} \in \dot{\boldsymbol{\Theta}}$, and let θ_m be its generic element. We define

$$\dot{p}_{h,t;\theta_m} = \frac{\partial \log P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)}{\partial\theta_m}, \qquad \ddot{p}_{h,t;\theta_m,\theta_n} = \frac{\partial^2 \log P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)}{\partial\theta_m\partial\theta_n},$$

for $h \in \mathbb{H}$, where $(\mathcal{S}_t = h) \equiv (S_t^{\alpha} = h_1, Z_t = h_2, S_t^{\eta} = h_3)$. With a slight adaptation of Proposition 3 of Freeland and McCabe (2004a),[9] we obtain the following representation for $\dot{p}_{h,t;\theta_m}$ and $\ddot{p}_{h,t;\theta_m,\theta_n}$:

$$\dot{p}_{h,t;\theta_m} = \begin{cases} \frac{1}{\alpha_{h_1}(1-\alpha_{h_1})}\left\{\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right] - \mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_{t-1},\mathcal{S}_t=h\right]\right\}, & \text{if } \theta_m = \alpha_{h_1}, \\ \frac{1}{\lambda_{h_2}}\left\{\mathbb{E}_{\boldsymbol{\theta}}\left[\eta_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right] - \mathbb{E}_{\boldsymbol{\theta}}\left[\eta_t|Y_{t-1},\mathcal{S}_t=h\right]\right\}, & \text{if } \theta_m = \lambda_{h_2}, \\ 0, & \text{otherwise,} \end{cases} \tag{A.6}$$

and if $\theta_m = \theta_n = \alpha_{h_1}$, then

$$\begin{aligned}\ddot{p}_{h,t;\alpha_{h_1},\alpha_{h_1}} = \frac{1}{\alpha^2_{h_1}(1-\alpha_{h_1})^2}&\left\{(2\alpha_{h_1}-1)\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right]\right.\\ &\left.+\mathbb{V}ar_{\boldsymbol{\theta}}[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h] - \alpha_{h_1}\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_{t-1},\mathcal{S}_t=h\right]\right\},\end{aligned}$$

if $\theta_m = \alpha_{h_1}$ and $\theta_n = \lambda_{h_2}$, then $\ddot{p}_{h,t;\alpha_{h_1},\lambda_{h_2}} = \frac{1}{\alpha_{h_1}\lambda_{h_2}}\mathbb{C}ov_{\boldsymbol{\theta}}[A_t,\eta_t|Y_t,Y_{t-1},\mathcal{S}_t=h]$, if $\theta_m = \lambda_{h_2}$ and $\theta_n = \lambda_{h_2}$, then $\ddot{p}_{h,t;\lambda_{h_2},\lambda_{h_2}} = \frac{1}{\lambda^2_{h_2}}\{\mathbb{V}ar_{\boldsymbol{\theta}}[\eta_t|Y_t,Y_{t-1},\mathcal{S}_t=h] - \mathbb{E}_{\boldsymbol{\theta}}[\eta_t|Y_t,Y_{t-1},\mathcal{S}_t=$

[9] The only difference between Freeland and McCabe (2004a) and this article is the conditioning event $\mathcal{S}_t = h$ which arises naturally when considering their Propositions 1 and 2.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

$h]\}$, and $\ddot{p}_{h,t;\theta_m,\theta_n} = 0$ otherwise. Given $A_t \leq Y_{t-1}$ a.s., it follows that $\mathbb{E}_{\boldsymbol{\theta}_0}[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\dot{p}^2_{h,t;\theta_m}] \leq 4\mathbb{E}_0[Y^2_{t-1}] < \infty$ (from Theorem 2) when $\theta_m = \alpha_{h_1}$. Regarding the case $\theta_m = \lambda_{h_2}$, we note that:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{E}_{\boldsymbol{\theta}}\left[\eta_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right]^2\right]\\
&\quad= \mathbb{E}_{\boldsymbol{\theta}_0}\left[\left(Y_t - \sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right]\right)^2\right]\\
&\quad\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_t^2\right] + \mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right]^2\right]\\
&\qquad+2\mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_t\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right]\right]\\
&\quad\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_t^2\right] + \mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_{t-1}^2\right] + 2\mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_t^2\right]\mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_{t-1}^2\right]\\
&\quad= 2\mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_t^2\right]\left(1+\mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_t^2\right]\right) < \infty,
\end{aligned}\tag{A.7}$$

where the first inequality follows by noting that $Y_t\cdot\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right] > 0$ a.s., and in the second one, we used the Cauchy–Schwarz inequality. The last equality follows from the stationarity of the model. Furthermore, $\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{E}_{\boldsymbol{\theta}}\left[\eta_t|Y_{t-1},\mathcal{S}_t=h\right] = \sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\lambda_{h_2} < \infty$, because $\boldsymbol{\Theta}$ is compact, such that $\mathbb{E}_{\boldsymbol{\theta}_0}[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\dot{p}^2_{h,t;\theta_m}] < \infty$ when $\theta_m = \lambda_{h_2}$. Regarding the second derivatives, we note that:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{V}ar_{\boldsymbol{\theta}}[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h]\right] &\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\mathbb{E}_{\boldsymbol{\theta}}[A_t^2|Y_t,Y_{t-1},\mathcal{S}_t=h]\right]\\
&\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_{t-1}^2\right] < \infty,
\end{aligned}$$

such that $\mathbb{E}_{\boldsymbol{\theta}_0}[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}|\ddot{p}_{h,t;\alpha_{h_1},\alpha_{h_1}}|] < \infty$. That $\mathbb{E}_{\boldsymbol{\theta}_0}[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}|\ddot{p}_{h,t;\alpha_{h_1},\lambda_{h_2}}|] < \infty$ follows from the fact that:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\left|\mathbb{C}ov_{\boldsymbol{\theta}}[A_t,\eta_t|Y_t,Y_{t-1},\mathcal{S}_t=h]\right|\right]\\
&\quad\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\left(\mathbb{E}_{\boldsymbol{\theta}}[A_t^2|Y_t,Y_{t-1},\mathcal{S}_t=h]\mathbb{E}_{\boldsymbol{\theta}}[\eta_t^2|Y_t,Y_{t-1},\mathcal{S}_t=h]\right)^{1/2}\right]\\
&\quad\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_{t-1}\left(Y_t^2+Y_{t-1}^2+2Y_tY_{t-1}\right)^{1/2}\right]\\
&\quad\leq 2\mathbb{E}_{\boldsymbol{\theta}_0}\left[Y_t^2\right]^2(1+\mathbb{E}_{\boldsymbol{\theta}_0}[Y_t^2]) < \infty,
\end{aligned}$$

by Theorem 2. Finally, that $\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}|\ddot{p}_{h,t;\lambda_{h_2},\lambda_{h_2}}|\right] < \infty$ follows by (A.7). This concludes the proof. □

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

LEMMA 4. *Under the assumptions of Theorem 3,*

$$\left|\left|\nabla_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)\right|\right| < 1+cY_{t-1}, \tag{A.8}$$

$$\left|\left|\nabla^2_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)\right|\right| < 1+cY^2_{t-1}, \tag{A.9}$$

a.s. for a finite constant $c>0$ for all $\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}$.

Proof. Let $\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}$, and let θ_m be its generic element. We define

$$\dot{v}_{h,t;\theta_m} = \frac{\partial P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)}{\partial\theta_m}, \qquad \ddot{v}_{h,t;\theta_m,\theta_n} = \frac{\partial^2 P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)}{\partial\theta_m\partial\theta_n},$$

for $h\in\mathbb{H}$, where $(\mathcal{S}_t=h)\equiv(S^{\alpha}_t=h_1, Z_t=h_2, S^{\eta}_t=h_3)$. Using results from Proposition 1 of Freeland and McCabe (2004a) and with the convention that $P_{\boldsymbol{\theta}}(-a|Y_{t-1},\mathcal{S}_t)=P_{\boldsymbol{\theta}}(Y_t|-b,\mathcal{S}_t)=P_{\boldsymbol{\theta}}(-a|-b,\mathcal{S}_t)=0$ for all $a,b>0$, we obtain

$$\dot{v}_{h,t;\theta_m} = \begin{cases} \frac{Y_{t-1}}{1-\alpha_{h_1}}\left[P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1}-1,\mathcal{S}_t=h)-P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)\right], & \text{if } \theta_m=\alpha_{h_1},\\ P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1},\mathcal{S}_t=h)-P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h), & \text{if } \theta_m=\lambda_{h_2},\\ 0, & \text{otherwise,}\end{cases} \tag{A.10}$$

and if $\theta_m=\theta_n=\alpha_{h_1}$, then

$$\begin{aligned}\ddot{v}_{h,t;\alpha_{h_1},\alpha_{h_1}} = \frac{Y_{t-1}}{(1-\alpha_{h_1})^2}\Bigg[&2P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1}-1,\mathcal{S}_t=h)\\ &+(Y_{t-1}-1)P_{\boldsymbol{\theta}}(Y_t-2|Y_{t-1}-2,\mathcal{S}_t=h)\\ &-Y_{t-1}\frac{P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1}-1,\mathcal{S}_t=h)^2}{P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)}-1\Bigg],\end{aligned}$$

if $\theta_m=\alpha_{h_1}$ and $\theta_n=\lambda_{h_2}$, then

$$\begin{aligned}\ddot{v}_{h,t;\alpha_{h_1},\lambda_{h_2}} &= \frac{Y_{t-1}}{1-\alpha_{h_1}}\\ &\times\left[P_{\boldsymbol{\theta}}(Y_t-2|Y_{t-1}-1,\mathcal{S}_t=h)-P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1},\mathcal{S}_t=h)\frac{P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1}-1,\mathcal{S}_t=h)}{P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)}\right],\end{aligned}$$

if $\theta_m=\lambda_{h_2}$ and $\theta_n=\lambda_{h_2}$, then

$$\ddot{v}_{h,t;\lambda_{h_2},\lambda_{h_2}} = P_{\boldsymbol{\theta}}(Y_t-2|Y_{t-1},\mathcal{S}_t=h)-\frac{P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1},\mathcal{S}_t=h)^2}{P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)},$$

and $\ddot{v}_{h,t;\theta_m,\theta_n}=0$ otherwise. Using Proposition 2 of Freeland and McCabe (2004a) it follows that

$$\begin{aligned}\frac{P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1},\mathcal{S}_t=h)}{P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)} &= \frac{1}{\lambda_{h_2}}\mathbb{E}_{\boldsymbol{\theta}}\left[\eta_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right],\\ \frac{P_{\boldsymbol{\theta}}(Y_t-1|Y_{t-1}-1,\mathcal{S}_t=h)}{P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)} &= \frac{1}{\alpha_{h_1}Y_{t-1}}\mathbb{E}_{\boldsymbol{\theta}}\left[A_t|Y_t,Y_{t-1},\mathcal{S}_t=h\right].\end{aligned}$$

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

Substituting these expressions into $\dot{v}_{h,t;\theta_m}$ and $\ddot{v}_{h,t;\theta_m,\theta_n}$, rearranging the terms, and applying the triangular inequality completes the proof. □

Consider the following alternative representation of the model. Let $\{\tilde{\mathcal{S}}_t\} = \{(S_t^{\alpha}, S_t^{\eta})\}$ be a Markov chain constructed by combining S_t^{α} and S_t^{η}. As in the proof of Lemma 2, it can be shown that $\{\tilde{\mathcal{S}}_t\}$ is a first-order homogeneous ergodic Markov chain with state space $\tilde{\mathbb{H}} = [1,\ldots,\tilde{H}]$, where $\tilde{H} = JK$. Let h also be a generic state of $\{\tilde{\mathcal{S}}_t\}$ and let h_1 and h_2 be the associated realizations of $\{(S_t^{\alpha}, S_t^{\eta})\}$, such that $P(\tilde{\mathcal{S}}_t = h) = P(S_t^{\alpha} = h_1, S_t^{\eta} = h_2)$, for $h \in \tilde{\mathbb{H}}$. By the $\tilde{\mathcal{S}}_t$-representation, it follows that $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = \tilde{h})$ for all $y_t, y_{t-1} \in \mathbb{N}$ is a mixture of K Poisson Binomial distributions with mixing weights given by $\omega_{k,\tilde{h}_2}$, for $k = 1,\ldots,K$, i.e.,

$$
\begin{aligned}
&P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = h)\\
&\quad = \sum_{k=1}^{K} \omega_{k,h_2} \sum_{q=0}^{y_t \wedge y_{t-1}} e^{-\lambda_{h_2}} \frac{\lambda_{h_2}^q}{q!} \binom{y_{t-1}}{y_t - q} \alpha_{h_1}^{y_t - q} (1-\alpha_{h_1})^{y_{t-1}-y_t+q}. \qquad \textbf{(A.11)}
\end{aligned}
$$

Let us represent $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \boldsymbol{\psi}')'$ with $\boldsymbol{\psi} = (\boldsymbol{\psi}_h', h = 1,\ldots,\tilde{H})'$, where $\boldsymbol{\gamma}$ includes all the elements of $\tilde{\boldsymbol{\Gamma}}$ (the transition probability matrix associated with $\{\tilde{\mathcal{S}}_t\}$) but those reported in the last column, and $\boldsymbol{\psi}_h = (\boldsymbol{\omega}_{h_2}', \alpha_{h_1}, \lambda_{h_2})'$.[10] Let $f_{h,t|t-1}(\boldsymbol{\psi}_h) = P_{\boldsymbol{\psi}_h}(Y_t|Y_{t-1}, \tilde{\mathcal{S}}_t = h)$, and $\tilde{v}_{h,t,m}(\boldsymbol{\psi}_h) = \frac{\partial f_{h,t|t-1}(\boldsymbol{\psi}_h)}{\partial [\boldsymbol{\psi}_h]_m}$.

LEMMA 5. *Under the assumptions of Theorem 3,*

$\{f_{h,t|t-1}(\boldsymbol{\psi}_{h,0}), \tilde{v}_{h,t,m}(\boldsymbol{\psi}_{h,0}), h = 1,\ldots,\tilde{H}, m = 1,\ldots,p_h\}$

are linearly independent for all $(Y_t, Y_{t-1}) = (y_t, y_{t-1}') \in \mathbb{N}^2$.

Proof. First, note that Assumption 2 implies that $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = h, Z_t = l_1)$ and $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = h, Z_t = l_2)$ for $l_1 \neq l_2$ are linearly independent for all $(y_t, y_{t-1}) \in \mathbb{N}^2$, such that mixtures $f_{h,t|t-1}(\boldsymbol{\psi}_h) = P_{\boldsymbol{\psi}_h}(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = h)$ for $h \in \tilde{\mathbb{H}}$ are linearly independent by Theorem 2.3 of Chandra (1977). Lemma 5 then follows by noting that expressions for $\tilde{v}_{h,t,m}(\boldsymbol{\psi}_h)$ are of the kind of those reported in (A.10), and involve the terms $\sum_{k=1}^{K} \omega_{k,h_2} P(Y_t = y_t - 1|Y_{t-1} = y_{t-1} - 1, \tilde{S}_t = h, Z_t = k)$ when $[\boldsymbol{\psi}_h]_m = \alpha_h$ and $[\boldsymbol{\psi}_h]_m = \lambda_h$ and $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{S}_t = h, Z_t = k)$ when $[\boldsymbol{\psi}_h]_m = \omega_{h_2,k}$. Linear independence then follows by distinctness of the mixture components (Assumption 2) and the condition $\omega_{h,l} > 0$ (Assumption 1.c). □

Let $f_{t|t-1}(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(Y_t|\mathbf{Y}_{-\infty:t-1})$ and note that:

$$
f_{t|t-1}(\boldsymbol{\theta}) = \sum_{h=1}^{\tilde{H}} \pi_{h,t|t-1}(\boldsymbol{\theta}) f_{h,t|t-1}(\boldsymbol{\psi}_h), \qquad \textbf{(A.12)}
$$

[10] The elements in the last column of $\tilde{\boldsymbol{\Gamma}}$ are excluded since they are computed by subtracting the sum of the other elements from one, i.e., $\tilde{\gamma}_{h\tilde{H}} = 1 - \sum_{i=1}^{\tilde{H}-1} \tilde{\gamma}_{hi}$ for all $h = 1,\ldots,\tilde{H}$.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

where $\pi_{h,t|t-1}(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(\tilde{\mathcal{S}}_t = h|\mathbf{Y}_{-\infty:t-1})$. Let now $\boldsymbol{\pi}_{t|t-1}(\boldsymbol{\theta}) = (\pi_{h,t|t-1}(\boldsymbol{\theta}), h = 1,\dots,\tilde{H})'$, then

$$\boldsymbol{\pi}_{t+1|t}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}^{\tilde{\mathcal{S}}}(\boldsymbol{\theta})'\tilde{\mathbf{P}}_t(\boldsymbol{\theta})\boldsymbol{\pi}_{t|t-1}(\boldsymbol{\theta}), \tag{A.13}$$

where $\tilde{\mathbf{P}}_t(\boldsymbol{\theta})$ is a diagonal matrix with generic element $f_{h,t|t-1}(\boldsymbol{\psi}_h)/f_{t|t-1}(\boldsymbol{\theta})$. Exploiting the fact that $\pi_{\tilde{H},t|t-1}(\boldsymbol{\theta}) = 1 - \sum_{h=1}^{\tilde{H}-1}\pi_{h,t|t-1}(\boldsymbol{\theta})$, starting from (A.13), we obtain the following representation for $\pi_{j,t|t-1}(\boldsymbol{\theta})$:

$$\pi_{j,t+1|t}(\boldsymbol{\theta}) = \gamma_{J,j} + \sum_{i=1}^{\tilde{H}-1}\frac{(\gamma_{i,j}-\gamma_{J,j})f_{i,t|t-1}(\boldsymbol{\theta})}{f_{t|t-1}(\boldsymbol{\theta})}\pi_{i,t|t-1}(\boldsymbol{\theta}), \tag{A.14}$$

for $j = 1,\dots,\tilde{H}-1$.

LEMMA 6. *Under the assumption of Theorem 3, $\mathcal{I}(\boldsymbol{\theta}_0)$ is positive definite.*

Proof. First, recall that

$$\mathcal{I}(\boldsymbol{\theta}_0) = \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\log f_{t|t-1}(Y_t;\boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}'}\log f_{t|t-1}(Y_t;\boldsymbol{\theta}_0)\right].$$

Hence, $\mathcal{I}(\boldsymbol{\theta}_0)$ is clearly positive semi-definite. To show that $\mathcal{I}(\boldsymbol{\theta}_0)$ is also positive definite, it is enough to show that $\mathbf{x}'\nabla_{\boldsymbol{\theta}}\log f_{t|t-1}(Y_t;\boldsymbol{\theta}_0) = 0$ a.s. implies that $\mathbf{x} = \mathbf{0}$. Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \boldsymbol{\psi}')'$, where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^{\alpha\prime}, \boldsymbol{\gamma}^{\eta\prime})'$, $\boldsymbol{\gamma}^{\alpha} = (\boldsymbol{\gamma}_1^{\alpha\prime},\dots,\boldsymbol{\gamma}_{J-1}^{\alpha\,\prime})'$, $\boldsymbol{\gamma}_i^{\alpha} = (\gamma_{j,i}^{\alpha}, j = 1,\dots,J)'$, $\boldsymbol{\gamma}^{\eta} = (\boldsymbol{\gamma}_1^{\eta\prime},\dots,\boldsymbol{\gamma}_{L-1}^{\eta\,\prime})'$, $\boldsymbol{\gamma}_i^{\eta} = (\gamma_{j,i}^{\eta}, j = 1,\dots,L)'$, and $\boldsymbol{\psi} = (\boldsymbol{\psi}_h', h = 1,\dots,\tilde{H})'$, and $\boldsymbol{\psi}_h$ is defined as below (A.11). Similarly, let $\mathbf{x} = (\mathbf{x}^{\gamma\prime}, \mathbf{x}^{\psi\prime})'$, where the elements of $\mathbf{x}^{\gamma}$ are of the kind $x_{i,j}^{\gamma^{\alpha}}$, $i = 1,\dots,J$ and $j = 1,\dots,J-1$, and $x_{i,j}^{\gamma^{\eta}}$, $i = 1,\dots,L$ and $j = 1,\dots,L-1$, and those of $\mathbf{x}^{\psi}$ are x_j^{α}, $j = 1,\dots,J$, x_k^{λ}, $k = 1,\dots,K$, $x_{l,k}^{\omega}$ for $l = 1,\dots,L$ and $k = 1,\dots,K$. From (A.12), we have that

$$\begin{aligned} f_{t|t-1}(Y_t,\boldsymbol{\theta}_0)\mathbf{x}'\nabla_{\boldsymbol{\theta}}\log f_{t|t-1}(Y_t;\boldsymbol{\theta}_0) &= \sum_{j=1}^{\tilde{H}}\mathbf{x}'\nabla_{\boldsymbol{\theta}}\pi_{j,t|t-1}(\boldsymbol{\theta}_0)f_{j,t|t-1}(\boldsymbol{\psi}_{j,0}) \\ &\quad + \sum_{j=1}^{\tilde{H}}\sum_{m=1}^{p}\pi_{j,t|t-1}(\boldsymbol{\theta}_0)x_m^{\psi_j}\tilde{v}_{j,t,m}(\boldsymbol{\psi}_{j,0}), \end{aligned}$$

where p is the dimension of $\boldsymbol{\psi}_h$ (which is the same for all $h = 1\dots,\tilde{H}$). Since $\{f_{j,t|t-1}(\boldsymbol{\psi}_{j,0}), v_{j,t,m}(\boldsymbol{\psi}_{j,0}), j = 1,\dots,\tilde{H}, m = 1,\dots,p\}$ are linearly independent with positive probability by Lemma 5, then $\mathbf{x}'\nabla_{\boldsymbol{\theta}}\log f_{t|t-1}(Y_t;\boldsymbol{\theta}_0) = 0$ a.s. implies that:

1) $\mathbf{x}'\nabla_{\boldsymbol{\theta}}\pi_{j,t|t-1}(\boldsymbol{\theta}_0) = 0$ a.s. for all $j \in \{1,\dots,\tilde{H}\}$,
2) $x_m^{\psi_j}\pi_{j,t|t-1}(\boldsymbol{\theta}_0) = 0$ a.s. for all $j \in \{1,\dots,J\}$ and $m \in \{1,\dots,p\}$,

where the latter implies that $x_m^{\psi_j} = 0$ for all $j \in \{1,\dots,\tilde{H}\}$ and $m \in \{1,\dots,p\}$. Hence, 1) becomes

1′) $\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}\pi_{j,t|t-1}(\boldsymbol{\theta}_0) = 0$ a.s. for all $j \in \{1,\dots,\tilde{H}\}$.

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

Provided that, associated with an index $j \in \{1,\ldots,\tilde{H}\}$, there are two indexes $j_1 \in \{1,\ldots,J\}$ and $j_2 \in \{1,\ldots,L\}$, and that a generic element of $\tilde{\boldsymbol{\Gamma}}$ is given by $\gamma_{i,j} = \gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2}$ for some indexes i_1, i_2, j_1, j_2, from (A.14), we get that

$$
\begin{aligned}
&\nabla_{\boldsymbol{\gamma}}\pi_{j,t|t-1}(\boldsymbol{\theta}_0)\\
&= \nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{J,j_1}\gamma^{\eta}_{L,j_2}) + \sum_{i=1}^{\tilde{H}-1}\left(\frac{(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2} - \gamma^{\alpha}_{J,j_1}\gamma^{\eta}_{L,j_2})f_{i,t|t-1}(\boldsymbol{\psi}_{i,0})}{f_{t|t-1}(Y_t;\boldsymbol{\theta}_0)}\right)\nabla_{\boldsymbol{\gamma}}\pi_{i,t|t-1}(\boldsymbol{\theta}_0)\\
&\quad + \sum_{i=1}^{\tilde{H}-1}\pi_{i,t|t-1}(\boldsymbol{\theta}_0)\left[\frac{\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2} - \gamma^{\alpha}_{J,j_1}\gamma^{\eta}_{L,j_2})f_{i,t|t-1}(\boldsymbol{\psi}_{i,0})}{f_{t|t-1}(Y_t;\boldsymbol{\theta}_0)}\right.\\
&\quad \left. - \frac{(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2} - \gamma^{\alpha}_{J,j_1}\gamma^{\eta}_{L,j_2})f_{i,t|t-1}(\boldsymbol{\psi}_{i,0})}{f_{t|t-1}(Y_t;\boldsymbol{\theta}_0)^2}\sum_{h=1}^{J}\nabla_{\boldsymbol{\gamma}}\pi_{h,t|t-1}(\boldsymbol{\theta}_0)f_{h,t|t-1}(\boldsymbol{\psi}_{h,0})\right], \qquad \textbf{(A.15)}
\end{aligned}
$$

for all $j = 1,\ldots,\tilde{H}-1$. Since $\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}\pi_{j,t|t-1}(\boldsymbol{\theta}_0) = 0$ for all $j = 1\ldots,\tilde{H}$ by 1′), and by premultiplying (A.15) by $f_{t|t-1}(Y_t : \boldsymbol{\theta}_0)\mathbf{x}^{\gamma\prime}$, we obtain

$$
0 = \sum_{i=1}^{\tilde{H}}\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2})f_{i,t|t-1}(\boldsymbol{\psi}_{i,0})\pi_{i,t|t-1}(\boldsymbol{\theta}_0)
$$

for all (j_1,j_2) associated with $j = 1,\ldots,\tilde{H}-1$. Furthermore, linear independence of $f_{1,t|t-1}(\boldsymbol{\psi}_{1,0}),\ldots,f_{\tilde{H},t|t-1}(\boldsymbol{\psi}_{\tilde{H},0})$ leads to the following result:

3) $\mathbf{x}^{\boldsymbol{\gamma}\prime}\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2}) = 0$ for all $j_1 \in \{1,\ldots,J\}$, $j_2 \in \{1,\ldots,L\}$ associated with $j \in \{1,\ldots,\tilde{H}-1\}$ and all $i_1 \in \{1,\ldots,J\}$, $i_2 \in \{1,\ldots,L\}$ associated with $i \in \{1,\ldots,\tilde{H}\}$.

Noting that

$$
\gamma_{i,j} = \begin{cases}
\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2}, & \text{if} \quad j_1 \neq J, j_2 \neq L\\
\left(1 - \sum_{l=1}^{J-1}\gamma^{\alpha}_{i_1,l}\right)\gamma^{\eta}_{i_2,j_2}, & \text{if} \quad j_1 = J, j_2 \neq L\\
\gamma^{\alpha}_{i_1,j_1}\left(1 - \sum_{l=1}^{L-1}\gamma^{\eta}_{i_2,l}\right), & \text{if} \quad j_1 \neq J, j_2 = L\\
\left(1 - \sum_{l=1}^{J-1}\gamma^{\alpha}_{i_1,l}\right)\left(1 - \sum_{l=1}^{L-1}\gamma^{\eta}_{i_2,l}\right), & \text{if} \quad j_1 = J, j_2 = L,
\end{cases}
$$

and

$$
\begin{aligned}
&\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2})\\
&= \begin{cases}
x^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2} + \gamma^{\alpha}_{i_1,j_1}x^{\eta}_{i_2,j_2}, & \text{if } j_1 \neq J, j_2 \neq L\\
-\sum_{l=1}^{J-1}x^{\alpha}_{i_1,l}\gamma^{\eta}_{i_2,j_2} + \left(1 - \sum_{l=1}^{J-1}\gamma^{\alpha}_{i_1,l}\right)x^{\eta}_{i_2,j_2}, & \text{if } j_1 = J, j_2 \neq L\\
x^{\alpha}_{i_1,j_1}\left(1 - \sum_{l=1}^{L-1}\gamma^{\eta}_{i_2,l}\right) - \gamma^{\alpha}_{i_1,j_1}\sum_{l=1}^{L-1}x^{\eta}_{i_2,l}, & \text{if } j_1 \neq J, j_2 = L\\
-\sum_{l=1}^{J-1}x^{\alpha}_{i_1,l}\left(1 - \sum_{l=1}^{L-1}\gamma^{\eta}_{i_2,l}\right) - \left(1 - \sum_{l=1}^{J-1}\gamma^{\alpha}_{i_1,l}\right)\sum_{l=1}^{L-1}x^{\eta}_{i_2,l}, & \text{if } j_1 = J, j_2 = L,
\end{cases}
\end{aligned}
$$

https://doi.org/10.1017/S0266466625100182 Published online by Cambridge University Press

and by $\left(\sum_{j_1=1}^{J-1}\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2})\right)+\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,J}\gamma^{\eta}_{i_2,j_2})=0$, we conclude that $x^{\eta}_{i_2,j_2}=0$ for all $i_2=1,\dots,L$, and $j_2=1,\dots,L-1$. Similarly, by $\left(\sum_{j_2=1}^{L-1}\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,j_2})\right)+\mathbf{x}^{\gamma\prime}\nabla_{\boldsymbol{\gamma}}(\gamma^{\alpha}_{i_1,j_1}\gamma^{\eta}_{i_2,L})=0$, we conclude that $x^{\alpha}_{i_1,j_1}=0$ for all $i_1=1,\dots,J$, and $j_1=1,\dots,J-1$, such that $\mathbf{x}^{\gamma}=\mathbf{0}$. All in all, $\mathbf{x}'\nabla_{\boldsymbol{\theta}}\log f_{t|t-1}(Y_t;\boldsymbol{\theta}_0)=0$ a.s. implies that $\mathbf{x}=\mathbf{0}$. Hence, $\mathcal{I}(\boldsymbol{\theta}_0)$ is also positive definite. □

A.5.2. *Proof of Theorem 3.*

Proof. To prove the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_{T,s_0}$ for any s_0, we apply Theorems 1 and 4 of Douc et al. (2004), respectively. For the strong consistency of the Fisher information matrix estimator, we apply their Theorem 3. For the proof, we illustrate below that their conditions (A1)–(A8) are valid in the context of ML estimation of the HMM-INAR model. We list below conditions (A1)–(A8), adapting them to the notation of the HMM-INAR model:

(A1) (a) $\gamma^{\mathcal{S}}_{ij}>0$
  (b) (i) for all $y,y'\in\mathbb{N}^2$: $(i)\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sum_{h=1}^{H}P_{\boldsymbol{\theta}}(Y_t=y|Y_{t-1}=y',\mathcal{S}_t=h)P_{\boldsymbol{\theta}}(\mathcal{S}_t=h)>0$ and $(ii)\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sum_{h=1}^{H}P_{\boldsymbol{\theta}}(Y_t=y|Y_{t-1}=y',\mathcal{S}_t=h)P(\mathcal{S}_t=h)<\infty$.

(A2) The sequence $\{(Y_t,Y_{t-1},\mathcal{S}_t)\}_{t\in\mathbb{N}}$ at $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ is a Markov chain on $\mathbb{N}^2\times\mathbb{H}$ with transition kernel $\Pi_{\boldsymbol{\theta}}$. For all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, the transition kernel $\Pi_{\boldsymbol{\theta}}$ is positive Harris recurrent and aperiodic with invariant distribution $\pi_{\boldsymbol{\theta}}$.

(A3) (a) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{(y,y')\in\mathbb{N}^2}\sup_{h\in\mathbb{H}}P_{\boldsymbol{\theta}}(Y_1=y|Y_0=y',S_1=h)<\infty$
  (b) $\mathbb{E}_{\boldsymbol{\theta}_0}\left(\left|\log\left(\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sum_{h=1}^{H}P_{\boldsymbol{\theta}}(Y_1|Y_0,S_1=h)P_{\boldsymbol{\theta}}(S_1=h)\right)\right|\right)<\infty$.

(A4) $(i)\boldsymbol{\theta}\to\boldsymbol{\Gamma}$ is continuous on $\boldsymbol{\Theta}$ and $(ii)\boldsymbol{\theta}\to P_{\boldsymbol{\theta}}(Y_1=y|Y_0=y',\mathcal{S}_1=h)$ is continuous on $\boldsymbol{\Theta}$ for all $(y,y',h)\in\mathbb{N}^2\times\mathbb{H}$.

(A5)′ $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, if and only if $\mathbb{E}\left[\log P_{\boldsymbol{\theta}_0}(\mathbf{Y}_{1:p}|\mathbf{Y}_0)\right]=\mathbb{E}\left[\log P_{\boldsymbol{\theta}}(\mathbf{Y}_{1:p}|\mathbf{Y}_0)\right]$ for all $p\geq 1$.

(A6) $(i)\boldsymbol{\theta}\to\boldsymbol{\Gamma}$ is twice continuously differentiable on $\dot{\boldsymbol{\Theta}}$ and $(ii)\boldsymbol{\theta}\to P_{\boldsymbol{\theta}}(Y_1=y|Y_0=y',\mathcal{S}_1=h)$ is twice continuously differentiable on $\dot{\boldsymbol{\Theta}}$ for all $(y,y',h)\in\mathbb{N}^2\times\mathbb{H}$.

(A7) (a) $\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{(i,j)\in\mathbb{H}^2}||\nabla_{\boldsymbol{\theta}}\log\gamma^{\mathcal{S}}_{i,j}||<\infty$ and $\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{(i,j)\in\mathbb{H}^2}||\nabla^2_{\boldsymbol{\theta}}\log\gamma^{\mathcal{S}}_{i,j}||<\infty$.
  (b) $\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}||\nabla_{\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)||^2\right]<\infty$ and $\mathbb{E}_{\boldsymbol{\theta}_0}\left[\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}\sup_{h\in\mathbb{H}}\left|\left|\nabla^2_{\boldsymbol{\theta}}\log P_{\boldsymbol{\theta}}(Y_t|Y_{t-1},\mathcal{S}_t=h)\right|\right|\right]<\infty$.

(A8) (a) There exists a function $f_{y,y'}:\mathbb{H}\to(0,\infty)$ such that $\sup_{\boldsymbol{\theta}\in\dot{\boldsymbol{\Theta}}}P_{\boldsymbol{\theta}}(Y_1=y|Y_0=y',\mathcal{S}_t=h)\leq f_{y,y'}(h)$ for almost all $(y,y')\in\mathbb{N}^2$, with $f_{y,y'}\in L^1$, the space of absolutely integrable functions.
  (b) There exist functions $f^1_{h,y}:\mathbb{N}\to(0,\infty)$ and $f^2_{h,y}:\mathbb{N}\to(0,\infty)$ such that $\left|\left|\nabla_{\boldsymbol{\theta}}P_{\boldsymbol{\theta}}(Y_t=y'|Y_{t-1}=y,\mathcal{S}_t=h)\right|\right|\leq f^1_{h,y}(y')$ and $||\nabla^2_{\boldsymbol{\theta}}P_{\boldsymbol{\theta}}(Y_t=y'|Y_{t-1}=y,\mathcal{S}_t=h)||\leq f^2_{h,y}(y')$, a.s., with $f^1_{h,y}(y')\in L^1$ and $f^2_{h,y}(y')\in L^1$.

Strong consistency of $\widehat{\boldsymbol{\theta}}_{T,s_0}$ for any s_0 follows from Theorem 1 of Douc et al. (2004) if (A1)–(A5) are satisfied. We note that (A1)(a) is satisfied by Lemma 1; (A1)(b)(i) is satisfied due to the fact that $P(\mathcal{S}_t=h)>0$ for all $h\in\mathbb{H}$ (Lemma 1) and $\min_{j=1,\dots,J}\alpha_j<1$, which ensures that for at least one $h\in\mathbb{H}$, we have $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}P_{\boldsymbol{\theta}}(Y_t=y|Y_{t-1}=y',\mathcal{S}_t=h)>0$ for all $(y,y')\in\mathbb{N}^2$; (A1)(b)(ii) is satisfied, since $P_{\boldsymbol{\theta}}(Y_t=y|Y_{t-1}=y',\mathcal{S}_t=h)\leq 1$ and $P(\mathcal{S}_t=h)<1$ for all y,y',h. Condition (A2) is satisfied, if the extended process is geometrically ergodic, which holds by Theorem 1. Condition (A3) is satisfied by the fact that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{(y,y')\in\mathbb{N}^2} \sup_{h\in\mathbb{H}} P_{\boldsymbol{\theta}}(Y_t = y|Y_{t-1} = y', \mathcal{S}_t = h) \leq 1$ and by the fact that $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} P_{\boldsymbol{\theta}}(Y_t|Y_{t-1}, \mathcal{S}_t = h') > 0$ a.s. for some $h' \in \mathbb{H}$. The latter inequality follows by selecting any index h' for which $\alpha_{h'_1} < 1$ (where h'_1 is the index for S_t^{α} selected according to h'). Indeed, it is easily shown that, for such h', $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} P_{\boldsymbol{\theta}}(Y_1|Y_0, \mathcal{S}_1 = h') > 0$ a.s. It then follows that, $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} P_{\boldsymbol{\theta}}(Y_1|Y_0) \geq \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} P_{\boldsymbol{\theta}}(Y_1|Y_0, \mathcal{S}_1 = h')P(\mathcal{S}_1 = h') > 0$ a.s., since $P(\mathcal{S}_1 = h') > 0$ by Assumption 1. Condition (A4) is satisfied by the HMM-INAR model because the conditional probability mass functions are smooth functions of $\boldsymbol{\theta}$. Condition (A5)$'$ is satisfied by Assumptions 1 and 2.

As shown in the proof of Lemma 5, under the $\tilde{\mathcal{S}}_t$ representation of the HMM-INAR model, $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = \tilde{h})$ for all $h = 1, \ldots, \tilde{H}$, are linearly independent. Condition (A5)$'$ then follows by Lemma 5 of Krishnamurthy and Rydén (1998). The only difference with Lemma 5 of Krishnamurthy and Rydén (1998) is that in the HMM-INAR case $Y_t|(Y_{t-1}, \tilde{\mathcal{S}}_t)$ is distributed according to a mixture of K distributions. However, as noted in the proof of Proposition 2 of Gassiat, Cleynen, and Robin (2016), a sufficient condition to ensure that the mixture distributions $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = \tilde{h})$, for $\tilde{h} \in \tilde{\mathbb{H}}$, are linearly independent is that the conditional probability mass functions $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = \tilde{h}, Z_t = l)$ for $l = 1, \ldots, L$ are linearly independent for all $y_{t-1} \in \mathbb{N}$ and that $\boldsymbol{\Omega}$ has rank L with $L \leq K$. Both conditions are satisfied by Assumption 2 and identifiability of $P(Y_t = y_t|Y_{t-1} = y_{t-1}, \tilde{\mathcal{S}}_t = \tilde{h}, Z_t = l)$, which follows from Theorem 2.3 of Chandra (1977), as discussed earlier. We then conclude that (A5)$'$ is satisfied by Lemma 5 of Krishnamurthy and Rydén (1998). So, by Theorem 1 of Douc et al. (2004) $\widehat{\boldsymbol{\theta}}_{T,s_0} \to \boldsymbol{\theta}_0$ a.s. as $T \to \infty$ and for any $s_0 \in \mathbb{H}$.

Asymptotic normality of the MLE follows by verifying conditions (A6)–(A8). We note that conditions (A6)(i) and (A6)(ii) are straightforward to verify in the HMM-INAR case. Conditions (A7)(a) and (A7)(b) are verified by applying Lemma 3. Condition (A8)(a) is satisfied by taking $f_{y,y'}(h) = 1$, while condition (A8)(b) is satisfied by taking $f^1_{h,y}(y') = 1 + cy'$ and $f^2_{h,y}(y') = 1 + c(y')^2$ for some finite $c > 0$ by an application of Lemma 4. Note that absolute integrability of $f^1_{h,y}(y')$ and $f^2_{h,y}(y')$ follows from the weak stationarity of $\{Y_t\}$, which is ensured by $\min_{j=1,\ldots,J} \alpha_j < 1$. So, by an application of Theorem 4 of Douc et al. (2004), we conclude that $\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,s_0} - \boldsymbol{\theta}_0) \to \mathcal{N}(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1})$ for $T \to \infty$, for any $s_0 \in \mathbb{H}$.

Strong consistency of the Fisher information matrix estimator follows by Theorem 3 of Douc et al. (2004) under their assumptions (A1)–(A3) and (A6)–(A8), which are satisfied by the HMM-INAR model. Positive definiteness of $\mathbf{I}(\boldsymbol{\theta}_0)$ follows by Lemma 6. □

Supplementary Material

Catania, L., Rossi, E., and Santucci de Magistris, P. (2025): Supplement to "Switching Regime Integer Autoregressions," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466625100182.

REFERENCES

Aknouche, A., & Franq, C. (2021). Count and duration time series with equal conditional stochastic and mean orders. *Econometric Theory*, 37(2), 248–280.

Aknouche, A., & Scotto, M. G. (2024). A multiplicative thinning-based integer-valued GARCH model. *Journal of Time Series Analysis*, 45(1), 4–26.

Al-Osh, M., & Alzaid, A. A. (1987). First-order integer-valued autoregressive INAR(1) process. *Journal of Time Series Analysis*, 8(3), 261–275.

Al-Osh, M. A., & Aly, E.-E. A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics-Theory and Methods*, 21(9), 2483–2492.

Alerini, J., Olteanu, M., & Ridgway, J. (2017). Markov and the Duchy of Savoy: Segmenting a century with regime-switching models. *Journal de la Société Française de Statistique*, 158, 89–116.

Alzaid, A. A., & Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27(2), 314–324.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56.

Andersen, T. G., Thyrsgaard, M., & Todorov, V. (2019). Time-varying periodicity in intraday volatility. *Journal of the American Statistical Association*, 114(2), 1695–1707.

Armillotta, M., Luati, L., & Lupparelli, M. (2022). Observation-driven models for discrete-valued time series. *Electronic Journal of Statistics*, 16(1), 1393–1433.

Barardehi, Y. H., Bernhardt, D., & Davies, R. J. (2019). Trade-time measures of liquidity. *The Review of Financial Studies*, 32(1), 126–179.

Bartolucci, F., & Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104(486), 816–831.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2012). *Latent Markov models for longitudinal data*. Chapman and Hall/CRC Press.

Bourguignon, M., & Weiß, C. H. (2017). An INAR(1) process for modeling count time series with equidispersion, underdispersion and over-dispersion. *TEST*, 26(4), 847–868.

Brännäs, K., & Hellström, J. (2001). Generalized integer-valued autoregression. *Econometric Reviews*, 20(4), 425–443.

Brockwell, A. E. (2007). Universal residuals: A multivariate transformation. *Statistics & Probability Letters*, 77(14), 1473–1478.

Bu, R., & McCabe, B. (2008). Model selection, estimation and forecasting in INAR(p) models: A likelihood-based Markov chain approach. *International Journal of Forecasting*, 24(1), 151–162.

Bu, R., McCabe, B., & Hadri, K. (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis*, 29(6), 973–994.

Catania, L., & Di Mari, R. (2021). Hierarchical Markov-switching models for multivariate integer-valued time-series. *Journal of Econometrics*, 221(1), 118–137.

Catania, L., Di Mari, R., & Santucci de Magistris, P. (2022). Dynamic discrete mixtures for high-frequency prices. *Journal of Business & Economic Statistics*, 40(2), 559–577.

Chan, K. (1990). Deterministic stability, stochastic stability, and ergodicity. In *Non-linear time series analysis: A dynamical system approach*. Oxford University Press.

Chandra, S. (1977). On the mixtures of probability distributions. *Scandinavian Journal of Statistics*, 4(3), 105–112.

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge University Press.

Csiszár, I., & Shields, P. C. (2000). The consistency of the BIC Markov order estimator. *The Annals of Statistics*, 28(6), 1601–1619.

Davis, R. A., Fokianos, K., Holan, S. H., Joe, H., Livsey, J., Lund, R., Pipiras, V., & Ravishanker, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association*, 116(535), 1533–1547.

Davis, R. A., Holan, S. H., Lund, R., & Ravishanker, N. (2016). *Handbook of discrete-valued time series*. CRC Press.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4), 863–883.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144.

Douc, R., Moulines, E., & Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *The Annals of Statistics*, 32(5), 2254–2304.

Du, J.-G., & Li, Y. (1991). The integer-valued autoregressive INAR(p) model. *Journal of Time Series Analysis*, 12(2), 129–142.

Easley, D., & O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19(1), 69–90.

Fokianos, K. (2012). Count time series models. In T. Subba Rao, S. Subba Rao & C. R. Rao (Eds.), *Handbook of statistics* (Vol. 30, pp. 315–347). Elsevier.

Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488), 1430–1439.

Fokianos, K., & Neumann, M. H. (2013). A goodness-of-fit test for Poisson count processes. *Electronic Journal of Statistics*, 7, 793–819.

Fokianos, K., Støve, B., Tjøstheim, D., & Doukhan, P. (2020). Multivariate count autoregression. *Bernoulli*, 26(1), 471–499.

Freeland, R. K., & McCabe, B. P. (2004a). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, 25(5), 701–722.

Freeland, R. K., & McCabe, B. P. (2004b). Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3), 427–434.

Gassiat, É., Cleynen, A., & Robin, S. (2016). Inference in finite state space non parametric hidden Markov models and applications. *Statistics and Computing*, 26(1), 61–71.

Gassiat, É., & Boucheron, S. (2003). Optimal error exponents in hidden Markov models order estimation. *IEEE Transactions on Information Theory*, 49(4), 964–980.

Gouriéroux, C., & Jasiak, J. (2004). Heterogeneous INAR(1) model with application to car insurance. *Insurance: Mathematics and Economics*, 34(2), 177–192.

Gouriéroux, C., & Lu, Y. (2019). Negative binomial autoregressive process with stochastic intensity. *Journal of Time Series Analysis*, 40(2), 225–247.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2), 357–384.

Harris, D., & McCabe, B. (2019). Semiparametric independence testing for time series of counts and the role of the support. *Econometric Theory*, 35(6), 1111–1145.

Jung, R. C., & Tremayne, A. R. (2006a). Binomial thinning models for integer time series. *Statistical Modelling*, 6(2), 81–96.

Jung, R. C., & Tremayne, A. R. (2006b). Coherent forecasting in integer time series models. *International Journal of Forecasting*, 22(2), 223–238.

Kang, Y., Zhu, F., Wang, D., & Wang, S. (2024). A zero-modified geometric INAR(1) model for analyzing count time series with multiple features. *Canadian Journal of Statistics*, 52(3), 873–899.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1), 109–126.

Krishnamurthy, V., & Rydén, T. (1998). Consistent estimation of linear and non-linear autoregressive models with Markov regime. *Journal of Time Series Analysis*, 19(3), 291–307.

Levinson, S. E., Rabiner, L. R., & Sondhi, M. M. (1983). An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Technical Journal*, 62(4), 1035–1074.

Livsey, J., Lund, R., Kechagias, S., & Pipiras, V. (2018). Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics*, 12(1), 408–431.

Lu, F., & Wang, D. (2022). First-order integer-valued autoregressive process with Markov-switching coefficients. *Communications in Statistics-Theory and Methods*, 51, 4313–4329.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21(4), 645–650.

Meng, X.-L., & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2), 267–278.

Nastić, A. S., Laketa, P. N., & Ristić, M. M. (2016). Random environment integer-valued autoregressive process. *Journal of Time Series Analysis*, 37(2), 267–287.

Pavlopoulos, H., & Karlis, D. (2008). INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics*, 19(4), 369–393.

Pedeli, X., Davison, A. C., & Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, 110(511), 1229–1238.

Qian, L., & Zhu, F. (2025). A flexible model for time series of counts with overdispersion or underdispersion, zero-inflation and heavy-tailedness. *Communications in Mathematics and Statistics*, 13(2), 431–454.

Quoreshi, A. M. M. S. (2014). A long-memory integer-valued time series model, INARFIMA, for financial application. *Quantitative Finance*, 14(12), 2225–2235.

Roick, T., Karlis, D., & McNicholas, P. D. (2021). Clustering discrete-valued time series. *Advances in Data Analysis and Classification*, 15(1), 209–229.

Ranaldo, A., & Santucci de Magistris, P. (2022). Liquidity in the global currency market. *Journal of Financial Economics*, 146(3), 859–883.

Rossi, E., & Fantazzini, D. (2014). Long memory and periodicity in intraday volatility. *Journal of Financial Econometrics*, 13(4), 922–961.

Roy, A., & Karmakar, S. (2021). Time-varying auto-regressive models for count time-series. *Electronic Journal of Statistics*, 15(1), 2905–2938.

Rudholm, N. (2001). Entry and the number of firms in the Swedish pharmaceuticals market. *Review of Industrial Organization*, 19, 351–364.

Rydberg, T. H., & Shephard, N. (2000). A modelling framework for the prices and times of trades made on the New York stock exchange. In W. J. Fitzgerald, R. L. Smith, A. T. Walden & P. Young (Eds.), *Nonlinear and nonstationary signal processing*. Cambridge University Press Cambridge.

Rydén, T., Teräsvirta, T., & Aasbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13(3), 217–244.

Schweer, S., & Weiß, C. H. (2014). Compound Poisson INAR(1) processes: Stochastic properties and testing for over-dispersion. *Computational Statistics & Data Analysis*, 77, 267–284.

Steutel, F., & Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7(5), 893–899.

Tang, M., & Wang, Y. (2014). Asymptotic behavior of random coefficient INAR model under random environment defined by difference equation. *Advances in Difference Equations*, 99(1), 1–9.

Tauchen, G. E., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, 51, 485–505.

Thyregod, P., Carstensen, J., Madsen, H., & Arnbjerg-Nielsen, K. (1999). Integer valued autoregressive models for tipping bucket rainfall measurements. *Environmetrics*, 10(4), 395–411.

Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Oxford University Press.

Tweedie, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes and their Applications*, 3(4), 385–403.

Vermunt, J. K., Langeheine, R., & Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24(2), 179–207.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts: A survey. *AStA Advances in Statistical Analysis*, 92(3), 319.

Weiß, C. H. (2015). A Poisson INAR(1) model with serially dependent innovations. *Metrika*, 78(7), 829–851.

Weiß, C. H., Zhu, F., & Hoshiyar, A. (2022). Softplus INGARCH models. *Statistica Sinica*, 32(2), 1099–1120.

Weiß, C. H., & Zhu, F. (2024). Conditional-mean multiplicative operator models for count time series. *Computational Statistics and Data Analysis*, 191, 107885.

Wong, C. S., & Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 62(1), 95–115.

Zheng, H., Basawa, I. V., & Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, 137(1), 212–229.

Zucchini, W., MacDonald, I. L., & Langrock, R. (2017). *Hidden Markov models for time series: An introduction using R*. CRC Press.