FLEXIBLE BIVARIATE INGARCH PROCESS WITH A BROAD RANGE OF CONTEMPORANEOUS CORRELATION

LUIZA S. C. PIANCASTELLI\textsuperscript{a} WAGNER BARRETO-SOUZA\textsuperscript{ab} AND HERNANDO OMBAO\textsuperscript{b} \textsuperscript{c}

\textsuperscript{a}School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
\textsuperscript{b}Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

We propose a novel flexible bivariate conditional Poisson (BCP) INteger-valued Generalized AutoRegressive Conditional Heteroscedastic (INGARCH) model for correlated count time series data. Our proposed BCP-INGARCH model is mathematically tractable and has as the main advantage over existing bivariate INGARCH models its ability to capture a broad range (both negative and positive) of contemporaneous cross-correlation, which is a non-trivial advancement. Properties of stationarity and ergodicity for the BCP-INGARCH process are developed. Estimation of the parameters is performed through conditional maximum likelihood (CML), and the finite-sample behavior of the estimators is investigated through simulation studies. Asymptotic properties of the CML estimators are derived. Hypothesis testing methods for the presence of contemporaneous correlation between the time series are presented and evaluated. A Granger causality test is also addressed. We apply our methodology to monthly counts of hepatitis cases in two nearby Brazilian cities, which are highly cross-correlated. The data analysis demonstrates the importance of considering a bivariate model allowing for a wide range of contemporaneous correlation in real-life applications.

Received 30 November 2020; Revised 5 June 2022; Accepted 13 July 2022

Keywords: asymptotics; cross-correlation; ergodicity; multivariate count time series; stability theory.

MOS subject classification: 2020; 62M10; 62F10; 62F12.

1. INTRODUCTION

Count time series data are collected and studied in many fields including public health, cybersecurity, criminology, and medical science. Univariate count time series models based on hidden Markov chains, INAR (INteger-valued AutoRegressive), and INGARCH (INteger-valued Generalized AutoRegressive Conditional Heteroscedastic) approaches, among others, were introduced and explored in several papers. For a general account on univariate count models, please see Kedem and Fokianos (2002), Fokianos (2011), Davis et al. (2015), and Weiß (2018).

Our goal here is to develop a model for multivariate count time series that can capture cross-dependencies between the different components. Recent works have emerged on this topic, but several limitations have yet to be addressed. As noted by Cameron and Trivedi (1998), Jung et al. (2011), and Karlis (2016), models for multivariate count time series are rather sparse mainly due to the analytical and computational challenges.

One approach to modeling multivariate count time series uses the idea of thinning operators (Steutel and van Harn, 1979) that were introduced to statistics by Latour (1997), Pedeli and Karlis (2011), Karlis and Pedeli (2013), Pedeli and Karlis (2013a), Pedeli and Karlis (2013b), and Scotto et al. (2014). A limitation of these thinning-based models is that in general they are not able to model negative cross-correlation and the associated likelihood function is cumbersome. Recent bivariate thinning-based count processes with autocorrelation structure are due to Livesey et al. (2018) and Darolles et al. (2019), but the same criticism on the cumbersome likelihood approach applies for these models. Another approach to analyzing multivariate count time series is the latent factor-based models by Jørgensen et al. (1999), Jung et al. (2011), and Wang and Wang (2018), just to name a few.

\textsuperscript{a}Correspondence to: Wagner Barreto-Souza, Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. Email: wagner.barretosouza@kaust.edu.sa

\textsuperscript{b}Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

\textsuperscript{c}Email: wagner.barretosouza@kaust.edu.sa

© 2022 John Wiley & Sons Ltd
The focus of the present article is the popular INGARCH approach, initially proposed by Heinen (2003), Ferland et al. (2006), and Fokianos et al. (2009), where a Poisson model is assumed for the conditional distribution of the counts given the past. Alternatives to the Poisson assumption motivated several works in the literature, as the negative binomial (Zhu, 2010; Christou and Fokianos, 2014), infinitely divisible (Gonçalves et al., 2015), exponential family (Davis and Liu, 2016), and mixed Poisson (Christou and Fokianos, 2015; Silva and Barreto-Souza, 2019) INGARCH models, to name a few. Further, the linearity assumption was relaxed in Fokianos (2011) and Fokianos and Tjøstheim (2012), which introduced the log-linear and nonlinear INGARCH models.

Although there is abundant work on univariate INGARCH models, multivariate extensions are still scarce in the field. The first work dealing with this topic is due to Liu (2012), where a bivariate Poisson INGARCH model was proposed. This bivariate process was also studied by Lee et al. (2018), where asymptotic properties of estimators and a parameter-change test were addressed. Recently, Cui and Zhu (2018) introduced another Poisson bivariate count time series model allowing for both negative and positive contemporaneous correlation (also known as cross-correlation). A drawback of the model by Liu (2012) and Lee et al. (2018) is that negative cross-correlation is not allowed in contrast with the model by Cui and Zhu (2018). On the other hand, in the former, the supported range of cross-correlation is very limited. In these papers, we have two major problems regarding that range: (i) natural restriction of the parameter space due to the baseline bivariate discrete distribution; (ii) the parameter space (related to the cross-correlation parameter) of the baseline count distribution depends on the marginal means. These points imply that there is a severe limitation in the correlation between the count time series, which these models can capture, limiting their practical applicability. Our proposed model deals with both issues as the parameter space of its correlation parameter is \( \mathbb{R} \)-valued (so it does not depend on the means) and our baseline distribution allows for a broad range of correlation. In Section 2.2, we provide a detailed discussion on these restrictions and how our model overcomes them.

We propose a novel bivariate INteger-valued Generalized AutoRegressive Conditional Heteroscedastic (INGARCH) model for the statistical analysis of correlated count time series data. More specifically, we introduce and study a new flexible bivariate conditional Poisson (BCP) INGARCH model, which is mathematically tractable and whose main advantage over the existing bivariate count time series models, by Liu (2012), Lee et al. (2018), and Cui and Zhu (2018), is its ability to capture a broad range of both negative and positive contemporaneous correlation. Along with the article, we argue that such a broad range is very important to model properly high correlated count time series. Besides, we derive the theoretical properties of the BCP-INGARCH model such as conditions to ensure stationarity and ergodicity, and asymptotics on the conditional maximum likelihood estimators, as well as a full discussion on the statistical modeling including point estimation, procedures to obtain standard errors, hypothesis testing on the presence of cross-correlation, and forecasting. It is worth mentioning two other related works about multivariate INGARCH models by Fokianos et al. (2020) and Cui et al. (2020). Although these models can also be flexible regarding the contemporaneous correlation, this is difficult to assess because explicit forms for the correlation structure are not provided. This is further discussed in Section 2.2.

The remainder of the article is organized in the following way. In Section 2, we define our proposed bivariate conditional Poisson INGARCH model, establish the properties of stationarity and ergodicity of the process, and compare it with existing bivariate INGARCH models. Section 3 is devoted to the estimation of the parameters via the conditional maximum likelihood method. Furthermore, we establish conditions to obtain consistency and asymptotic normality of the proposed estimators. Simulation studies are conducted to assess the finite-sample performance of the proposed estimators in Section 4. We also compare methods to obtain standard errors of the parameters. Hypothesis testing and simulated results involving the cross-correlation parameter are also addressed. A Granger causality test is addressed in Section 5. A full data analysis of the bivariate counts of hepatitis, which extracts the contemporaneous correlation in two nearby Brazilian cities, is presented in Section 6, which demonstrates the importance of considering a bivariate model allowing for a wide range of contemporaneous correlation in real-life applications. Concluding remarks and future research are discussed in Section 7. This article has a Supplementary Material containing the proofs of the theorems and additional numerical results as well. We make available in the GitHub page https://github.com/luizapiancastelli/rbcpingarch the codes for fitting our proposed bivariate INGARCH model.
2. BIVARIATE CONDITIONAL POISSON INGARCH PROCESS

We begin by presenting the BCP distribution introduced by Berkhout and Plug (2004). We say that a random vector \((Z_1, Z_2)\) follows a bivariate conditional Poisson distribution with parameters \(\lambda_1, \lambda_2 > 0\) and \(\phi \in \mathbb{R}\) if it satisfies the stochastic representation: \(Z_1 \sim \text{Poisson}(\lambda_1)\) and \(Z_2|Z_1 = z_1 \sim \text{Poisson}(\mu_2 e^{\phi z_1})\), where \(\mu_2 \equiv \lambda_2 \exp\{-\lambda_1 (e^\phi - 1)\}\). We denote \((Z_1, Z_2) \sim \text{BCP}(\lambda_1, \lambda_2, \phi)\). Note that the marginal of \(Z_2\) is not Poisson but mixed Poisson distributed. The mean and variance of \(Z_2\) are given by \(E(Z_2) = E(E(Z_2|Z_1)) = \mu_2 E(e^{\phi Z_1}) = \lambda_2\) and \(\text{Var}(Z_2) = E(\text{Var}(Z_2|Z_1)) + \text{Var}(E(Z_2|Z_1)) = \mu_2 E(e^{\phi Z_1}) + \mu_2^2 \text{Var}(e^{\phi Z_1}) = \lambda_2 + \mu_2^2 \left\{\exp(\lambda_1 (e^\phi - 1)^2) - 1\right\}\) respectively. As expected from the definition, \(Z_2\) is overdispersed (variance greater than mean). Evidently, the marginal moments of \(Z_1\) are obtained from the Poisson ones.

Remark 2.1. In this article, we develop a different parameterization of the BCP distribution with \(\lambda_2\) being the marginal mean of \(Z_2\), differently from Berkhout and Plug (2004). This will be important for the definition of our bivariate INGARCH process in terms of mean parameters and has the advantage of being easier to establish first-order stationarity of the bivariate count process.

The joint probability function of \((Z_1, Z_2)\), say \(p(x, y) \equiv P(Z_1 = x, Z_2 = y)\), is given by

\[
p(x, y) = \frac{\lambda_1^x \lambda_2^y}{x!y!} \exp \left\{ -\lambda_1 (1 + y(e^\phi - 1)) - \lambda_2 \exp \left\{ -\lambda_1 (e^\phi - 1) + \phi x\right\} + \phi xy \right\},
\]

for \(x, y \in \{0, 1, 2, \ldots\}\). Joint moments for the BCP distribution are given in Berkhout and Plug (2004). The covariance between \(Z_1\) and \(Z_2\) is \(\text{cov}(Z_1, Z_2) = \lambda_1 \lambda_2 (e^\phi - 1)\) and therefore the correlation takes the form

\[
\text{corr}(Z_1, Z_2) = (e^\phi - 1) \sqrt{\frac{\lambda_1 \lambda_2}{1 + \lambda_2 (e^\phi (e^\phi - 1) - 1)}}.
\]

Remark 2.2. The parameter \(\phi\) controls the dependence of the model. For \(\phi = 0\), \(\phi > 0\), and \(\phi < 0\), we have respectively independence, positive and negative correlations. Another remarkable point is that this parameter does not have restrictions depending on the means, in contrast with the previous bivariate models considered for constructing bivariate INGARCH models. This will enable us to deal with highly correlated bivariate count time series.

Remark 2.3. Figure 1 illustrates the features of the model. For fixed values of \(\lambda_1\) and \(\lambda_2\), the cross-correlation value is high when \(\phi\) is close to 0, where positive small values of \(\phi\) imply a high positive correlation while small negative values imply a high negative correlation. From this figure, we can see that the BCP distribution accommodates a wide range of cross-correlation for various values of the mean parameters.

Remark 2.4. We now obtain explicitly the maximum and minimum points of (2) as function of \(\phi\), which can be expressed in terms of the Lambert function \(W(x)\), which is defined by the solution of the equation \(W(x) \exp(W(x)) = x\), for \(x \in \mathbb{R}\). This point has not been discussed in Berkhout and Plug (2004). As argued by Corless et al. (1996), for \(-e^{-1} \leq x \leq 0\), there are two possible solutions for \(W(\cdot)\): the solution satisfying \(-1 \leq W(x)\) is denoted by \(W_0(x)\) (which is known as the principal branch of the Lambert function), the other one satisfying \(W(x) \leq -1\) is denoted by \(W_1(x)\) (\(W(x)\) is used when the solution is unique). By taking the first derivative of (2) with respect to \(\phi\) and equating 0, we obtain that \(z e^z = e^{-1}(\lambda_2^{-1} - 1)\), where \(z = \lambda_1 (e^\phi - 1)^2 - 1\). From the results given in Corless et al. (1996), we have that \(z = W_0(e^{-1}(\lambda_2^{-1} - 1))\) when \(\lambda_2 \leq 1\). For \(\lambda_2 > 1\), we obtain the real solutions \(z = W_0(e^{-1}(\lambda_2^{-1} - 1))\) and \(z = W_1(e^{-1}(\lambda_2^{-1} - 1))\). Hence, explicit solutions in terms of \(\phi\) can be obtained as well as the theoretical range of correlation. Further, note that the parameter \(\phi\) is involved in the conditional distribution of \(Z_2\) given \(Z_1\). In particular, the parameter \(\phi\) also controls the variance of \(Z_2\). This results in a correlation between \(Z_1\) and \(Z_2\) being a non-monotonic function of \(\phi\). On the other hand, the covariance between \(Z_1\) and \(Z_2\) is \(\text{Cov}(Z_1, Z_2) = \lambda_1 \lambda_2 (e^\phi - 1)\), which is a monotonic function more easy to interpret.
With the above bivariate conditional Poisson distribution, we can define our proposed bivariate count process with a flexible range of contemporaneous correlation as follows.

**Definition 2.1.** (BCP-INGARCH process). Let \( Y_t = (Y_{t1}, Y_{t2})^\top \) be a bivariate count time series where \( t \geq 1 \). We say that \( \{Y_t\}_{t \geq 1} \) is a bivariate conditional Poisson INGARCH(1,1) process if it satisfies

\[
Y_t | \mathcal{F}_{t-1} \sim \text{BCP}(\lambda_{t1}, \lambda_{t2}, \phi),
\]

with

\[
\lambda_t \equiv E(Y_t | \mathcal{F}_{t-1}) = \omega + A\lambda_{t-1} + BY_{t-1},
\]

where \( \mathcal{F}_{t-1} = \sigma(Y_{t-1}, \ldots, Y_1, \lambda_1) \), for \( t \geq 2 \), \( \omega = (\omega_1, \omega_2)^\top \in \mathbb{R}^2 \) is the intercept vector, \( A = \{a_{ij}\}_{i,j=1,2} \) and \( B = \{b_{ij}\}_{i,j=1,2} \) are two \( 2 \times 2 \) matrices with non-negative entries/parameters, \( \phi \in \mathbb{R} \) is the contemporaneous dependence parameter, and \( Y_{t1} | F_0 \sim \text{BCP}(\lambda_{11}, \lambda_{12}, \phi) \), with \( F_0 = \sigma(\lambda_1) \) and \( \lambda_1 = (\lambda_{11}, \lambda_{21})^\top \).

**Remark 2.5.** Note that we have defined above explicitly the conditional distribution of \( Y_{t1} | F_0 \) following our baseline bivariate conditional Poisson distribution. In the existing bivariate INGARCH models, such an assumption is not mentioned, but it is implicitly used for showing, for example, that the bivariate Markov process \( \{\lambda_t\}_{t \geq 1} \) is an e-chain (for the definition of an e-chain, please see the beginning of Section 2.1), which is, in particular, important to obtain desirable theoretical properties of the bivariate count process \( \{Y_{t1}\}_{t \geq 1} \).

We have that \( E(Y_t) = E(\lambda_t) = \omega + AE(\lambda_{t-1}) + BE(Y_{t-1}) \), and under first-order stationarity, we obtain \( E(Y_t) = (I - A - B)^{-1} \omega \), for \( t \geq 1 \), where \( I \) is the identity matrix. Detailed discussion on the stationarity and ergodicity of the count process is provided in the following subsection.
2.1. Stability Theory

We now introduce some matrix notations to state the main results on stability theory for the proposed bivariate count process. We follow the notation used in Liu (2012). For a matrix $J \in \mathbb{C}^{m \times n}$ and $p \in [1, \infty]$, we denote $\|J\|_p = \max_{x \neq 0} \{\|Jx\|_p/\|x\|_p : x \in \mathbb{C}^n\}$, with $\|x\|_p^p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \in [1, \infty)$ and $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$ for $p = \infty$ being the $p$-norm of the vector $x$. For $p = 1$ and $p = \infty$, we have respectively that $\|J\|_p = \max_{1 \leq i \leq m} \sum_{j=1}^n |J_{ij}|$ and $\|J\|_\infty = \max_{1 \leq i \leq m} |J_{ij}|$, with $J_{ij}$ denoting the $(i,j)$th element of $J$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Further, define $\mathcal{H}$ to be the class of continuous real functions with compact support on $[0, \infty) \times [0, \infty]$. We now have all the ingredients to establish that the joint random bivariate vector $\{\lambda_t\}_{t \geq 1}$ is an e-chain. By $\{\lambda_t\}_{t \geq 1}$, be an e-chain it means that for any function $g \in \mathcal{H}$ and for every $\epsilon > 0$, there exists $\eta > 0$ such that $\|x - z\| < \eta$ implies $\left|E \left( g(\lambda_t) | \lambda_0 = x \right) - E \left( g(\lambda_t) | \lambda_0 = z \right) \right| < \epsilon$ for all $k \geq 1$, where $\|\cdot\|$ is some norm and $z, x \in (0, \infty) \times (0, \infty)$. To save space in the article, all the proofs of the theoretical results given in the article are presented in the Supplementary Material.

**Theorem 2.6.** If $\|A\|_p < 1$ for some $p \in [1, \infty]$, then $\{\lambda_t\}_{t \geq 1}$ is an e-chain.

**Remark 2.7.** A key ingredient to establish the e-chain property of $\{\lambda_t\}_{t \geq 1}$ is inequality (3) given in the Supplementary Material. Our approach given in (2) (Supplementary Material) only uses that one of the marginals (of the baseline bivariate distribution) is Poisson distributed. We mean the result holds whatever is the conditional distribution $p (m|n)$. Therefore, the argument used here is simpler and more general than those used in Liu (2012), Cui and Zhu (2018), and Cui et al. (2020), where other bivariate Poisson distributions are considered.

With Theorem 2.6 at hand, we can use the results given in Liu (2012) to obtain conditions ensuring stationarity and ergodicity for $\{(Y_t, \lambda_t)\}_{t \geq 1}$. Denote by $\rho(C)$ the largest absolute eigenvalue of a matrix $C$. Under the conditions $\rho(A + B) < 1$ and $\|A\|_p < 1$ for some $p \in [1, \infty]$, $\{(Y_t, \lambda_t)\}_{t \geq 1}$ has a unique stationary solution. If $\|A\|_p + 2^{1-1/p} \|B\|_p < 1$ for some $p \in [1, \infty]$, $\{(Y_t, \lambda_t)\}_{t \geq 1}$ has a unique stationary and ergodic solution.

**Remark 2.8.** It is noteworthy that Liu (2012) has established seminal results on the stationarity and ergodicity of bivariate INGARCH models. These results have been used for instance by Lee et al. (2018), Cui and Zhu (2018), and Cui et al. (2020).

2.2. Existing Bivariate INGARCH Models and their Limitations

We here discuss the existing bivariate INGARCH models and present some problems regarding the contemporaneous correlation, which relies on the ability for capturing the dependence of the baseline bivariate discrete distribution. The model in Liu (2012) and Lee et al. (2018) is defined by $Y_t | F_{t-1} \sim \text{BP}^* (\lambda_{1t}, \lambda_{2t}, \phi)$ and the dynamics for $\lambda_t$ as in (3), where $\text{BP}^*$ stands for the bivariate Poisson distribution obtained via the trivariate reduction method, assuming the form

$$P(Y_{1t} = y_1, Y_{2t} = y_2 | F_{t-1}) = e^{-\lambda_{1t} - \lambda_{2t} - \phi} (\lambda_{1t} - \phi)^{y_1} (\lambda_{2t} - \phi)^{y_2} y_1! y_2! \times \min_{s \in \{0, 1, \ldots, y_1\}} \left(\begin{array}{c} y_1 \\ s \end{array} \right) \left(\begin{array}{c} y_2 \\ s \end{array} \right) s! \left(\frac{\phi}{(\lambda_{1t} - \phi)(\lambda_{2t} - \phi)}\right)^s, \quad y_1, y_2 \in \{0, 1, \ldots, \},$$

with $\phi = \text{cov}(Y_{1t}, Y_{2t}) \in [0, \min(\lambda_{1t}, \lambda_{2t}))$ deterministic and does not depend on $t$. To ensure this last condition, the authors assumed that $\phi < \min(\lambda_{1t}, \lambda_{2t})$, with $(a_1, a_2)^T = (I - A)^{-1} \omega$ since $\lambda_t \geq (I - A)^{-1} \omega \forall t \geq 1$ when $\rho(A) < 1$ (Liu, 2012). The bivariate distribution in (4) has Poisson marginals conditional on $F_{t-1}$. One of the limitations of this model is that it does not allow for negative contemporaneous correlation and the parameter value that corresponds to independence lies on the boundary of the parameter space. As argued by Berkhourt and Plug (2004),
this bivariate Poisson distribution also does not accommodate higher values of positive correlation, especially for large values of the marginal means (see eq. (12) from that paper). All these restrictions on the parameter \( \phi \) imply compromised practical applicability, where a broad range of correlation is required.

We now discuss the model in Cui and Zhu (2018) given by \( Y_t | F_{t-1} \sim \text{BP}(\lambda_1, \lambda_2, \phi) \), with \( \lambda \), satisfying the dynamics as in (3). Here \( \text{BP} \) denotes the bivariate Poisson distribution with probability function

\[
P(Y_t = y_1, Y_{t+1} = y_2 | F_{t-1}) = e^{-c(\lambda_1 + \lambda_2)} \frac{\lambda_1^{y_1} \lambda_2^{y_2}}{y_1! y_2!} \{ 1 + \phi(e^{-\gamma_1} - e^{-c\lambda_1})(e^{-\gamma_2} - e^{-c\lambda_2}) \},
\]

for \( y_1, y_2 \in \{0, 1, \ldots\} \), where \( c = 1 - e^{-1} \). The parameter space related to \( \phi \) is incorrectly stated in that paper. The bivariate Poisson distribution in (5) belongs to a more general class of distributions proposed by Sarmanov (1966) and the correct range of \( \phi \) is well known in the literature. For instance, the correct range can be found in section 2.2 (after eq. (1)) from Hofer and Leitner (2012); see also Lee (1996). Using that result in the particular case given in (5), we obtain that

\[
-1 < \phi < \frac{1}{\max\{e^{-c\lambda_1}(1 - e^{-c\lambda_2}), e^{-c\lambda_2}(1 - e^{-c\lambda_1})\}},
\]

in contrast with the wrong range considered by Cui and Zhu (2018) \( |\phi| < \frac{1}{(1 - e^{-c\lambda_1})(1 - e^{-c\lambda_2})} \). The obvious implication of this incorrect bound is that the model there is not well defined. Moreover, note that (6) needs to be deterministic and independent of \( t \) as done by Lee et al. (2018) and therefore an additional restriction is necessary. Such restriction is not clearly discussed by Cui and Zhu (2018). It is worth noting that, although both negative and positive contemporaneous correlation are allowed in that model, its range is extremely limited due to the simplex structure in (5), as discussed by Cui et al. (2020).

In the papers above, one can observe two major problems regarding the range of cross-correlation: (1) natural restriction of the parameter space due to the baseline bivariate discrete distribution; (2) bounds of the parameter space depending on the conditional means, which are driven by stochastic dynamics. Our proposed model deals directly with both issues and naturally overcomes the limitations because the parameter space of \( \phi \) is \( \mathbb{R} \)-valued (so it does not depend on the means) and the baseline distribution allows for a broad range of correlation. Recently, Cui et al. (2020) proposed a bivariate Poisson INGARCH model having a more general structure as an alternative to the simplex form in (5). Although this provides greater flexibility for modeling dependence, it adds a cumbersome normalizing constant in the joint probability function, which involves a double infinite summation. Further, it is difficult to assess the possible range of correlation.

Another related work is due to Fokianos et al. (2020), where a multivariate INGARCH model is elegantly introduced based on a latent copula approach. The paper studied the stochastic properties of the multivariate count process and estimate parameters via a quasi-likelihood approach, which is equivalent to modeling the multivariate counts under the assumption of contemporaneous independence (see section 4 of that paper). A tricky point that arises in this approach is the estimation of the parameter responsible for controlling the cross-correlation. Further, like the model proposed by Cui et al. (2020), it is hard to assess the possible range of contemporaneous correlation.

As demonstrated in the following section, our likelihood function assumes a very simple form, and the parameters are jointly estimated (including the cross-correlation parameter) via the conditional maximum likelihood method.

### 3. STATISTICAL INFERENCE AND ASYMPTOTIC RESULTS

To estimate the parameters in the BCP-INGARCH model, we consider the conditional maximum likelihood (CML) approach. Here, we also determine conditions that ensure asymptotic normality of the CML estimators. Denote by \( \theta = (\text{vec}(A), \text{vec}(B), \phi)^\top \) the parameter vector, where \( \text{vec}(A) \equiv (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})^\top \) and \( \text{vec}(B) \equiv (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^\top \).

J. Time Ser. Anal. 44: 206–222 (2023) © 2022 John Wiley & Sons Ltd wileyonlinelibrary.com/journal/jtasa

DOI: 10.1111/jtsa.12663
Due to the Markovian property of the process, the conditional joint probability function of \( \{ \mathbf{Y}_t \}_{t=2}^{n} \) given \( \mathbf{Y}_1 = \mathbf{y}_1 \) is 
\[ P(\mathbf{Y}_n = \mathbf{y}_n, \mathbf{Y}_{n-1} = \mathbf{y}_{n-1}, \ldots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1) = \prod_{t=2}^{n} P(\mathbf{Y}_t = \mathbf{y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}), \]
where \( n \) is the sample size. The conditional log-likelihood function is given by 
\[ \ell'(\theta) = \sum_{t=2}^{n} \ell'_t(\theta), \]
with
\[ \ell'_t(\theta) = \log \lambda_{1t} + \log \phi_{\tau t} - \log \lambda_{2t} - \log (1 + \phi_{\tau t}(e^\phi - 1)) - \lambda_{2t} \exp \{-\lambda_{1t}(e^\phi - 1) + \phi y_{1t}\} + \phi y_{1t} y_{2t}. \]

where \( y_t = (y_{1t}, y_{2t})^T \) denotes the observed value of \( Y_t = (Y_{1t}, Y_{2t})^T \) for \( t = 1, \ldots, n \). The conditional maximum likelihood estimator (CMLE) of \( \theta \) is given by \( \hat{\theta} = \arg \max_{\theta} \ell'(\theta) \), with \( \Theta \) denoting the parameter space. To derive \( \hat{\theta} \), we employ the numerical optimization routine provided by the Stan software (Stan Development Team, 2020) through the R (R Core Team, 2020) package rstan. Stan is a platform for high-performance statistical computation where fast results are achieved through compilation in C++. Additionally to the efficiency gain, we have found in this and other works that Stan’s numerical maximization can yield superior results (in terms of finding the CML estimators) in comparison to the standard optim command in R.

The associated score function to the log-likelihood \( \ell'(\theta) \) is denoted by \( U(\theta) = \partial \ell'(\theta)/\partial \theta \), where its components are given in the Supplementary Material. The next result gives us some properties of the score function useful to establish the asymptotic normality of the CML estimators. To establish the asymptotic properties of the CMLEs, we will assume that \( \mathbf{A} \) (see equation (3)) is a diagonal matrix with elements \( a_1 \equiv a_{11} \) and \( a_2 \equiv a_{22} \). As discussed by Lee et al. (2018), this facilitates the study of asymptotic properties of the maximum likelihood estimators. Anyway, we allow for the matrix \( \mathbf{A} \) to be non-diagonal in our numerical experiments and \( \mathbb{R} \) code as well.

**Theorem 3.1.** Let \( U(\theta) = \partial \ell'(\theta)/\partial \theta \). We have that \( \{ U(\theta); F_{r-1} \} \) is a martingale difference sequence. Further, the total score function \( U(\theta) \) satisfies the information matrix equality
\[ -E(\nabla U(\theta)) = E(U(\theta)U(\theta)^T). \quad (7) \]

The next point is to develop the asymptotic normality of the conditional maximum likelihood estimator, where some regularity conditions are necessary as follows.

**Assumption 3.2.** There exists \( p \in [1, \infty) \) such that \( \| A \|_p + 2^{1-1/p}\| B \|_p < 1 \).

**Assumption 3.3.** The true parameter value \( \theta_0 \) is an interior point of \( \Theta \), with \( \Theta \) being a compact set.

**Remark 3.4.** In the simulated and real data analyses, we consider that Assumption 3.2 is in force with \( p = 1 \).

Expanding \( U(\hat{\theta}) \) in Taylor’s series around \( \theta_0 \), we obtain that \( 0 = U(\hat{\theta}) = U(\theta_0) + (\hat{\theta} - \theta_0)\nabla U(\hat{\theta}) \), where \( \hat{\theta} \) belongs to the segment connecting the points \( \hat{\theta} \) and \( \theta_0 \). We rearrange the terms to obtain that
\[ \sqrt{n}(\hat{\theta} - \theta_0) = \frac{U(\theta_0)}{\sqrt{n}}. \quad (8) \]

Under Assumption 3.2, the stationarity and ergodicity of \( \{ (Y_t, \lambda_t) \}_{t=1} \) implies in the same properties for \( U(\theta_0) \). Hence, we use the Central Limit Theorem for Martingales (Hall and Heyde, 1980) and get that \( \frac{U(\theta_0)}{\sqrt{n}} \stackrel{d}{\longrightarrow} N(0, I(\theta_0)) \) as \( n \to \infty \), where \( I(\theta_0) \) is the Fisher information matrix which can be obtained as the limit in probability \( I(\theta_0) = \lim_{n \to \infty} n^{-1} \sum_{t=2}^{n} E(\nabla U(\theta_0)|F_{t-1}) \).

Now, under Assumptions 3.2 and 3.3, we apply the Law of Large Number for Martingales and follow the steps of the proof of Proposition 5 by Lee et al. (2016) (see also Lemma 6 by Lee et al., 2018), to obtain that
\[ \frac{-\nabla U(\theta_0)}{n} \mathop{\longrightarrow}^{a.s.} I(\theta_0) \text{ as } n \to \infty. \]

By combining the above results in (8), we have that \( \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{n}} \stackrel{d}{\longrightarrow} N(0, I^{-1}(\theta_0)) \).

With this asymptotic normality at hand, we can assess standard errors, construct confidence intervals for the parameters, and test the hypothesis of interest. In the next section, we provide some simulated results aiming at
(i) study of the finite-sample performance of the CMLEs; (ii) the evaluation of two strategies to get the standard errors of the parameter estimates; and (iii) hypothesis testing for $H_0 : \phi = 0$ against $H_1 : \phi \neq 0$ (testing the presence of contemporaneous correlation). For this last aim, we compare the performance of the likelihood ratio and score tests. Due to the asymptotic normality of the conditional maximum likelihood estimators and the fact that $\phi = 0$ does not belong to the boundary space, these statistics are asymptotically $\chi_1^2$-distributed under the null hypothesis.

4. SIMULATION STUDIES

4.1. Point Estimation

Here, we assess the finite-sample behavior of the conditional maximum likelihood estimators for the BCP-INGARCH(1,1) model through Monte Carlo simulations. In this study, 1000 replications are used, and the sample sizes $n = 200, 500$ are investigated. The study is conducted considering the diagonal and non-diagonal versions of our model, which we denote as Configurations (a) and (b) respectively. Further, two different settings for the parameter values are considered. In the first one, we set Configuration I (b) to be $\alpha_1 = 0.4$, $\alpha_2 = 0.3$, $\beta_{11} = 0.3$, $\beta_{12} = 0.2$, $\beta_{22} = 0.2$, $\omega_1 = 1$, $\omega_2 = 2$, $\phi = 0.3$, where setting $\beta_{21} = 0$, $\beta_{12} = 0$ yields Configuration I (a). Details on Configuration II and their results are reported in the Supplementary Material. We would like to grab attention that under $\phi$ values ‘close’ to zero, a high correlation between the components is produced as illustrated in Figure 1; see also Remark 2.3. For instance, the average correlation from the 1000 trajectories simulated under I (a) is 0.715.

We simulate the BCP-INGARCH process by employing a burn-in period of length 300 to reduce the influence of the initial values in the simulated trajectories. That is, we generate bivariate count time series with sample size $n + 300$ and discard the first 300 pairs, so we end up with $n$ pairs. Results of parameter configurations of type I are reported in Table I, which displays the empirical mean and standard errors (SD) of the parameters as well as the mean squared error (MSE). It is shown that the conditional maximum likelihood estimation behaves well, displaying average estimates that are close to the true values used to generate the data. In particular, the correlation parameter $\phi$ is very well estimated under both sample sizes and configurations. As expected, the standard deviations and MSEs decrease as the sample size increases. In the simulations, we have used $\phi = 0$ as an initial guess for the optimization. For the remaining parameters, the fit of univariate INGARCH models provides good starting values. We now investigated the influence of the initial guess on $\phi$ through a simulated dataset under Configuration I (a) with the true values $\phi = -0.3, 0.3$. For the case $\phi = 0.3$, we obtained the very same estimate $\hat{\phi} = 0.337$ and $\hat{\phi} = 0.330$ for $n = 200$ and $n = 500$ respectively, for all the initial guesses $-2, -1, 0, 1, 2$. Similarly, for the case $\phi = -0.3$, we got $\hat{\phi} = -0.257$ and $\hat{\phi} = -0.315$ for $n = 200$ and $n = 500$ respectively, for all the initial guesses $-2, -1, 0, 1, 2$ considered. The good estimation of the parameter $\phi$ can be explained by the simple form assumed by the likelihood function. For the simulated datasets above, we display the profile likelihood of $\phi$ in Figure 2, which is a well-behaved concave function.

An additional parameter configuration, denoted by II, is investigated with results reported in the Supplementary Material. Similar to I, II (a) and (b) are diagonal and full variations of the BCP-INGARCH model. Results under II show the same behavior allowing us to conclude that conditional maximum likelihood estimation works well for all parameters of the BCP-INGARCH(1,1) process. An excellent average estimation was achieved even for the small sample size of $n = 200$ under all configurations, and increasing the sample size produces estimates with low variability, as reflected by the decrease in standard deviation and mean squared errors.

4.2. Standard Errors

Alternative asymptotic methods proposed in the literature to obtain the standard errors of INGARCH model parameters are investigated in this section. We consider the following consistent estimators for the Fisher information
Table I. Empirical mean, standard errors (SD), and mean squared error (MSE) of the Monte Carlo estimates for the BCP-INGARCH model under Configurations I (a) and I (b).

| n    | \(a_1 = 0.4\) | \(a_2 = 0.3\) | \(\beta_{11} = 0.3\) | \(\beta_{21} = 0.2\) | \(\beta_{12} = 0.1\) | \(\beta_{22} = 0.2\) | \(\omega_1 = 1\) | \(\omega_2 = 2\) | \(\phi = 0.3\) |
|------|----------------|----------------|---------------------|---------------------|---------------------|---------------------|-----------------|----------------|----------------|
| I (a) | 200            | Mean           | 0.388               | 0.261               | 0.292               | –                   | –               | 0.196          | 1.065          | 2.178          | 0.302          |
|       |                | SD             | 0.126               | 0.193               | 0.054               | –                   | –               | 0.061          | 0.339          | 0.712          | 0.017          |
|       |                | MSE            | 0.016               | 0.039               | 0.003               | –                   | –               | 0.004          | 0.119          | 0.539          | 0.000          |
| 500   | Mean           | 0.395          | 0.285               | 0.296               | –                   | –                   | 0.196           | 1.032          | 2.080          | 0.300          |
|       |                | SD             | 0.078               | 0.137               | 0.033               | –                   | –               | 0.040          | 0.203          | 0.499          | 0.011          |
|       | MSE            | 0.006          | 0.019               | 0.001               | –                   | –                   | 0.002           | 0.042          | 0.255          | 0.000          |
| I (b) | 200            | Mean           | 0.393               | 0.264               | 0.288               | 0.171               | 0.101           | 0.206          | 1.095          | 2.327          | 0.300          |
|       |                | SD             | 0.084               | 0.174               | 0.07                | 0.116               | 0.026           | 0.086          | 0.278          | 0.893          | 0.012          |
|       | MSE            | 0.007          | 0.032               | 0.005               | 0.014               | 0.001               | 0.007           | 0.086          | 0.905          | 0.000          |
| 500   | Mean           | 0.402          | 0.292               | 0.293               | 0.192               | 0.099               | 0.196           | 1.026          | 2.104          | 0.301          |
|       |                | SD             | 0.053               | 0.125               | 0.045               | 0.089               | 0.015           | 0.058          | 0.158          | 0.566          | 0.007          |
|       | MSE            | 0.003          | 0.016               | 0.002               | 0.008               | 0.000               | 0.003           | 0.026          | 0.331          | 0.000          |

Figure 2. Profile likelihood function of \(\phi\) for simulated datasets with \(n = 200, 500\) and true values \(\phi = -0.3, 0.3\)

matrix (Ferland et al., 2006):

\[
\hat{S}_n = \frac{1}{n} \sum_{t=2}^{n} U_t(\hat{\Theta}) U_t^T(\hat{\Theta}) \quad \text{and} \quad \hat{D}_n = -\frac{1}{n} \sum_{t=2}^{n} H_t(\hat{\Theta}),
\]

where \(U_t(\cdot)\) and \(H_t(\cdot)\) denote the score function and Hessian matrix associated to the \(t\)th bivariate count observation respectively, for \(t = 2, \ldots, n\), and \(\hat{\Theta}\) is the CML estimator of \(\Theta\).

The same trajectories simulated under Configurations I (a) and II (a) in Section 4.1 are now used for investigation of the competing methods to derive the CMLEs’ standard errors and confidence intervals. Tables 2 and 3 from the Supplementary Material are dedicated to comparing the average standard errors from \(\hat{S}_n\) and \(\hat{D}_n\) to the Monte Carlo standard deviation. It is observed that, on average, both methods approach well the Monte Carlo standard deviation.
under both sample sizes and show similar variability. Moreover, the \( \hat{S}_n \) and \( \hat{D}_n \) standard errors are also used to construct confidence intervals for the BCP-INGARCH model parameters based on the asymptotic normality of the estimators. The normal approximation is investigated via histograms of the standardized CML estimates that are shown alongside a standard normal density curve in the Supplementary Material (see Figures 1 and 2 there). Given that the normal approximation is satisfactory, it is reasonable to use normal quantiles to derive confidence intervals for the model parameters.

Section 1.2 from the Supplementary Material also includes the empirical coverage of intervals at 1%, 5%, and 10% significance levels derived from using \( \hat{S}_n \) or \( \hat{D}_n \) standard errors; see Tables 4–7 there. Under both parameter configurations, our simulation studies revealed that the empirical coverage of the \( \hat{S}_n \) and \( \hat{D}_n \)–based intervals are close to the nominal levels for all model parameters. Moreover, excellent results are achieved even under a small sample size, which demonstrates that the two asymptotic methods work very well for the uncertainty quantification of the proposed model.

### 4.3. Hypothesis Testing

Methods to test for the presence of contemporaneous correlation between the count time series are evaluated in this section. Under a BCP-INGARCH model, we would like to test the hypothesis \( H_0 : \phi = 0 \) versus \( H_1 : \phi \neq 0 \). We evaluate two asymptotic tests, the likelihood ratio test (LRT) and the score test via a simulation study. The Monte Carlo probability of rejecting \( H_0 \) in favor of \( H_1 \) is calculated with 1000 replications for each \( \phi \) in the range \([-1, 1]\) with 0.1 spacing. In each Monte Carlo replication, model parameters are estimated under \( H_0 \) and \( H_1 \), denoted as \( \hat{\theta} \) and \( \hat{\theta} \) respectively. The LRT statistic is calculated as \(-2(\ell(\hat{\theta}) - \ell(\hat{\theta}))\), where \( \ell(\cdot) \) is the log-likelihood function. The score test relies only on the model parameters estimated under \( H_0 \) and its test statistic is given by \( U(\hat{\theta})^T F^{-1}(\hat{\theta}) U(\hat{\theta}) \), where \( U(\cdot) \) and \( F(\cdot) \) denotes the score function and the model’s Fisher Information matrix respectively, with the first being calculated analytically via expressions provided in Section 2 from the Supplementary Material and the former by numerical differentiation.

The simulation study is carried for a setting where \( A \) and \( B \) are diagonal matrices with the true parameter values \((\alpha_1, \alpha_2, \beta_{11}, \beta_{22}, \omega_1, \omega_2) = (0.4, 0.3, 0.2, 0.4, 1, 1)^T\). We refer to this specification as Scenario I. A configuration where the matrix \( B \) is non-diagonal is chosen in Scenario II, with the parameter vector \((\alpha_1, \alpha_2, \beta_{11}, \beta_{21}, \beta_{22}, \omega_1, \omega_2) = (0.3, 0.2, 0.3, 0.1, 0.2, 0.2, 1, 0.5)\). We set the significance level of both tests at 5%. Figure 3 displays the power of the likelihood ratio and score tests as function of \( \phi \) under both Scenarios 1 and 2 and sample sizes equal to 200 and 500.

Both the likelihood ratio and score tests demonstrate the ability to reject the null hypothesis (power) with a high probability when \( \phi \neq 0 \). However, the score test suffers from numerical problems at high positive values of \( \phi \), as can be observed from Figure 3. This issue arises from numerical differentiation employed to calculate the Hessian matrix, causing the rejection probability to decrease in this region. Notably, this is more severe in Scenario II where the number of parameters increases.

### 5. TESTING GRANGER CAUSALITY

We can test Granger (1969) causality involving counts based on the BCP process. It is said that \( \{Y_{1t}\}_{t \geq 1} \) causes to \( \{Y_{2t}\}_{t \geq 1} \) in mean if \( P\{E(Y_{2t} | F_{2,t-1}) \neq E(Y_{2t} | F_{1,t-1})\} > 0 \), where \( F_{1,t} = \sigma(Y_{1t}, \ldots, Y_{11}, \lambda_{11}) \), \( F_{2,t} = \sigma(Y_{2t}, \ldots, Y_{21}, \lambda_{21}) \), and \( F_t \) as before. From the formulation in (3), we have that \( \{Y_{1t}\}_{t \geq 1} \) does not cause to \( \{Y_{2t}\}_{t \geq 1} \) in mean if \( \phi = 0, \alpha_{21} = 0, \) and \( \beta_{21} = 0 \). The parameter \( \phi \) here characterizes the instantaneous causality in mean. Further, as the conditional variance in the BCP model depends on the mean, we are able to simultaneously test the causality in mean and variance; see Guo et al. (2014) for a simultaneous test in mean and variance under a (continuous) factor double autoregressive model. For a Granger causality test based on a univariate INGARCH approach and under a Bayesian perspective, see Chen and Lee (2017).

The hypotheses of interest here are \( H_0 : \phi = 0, \alpha_{21} = 0, \beta_{21} = 0 \) and \( H_1 : \text{at least one of the following parameters } \phi, \alpha_{21}, \beta_{21} \text{ different from zero} \). The null hypothesis involving the \( \alpha \)'s and \( \beta \)'s lie on the boundary of the...
Figure 3. Power of the likelihood ratio test (LRT) and score test as function of \( \phi \) with data generated under Scenarios I and II and sample sizes \( n = 200 \) (solid line) and \( n = 500 \) (dashed line)

parameter space, and therefore the classical inference results do not hold. We propose using the likelihood ratio test combined with the restricted parametric bootstrap by Cavaliere et al. (2017) to deal with this problem. While the usual (or unrestricted) procedure simulates trajectories with CMLEs obtained under the alternative hypothesis, model parameters fitted under \( H_0 \) are used to replicate the data in restricted bootstrap. We will use this test to set which time series will be considered as \( \{Y_1 t\} \) and the other to be \( \{Y_2 t\} \). A practical illustration of this Granger causality test is given in the next section. We call the attention that a similar restricted bootstrap can also be employed for testing if any of the matrices \( A \) or \( B \) (or even both simultaneously) are diagonal.

6. BIVARIATE HEPATITIS COUNT DATA ANALYSIS

Here, the proposed methodology is applied to the confirmed monthly cases of viral hepatitis recorded at two nearby Brazilian cities. Hepatitis is an inflammation of the liver, most commonly caused by a viral infection. Symptoms can take some time to develop, only manifesting after the liver function has been affected. In Brazil, the most common types of hepatitis are A, B, and C. The data are made available by the Brazilian public healthcare system SUS through the site https://datasus.saude.gov.br (DATASUS platform). It currently comprises the period of 2001–2018, giving a total of \( n = 216 \) observations per city. We analyze the data of Brazil’s capital Brasília, which is located in the Federal District within Goiás state. Due to proximity, it is natural to expect that Brasília’s counts are correlated to the Goiás’s capital, Goiânia. Their geographic locations are shown on the right side of Figure 4 with corresponding time series of confirmed viral hepatitis cases displayed on the left, which shows how the counts tend to be correlated over time, with their Pearson’s empirical correlation being 0.50.

Our goal here is to jointly model and predict the monthly counts of hepatitis cases in Brasília and Goiânia. As we will discuss in what follows, most existing bivariate INGARCH models cannot handle this problem due to a constrained range of supported cross-correlation (see also discussion in Section 2.2).
Table II. Parameter estimates and standard errors (under parenthesis) of BCP-INGARCH(1,1) model fits to the hepatitis data in the Brazilian capitals of Goiânia ($Y_1$) and Brasília ($Y_2$)

|       | $\alpha_{11}$   | $\alpha_{22}$   | $\beta_{11}$   | $\beta_{12}$   | $\beta_{21}$ |
|-------|-----------------|-----------------|----------------|----------------|----------------|
| BCP-D | 0.466 (0.057)   | 0.482 (0.040)   | 0.430 (0.042)  | - (-)         | - (-)         |
| BCP-ND| 0.497 (0.059)   | 0.448 (0.044)   | 0.398 (0.046)  | 0.002 (0.007)  | 0.019 (0.032) |

|       | $\beta_{22}$   | $\omega_1$     | $\omega_2$    | $\phi$        |
|-------|----------------|----------------|---------------|---------------|
| BCP-D | 0.384 (0.026)  | 2.310 (0.578)   | 6.519 (0.910) | 0.010 (0.001) |
| BCP-ND| 0.399 (0.030)  | 2.216 (0.612)   | 7.012 (1.134) | 0.010 (0.001) |

Note: BCP-D and BCP-ND correspond to different setting of $B$ matrix as diagonal or non-diagonal respectively.

The causality approach introduced in Section 5 is taken for choosing how the cities are assigned to the model components, where we test for causation from $\{Y_{1t}\}$ to $\{Y_{2t}\}$ with the null hypothesis $H_0 : \phi = \alpha_{21} = \beta_{21} = 0$. Setting Brasília to $\{Y_{1t}\}$ yields a restricted bootstrap $p$-value (100K replications) of 0.089, which becomes 0.041 when Goiânia is assigned to the first component in the model. We can reject $H_0$ in favor of $H_1 : \text{at least one of } \phi, \alpha_{21}, \beta_{21} \neq 0$ at 5% confidence level in the former case, concluding that there is evidence of Goiânia hepatitis counts Granger-causing those of Brasília. As proposed by Berkhout and Plug (2004) (page 353) for the bivariate conditional Poisson distribution, the choice of the components can be determined by the data through the best fitting. On the other hand, if there is a clear causality relation between both counts, this can be used to determine the components. Because we have found that the Goiânia hepatitis counts Granger-cause those of Brasilia but not the opposite, we assign $\{Y_{1t}\}$ to Goiânia and $\{Y_{2t}\}$ to Brasília.

The diagonal and non-diagonal BCP-INGARCH models are fitted to the data. Estimates and standard errors of the BCP and BP model parameters are reported in Table II with R code fitting the proposed process available from the github page https://github.com/luizapiancastelli/rbcpingarch.

Both models capture a positive cross-correlation between the count time series once $\phi > 0$ in all cases. The ‘small’ $\phi$ estimate suggests that the cross-correlation is high, but we can formally test the hypothesis $H_0 : \phi = 0$
versus $H_1: \phi \neq 0$ with the methodology introduced in Section 4.3. The LRT statistics due to the BCP-D and BCP-ND models (and $p$-values in parenthesis) are 68.06 ($1 \times 10^{-16}$) and 61.02 ($5 \times 10^{-15}$) respectively. Those from the score test are 69.13 ($1 \times 10^{-16}$) and 66.48 ($3 \times 10^{-16}$), so we can conclude that there is a statistically significant contemporaneous correlation between the count time series from both methods.

Such serial cross-correlation can be explained by the forms of transmission of viral hepatitis and the cities’ close geographical proximity. While hepatitis B and C are mainly contracted via sexual contact and parenteral form, hepatitis A contagion is predominantly fecal-oral. Hence, one possible explanation for a high cross-correlation among the confirmed cases of the disease at Brasília and Goiânia is the sharing of food and water sources that relate directly to the transmission of hepatitis A. From an epidemiological perspective, it would be interesting to repeat this study by considering separately the hepatitis cases of each type and exploring whether the contemporaneous behavior maintains as strong when modeling the three types of the disease separately.

For the BCP models, we can assess how the between-cities dependency varies over time by computing the estimated contemporaneous correlation (conditional on the past) with equation (2) and CMLEs. This is displayed in Figure 5 where a decreasing tendency of cross-correlation over time is shown, although peaks occur around 2005 and 2016. These are periods with pronounced peaks of cases in Brasília that are accompanied by an increased number of counts at Goiânia. The conditional contemporaneous correlation estimated by the diagonal and non-diagonal BCP-INGARCH model fits are very similar and mostly overlap.

To compare the competing model fits to the data, one possibility is to use information criteria. We consider the AIC and BIC as selection criteria, and both of them indicate the diagonal version of the proposed model to be preferred over the BCP-ND model. The computed AIC and BIC values are provided in the Supplementary Material (Section 4.1). Choosing the BCP-D model over the non-diagonal case suggests that a leading/lagging relationship among the counts is not statistically significant. In other words, the effects of lagged Brasília counts in Goiânia and vice versa are not statistically different than zero. This implies that cross-correlation is purely contemporaneous and not driven by a leading/lagging relationship among the count time series.

In Figure 6, we provide the histograms of empirical 1-lag autocorrelations and cross-correlation for Brasília and Goiânia based on generated samples (1000 replications) from the fitted BCP model. As seen, our methodology captures well the first-order autocorrelation and the correlation between both count series, which is the main focus of the present article.
Moreover, in Section 4.2 of the Supplementary Material, models are compared via out-of-sample prediction, a criterion that further supports modeling this data application with the BCP-D process. Here, joint prediction of a pair \( (\hat{Y}_{t+1}, \hat{Y}_{t+1}) \) is done, but there is also the possibility of performing the predictions based on the conditional distribution of \( Y_{t+1} | Y_{t+1} \) under BCP-INGARCH modeling, something that we detail in Section 4.3 of the article Supplementary Material. In Section 4.4 of the Supplementary Material, we provide a goodness-of-fit assessment of the chosen model, which is done via Pearson and pseudo-Normal (Dunn and Smyth, 1996) residuals analysis.

We aimed to include other existing BCP-INGARCH models such as those by Liu (2012); Lee et al. (2018) and Cui and Zhu (2018) to the comparison but were unable to fit them to our data application of interest. For the former, we found that it can be problematic to estimate the cross-correlation parameter \( \delta \) due to its small contribution to the likelihood relative to the other parameters in the model. Failure to fit the bivariate Poisson INGARCH by Liu (2012) and Lee et al. (2018) to this data set is most likely due to the limited range of correlation of this model under higher \( \lambda \). As discussed in Section 2.2, the maximum value of \( \phi \) (hereby \( \phi_{\text{max}} \)) is \( \min(a_1, a_2) \), where \( (a_1, a_2)^T = (I - A)^{-1} \omega \), so the maximum cross-correlation supported at time \( t \) under this model is \( \text{corr}_{t}^{\text{max}} = \phi_{\text{max}} / (\lambda_1 \lambda_2) \). We can get a rough idea on the \( \text{corr}_{t}^{\text{max}} \) limits employing the BCP-D fitted \( \hat{\lambda}, \hat{\phi} \) and \( \hat{\lambda} \). This is reasonable given that the INGARCH models share the same \( \lambda \) specification and gives us that the maximum possible value of \( \text{corr}_{t}^{\text{max}} \) is 0.29 for \( t = 1, \ldots, 216 \), given the parameter restrictions. Figure 5 illustrates that the fitted (conditional) cross-correlation under the BCP models can be as high as 0.6, and is often above the constrained 0.29 value.

For a careful assessment, we tried replicating the model fit to the publicly available data set on the weekly number of syphilis cases in Pennsylvania and Maryland from 2007 to 2010, reported by the authors in Cui et al. (2020). These data are made available by the R package ZIM (https://cran.r-project.org/web/packages/ZIM/ZIM.pdf). Our conditional maximization routine for this model, available in Section 3 of the Supplementary Material, returned \( \hat{\phi} = 0.923 \) from various parameter initializations, with log-likelihood up to proportionally above 388.583 in all cases. This is inconsistent with results reported by Cui et al. (2020) (table 9), where \( \hat{\delta} \equiv \hat{\phi} = 0.747 \) and maximized log-likelihood value of 388.144. In addition to the observation that the authors’ maximization procedure was unable to find the true CMLEs, some further comments are made below.

1. The reported point estimate of \( \phi \) is positive in a scenario where there is a negative dependency between the count time series (the empirical Pearson’s correlation is \( -0.1355 \) for this data set). It is expected that \( \phi < 0 \), given that the sign of this parameter determines the sign of the cross-correlation in their model.
2. It is unclear whether the stationarity and ergodicity condition in Assumption 3.2 was met for the CMLEs. In practice, it is common to verify that with $p = 1$, in which case it is not satisfied for the reported $\hat{A}$ and $\hat{B}$.

In our view, the estimation of the cross-correlation parameter could be improved by performing the maximization in two steps, first estimating the parameters related to the conditional mean and $\phi$ subsequently, but this is outside of our scope.

The hepatitis data analysis explored in this article evidenced how other competing bivariate INGARCH models can be of limited practical application due to constricted ranges of supported cross-correlation. For both the bivariate Poisson INGARCH models by Liu (2012), Lee et al. (2018) and Cui and Zhu (2018), the limitation arises from the dependency parameter space being linked to the conditional means. Not only does this become a more severe restriction as the bivariate counts increase in value, but maximization is also complicated. Such challenges are overcome by the proposed BCP-INGARCH process where the cross-correlation parameter lies on the real line and a simple likelihood specification does not pose any complications to numerical maximization.

7. CONCLUDING REMARKS

We developed a novel bivariate conditional Poisson INGARCH process for modeling correlated count time series data having as the main advantage of its capability of capturing a wide range of contemporaneous correlation. This flexibility is important because bivariate/multivariate count time series data are prevalent in many fields and that there is a lack of flexible bivariate models based on the INGARCH approach, which is a relevant tool for dealing with univariate count time series. We here showed that it is possible to construct promising models based on such an approach.

The stability theory of our bivariate count process was established. Through simulation studies, we demonstrated that the proposed conditional maximum likelihood estimation works well and evaluated different methods of obtaining parameter standard errors. Asymptotic properties of the estimators were also derived. Hypothesis testing for the presence of cross-correlation under our model was presented and evaluated through likelihood ratio and score tests, demonstrating the power of such tests. A Granger causality test was also discussed and applied. Finally, the proposed methodology was employed in an application to counts of hepatitis cases in nearby Brazilian cities. The series showed to be highly positively correlated and modeling the data jointly was successfully done through our proposed model. The limitation of some bivariate INGARCH models was discussed both theoretically and empirically.

We now discuss some possible points for future research. As demonstrated along with this article, a key ingredient to propose a bivariate INGARCH process being mathematically tractable and having flexible contemporaneous correlation structure relies on the baseline bivariate count distribution. For example, other models can be proposed by assuming $Z_1 \sim \text{NB}(\lambda_1, \sigma)$ (see beginning of Section 2) if an exceeding overdispersion needs to be accounted for the count time series $\{Y_{1t}\}$, where $\text{NB}(\lambda_i, \sigma_i)$ stands for a negative binomial distribution with mean $\lambda_i$ and dispersion parameter $\sigma_i$. In a similar fashion, by assuming $Z_2|Z_1 = z_1 \sim \text{NB}(\mu_2 e^{\phi z_1}, \sigma_2)$, with $\mu_2$ as defined in Section 2, we can account for a wider range of overdispersion related to $\{Y_{2t}\}$. Other alternatives to the negative binomial assumption like COM-Poisson, zero-inflated/deflated count models, or Poisson inverse-Gaussian distributions can be considered for attacking underdispersion, overdispersion, zero-inflation/deflation, and heavy-tailed counts, just to name a few. We call the attention that in any of these extensions, many of the developed methodologies given in this article can be straightforwardly adapted. It is worth mentioning that codes are available, which is useful for practitioners and applied statisticians as well as for comparison purposes of new emerging multivariate count time series methods.

Another important point to be addressed is a multivariate extension allowing for a higher dimension rather than 2. The model proposed by Fokianos et al. (2020) allows for dealing with $d \geq 2$ correlated count time series, which is an attractive feature over the existing multivariate INGARCH models (including our proposed process). We hope to address this problem in a future paper. Other topics deserving further research are (a) inclusion of covariates, which can be done via a log-linear structure like in Fokianos et al. (2020); (b) BCP-INGARCH($p, q$) high-order extension with $p, q \geq 1$; and (c) nonlinear BCP-INGARCH generalization.
The authors thank the Associate Editor and two Referees for their constructive criticisms and comments that lead to an improvement of the paper. Luiza S. C. Piancastelli thanks the financial support of Science Foundation Ireland under Grant number 18/CRT/6049. Wagner Barreto-Souza and Hernando Ombao acknowledge the financial support by KAUST Research Fund. Wagner Barreto-Souza also thanks the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil, Grant number 305543/2018-0).

Data Availability Statement
The data are made available by the Brazilian public healthcare system SUS through the site https://datasus.saude.gov.br (DATASUS platform). Further, the data are available in the Supplementary Material as well as the other R codes. The codes for fitting our proposed bivariate INGARCH model are available in the github page https://github.com/luizapiancastelli/rbcpingarch.

Supporting Information
Additional Supporting Information may be found online in the supporting information tab for this article.

References
Berkhout P, Plug E. 2004. A bivariate Poisson count data model using conditional probabilities. *Statistica Neerlandica* 58:349–364.

Cameron AC, Trivedi PK. 1998. *Regression Analysis of Count Data* Cambridge University Press, Cambridge.

Cavaliere G, Nielsen H, Rahbek A. 2017. On the consistency of bootstrap testing for a parameter on the boundary of the parameter space. *Journal of Time Series Analysis* 38:513–534.

Chen CWS, Lee S. 2017. Bayesian causality test for integer-valued time series models with applications to climate and crime data. *Journal of the Royal Statistical Society - Series C* 66:797–814.

Christou V, Fokianos K. 2014. Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis* 35:55–78.

Christou V, Fokianos K. 2015. Estimation and testing linearity for non-linear mixed Poisson autoregressions. *Electronic Journal of Statistics* 9:1357–1377.

Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ, Knuth DE. 1996. On the Lambert W function. *Advances in Computational Mathematics* 5:329–359.

Cui Y, Zhu F. 2018. A new bivariate integer-valued GARCH model allowing for negative cross-correlation. *TEST* 27:428–452.

Cui Y, Li Q, Zhu F. 2020. Flexible bivariate Poisson integer-valued GARCH model. *Annals of the Institute of Statistical Mathematics* 72:1449–1477.

Darolles S, Le Fol G, Lu Y, Sun R. 2019. Bivariate integer-autoregressive process with an application to mutual fund flows. *Journal of Multivariate Analysis* 173:181–203.

Davis RA, Holan SH, Lund RB, Ravishanker N. 2015. *Handbook of Discrete-Valued Time Series* CRC Press, Boca Raton, Florida, USA.

Davis RA, Liu H. 2016. Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* 26:1673–1707.

Dunn P, Smyth GK. 1996. Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics* 5:236–244.

Ferland R, Latour A, Oraichi D. 2006. Integer-valued GARCH processes. *Journal of Time Series Analysis* 27:923–942.

Fokianos K. 2011. Some recent progress in count time series. *Statistics* 45:49–58.

Fokianos K, Tjøstheim D. 2012. Nonlinear Poisson autoregression. *Annals of the Institute of Statistical Mathematics* 64:1205–1225.

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

Fokianos K, Stove B, Tjøstheim D, Doukhan P. 2020. Multivariate count autoregression. *Bernoulli* 26:471–499.

Gonçalves E, Mendes-Lopes N, Silva F. 2015. Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis* 36:503–527.

Granger CWJ. 1969. Investigating causal relations by econometric models and cross spectral methods. *Econometrica* 37:424–438.
Guo S, Ling S, Zhu K. 2014. Factor double autoregressive models with application to simultaneous causality testing. *Journal of Statistical Planning and Inference* 148:82–94.

Hall P, Heyde CC. 1980. *Martingale Limit Theory and its Applications* Academic Press, New York.

Heinen A. 2003. *Modelling time series count data: an autoregressive conditional Poisson model*. In *CORE Discussion Paper 2003/62*, Université Catholique de, Louvain.

Hofer V, Leitner J. 2012. A bivariate Sarmanov regression model for count data with generalised Poisson marginals. *Journal of Applied Statistics* 39:2599–2617.

Jørgensen B, Lundbye-Christensen S, Song P, Sun L. 1999. A state space model for multivariate longitudinal count data. *Biometrika* 86:169–181.

Jung RC, Liesenfeld R, Richard JF. 2011. Dynamic factor models for multivariate count data: An application to stock-market trading activity. *Journal of Business and Economic Statistics* 29:73–85.

Karlis D, Pedeli X. 2013. Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics – Theory and Methods* 42:723–740.

Karlis D. 2016. *Models for multivariate count time series*. In *Handbook of Discrete-Valued Time Series*, CRC Press, Boca Raton; 407–424.

Kedem B, Fokianos K. 2002. *Regression Models for Time Series Analysis* Wiley, New Jersey, USA.

Latour A. 1997. The multivariate GINAR(p) process. *Advances in Applied Probability* 29:228–248.

Lee S, Lee Y, Chen CWS. 2016. Parameter change test for zero-inflated generalized Poisson autoregressive models. *Statistics* 50:540–557.

Lee Y, Lee S, Tjøstheim D. 2018. Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. *TEST* 27:52–69.

Lee MLT. 1996. Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics - Theory and Methods* 25:1207–1222.

Liu H. 2012. Some models for time series of counts. Ph.D. thesis, Columbia University.

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

Pedeli X, Karlis D. 2011. A bivariate INAR(1) process with application. *Statistical Modelling* 11:325–349.

Pedeli X, Karlis D. 2013a. On composite likelihood estimation of a multivariate INAR(1) model. *Journal of Time Series Analysis* 34:206–220.

Pedeli X, Karlis D. 2013b. Some properties of multivariate INAR(1) processes. *Computational Statistics and Data Analysis* 67:213–225.

R Core Team. 2020. *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ Accessed November, 2020.

Sarmanov O. 1966. Generalized normal correlation and two-dimensional Frechet classes. *Doklady Akademii Nauk* 168:32–35 (in Russian).

Scotto MG, Weiß CH, Silva ME, Pereira I. 2014. Bivariate binomial autoregressive models. *Journal of Multivariate Analysis* 125:233–251.

Silva RB, Barreto-Souza W. 2019. Flexible and robust mixed Poisson INGARCH models. *Journal of Time Series Analysis* 40:788–814.

Stan Development Team. 2020. *Stan Modeling Language Users Guide and Reference Manual* 2.25. https://mc-stan.org/. Accessed November, 2020.

Steutel F, van Harn K. 1979. Discrete analogues of self–decomposability and stability. *Annals of Probability* 7:893–899.

Wang F, Wang H. 2018. Modelling non-stationary multivariate time series of counts via common factors. *Journal of the Royal Statistical Society - Series B* 80:769–791.

Weiß C. 2018. *An Introduction to Discrete-Valued Time Series* Wiley, Chichester, UK.

Zhu F. 2010. A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis* 32:54–67.