Semiparametric time series models driven by latent factor

Gisele de Oliveira Maia\textsuperscript{a}, Wagner Barreto-Souza\textsuperscript{b,a}, Fernando de Souza Bastos\textsuperscript{c} and Hernando Ombao\textsuperscript{b}

\textsuperscript{a}Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
\textsuperscript{b}Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
\textsuperscript{c}Instituto de Ciências Exatas e Tecnológicas, Universidade Federal de Viçosa, Florestal, Brazil

\textbf{ARTICLE INFO}

\textbf{Keywords:} Bounded time series, Gaussian process, Regression analysis, Shifted gamma process, Quasi-likelihood estimation

\textbf{ABSTRACT}

We introduce a class of semiparametric time series models (SemiParTS) driven by a latent factor process. The proposed SemiParTS class is flexible because, given the latent process, only the conditional mean and variance of the time series are specified. These are the primary features of SemiParTS: (i) no parametric form is assumed for the conditional distribution of the time series given the latent process; (ii) it is suitable for a wide class of data: non-negative, count, bounded, binary and real-valued time series; (iii) it does not constrain the dispersion parameter to be known. The quasi-likelihood inference is employed for estimating the parameters in the mean function. Here, we derive explicit expressions for the marginal moments and for the autocorrelation function of the time series process so that a method of moments can be employed for estimating the dispersion parameter and also the parameters related to the latent process. Simulated results aiming to check the proposed estimation procedure are presented. Forecasting procedures are proposed and evaluated in simulated and real data. Analyses of the number of admissions in a hospital due to asthma and total insolation time series illustrate the potential for practical situations of the proposed models.

1. Introduction

In Cox (1981) two classes of models for time-dependent data are described: observation-driven and parameter-driven models. Let \( \{Y_t\}_{t \in \mathbb{N}} \) denote a time series along this section. In the observation-driven approach, some conditional distribution for \( Y_t \) given \( \mathcal{F}_t = \sigma\{Y_{t-1}, Y_{t-2}, \ldots\} \) is assumed. These types of processes are described in Zeger and Qaqish (1988), Benjamin et al. (2003), Davis et al. (2003b), Rocha and Cribari-Neto (2009), Davis and Liu (2016) and Ombao et al. (2018), among others. The focus of this present paper is on time series driven by a latent process, which is one example of a parameter-driven model. Here, \( \{Y_t\}_{t \in \mathbb{N}} \) given a latent process \( \{\alpha_t\}_{t \in \mathbb{N}} \) is assumed to be conditionally independent but not identically distributed; a regression
structure is considered for modeling the mean of the process.

Pioneering work is due to Zeger (1988) which proposed a semiparametric count time series model so that, given the latent process, only the two first moments of the conditional distribution of the counts are specified. Estimation of the parameters related to the mean for this semiparametric model is performed through a quasi-likelihood function and a method of moments is considered for estimating parameters related to the latent process. In our current work, we develop a more general semiparametric time series model that can accommodate outcomes beyond just counts. Here, we will consider various types of outcomes including bounded, binary, positive continuous, and real-valued time series.

In Davis et al. (2000), a Poisson count time series model was proposed which is driven by a Gaussian latent factor. The focus of that paper was on checking the presence of such a latent process in practical situations. The regression coefficients are estimated based on a generalized linear model (GLM) approach. Moreover, conditions for consistency and asymptotic normality of the GLM estimators are derived along with an explicit form for the asymptotic covariance of these estimators. Without imposing a full distribution specification for the time series, a different strategy is needed to estimate parameters as we will propose here based on a quasi-likelihood approach. Another important work is due to Davis and Wu (2009), where a one-parameter exponential family is considered for the response time series given the latent process and thus extending the paper by Davis et al. (2000). Here, the focus is on the negative binomial case (with known dispersion parameter and so belonging to this family) and derived asymptotic properties of the GLM estimators. Estimation of the parameters related to the latent factor used some version of the ordinary least squares method.

In this paper, we introduce a class of semiparametric time series (SemiParTS) models that are driven by a latent factor process where estimation and inference will be conducted using the quasi-likelihood approach. One advantage of the SemiParTS model is that we only need to specify the conditional mean and variance of the time series (given the latent process). The proposed SemiParTS model has three remarkable features: (i) it does not impose any parametric form for the conditional distribution of the time series given the latent process; (ii) it can model a broad class of outcomes including non-negative, count, bounded, binary and real-valued time series; (iii) it does not assume that the dispersion parameter is known. Thus, the proposed SemiParTS model is general and includes, as special cases, the popular models in Zeger (1988), Davis et al. (2000) and Davis and Wu (2009).

As a final note, the quasi-likelihood approach for estimation and inference has been used in time series analysis, see for example Zeger and Qaqish (1988), Heyde (1997), Berkes and Horváth (2003), Francq and Zakoïan (2004), Straumann and Mikosch (2006), Christou and Fokianos (2014) and Christou and Fokianos (2015). The models proposed in these papers belong to the observation-driven class. However, SemiParTS goes in a different direction because this falls under the class of parameter-driven models and has the advantage of applying to various types of time series data.

The remainder of the paper is organized in the following manner. In Section 2 we introduce our class of semiparametric time series models and develop specific applications for the non-negative continuous, count, bounded, binary and $\mathbb{R}$-valued time series. Further, we derive marginal moments and the autocorrelation function of the proposed models, which will be utilized in subsequent steps for estimating the model parameters. Section 3 is devoted to the quasi-likelihood estimation for the mean parameters combined with the method of moments for estimating the dispersion and
parameters related to the latent process. We also develop a bootstrap procedure to obtain the standard errors of the parameter estimates. Monte Carlo simulations are reported in Section 4 to check the finite-sample behavior of the proposed estimators. Further, forecasting procedures based on our class of semiparametric time series models are proposed and discussed. To demonstrate practical utility, we analyze two datasets, namely the number of asthma cases and insolation (amount total solar energy at a particular location) time series in Section 5 using the SemiParTS models. Concluding remarks are addressed in Section 6.

2. Model definition

In this section, we define our proposed class of semiparametric time series (SemiParTS) models and develop basic properties that will be essential for estimating parameters. Broadly, under the proposed SemiParTS model, the observed time series is driven by a latent process, and estimation is conducted via a quasi-likelihood approach (conditioned on the latent process). This approach is flexible because it requires specifying only the first two moments of the conditional distribution. Our starting point is inspired by Zeger (1988) but the SemiParTS model is more general because it offers a broad family of link mean and variance functions whereas Zeger (1988) constrains these to be the identity function.

**Definition 2.1.** Let \( \{ \alpha_t \}_{t \in \mathbb{N}} \) be a latent stationary strongly mixing process. Our proposed class of semiparametric time series (SemiParTS) processes, say \( \{ Y_t \}_{t \in \mathbb{N}} \), is defined by assuming conditional independence of \( \{ Y_t \}_{t \in \mathbb{N}} \) given \( \{ \alpha_t \}_{t \in \mathbb{N}} \) and by the following specifications:

\[
\begin{align*}
g(\tilde{\mu}_t) &= x_{nt}^T \beta + \alpha_t, \\
E(Y_t|\alpha_t) &= \tilde{\mu}_t, \\
Var(Y_t|\alpha_t) &= \phi V(\tilde{\mu}_t),
\end{align*}
\]

where \( \beta = (\beta_1, ..., \beta_q)^T \) is the vector of regression coefficients, \( x_{nt} \) is an observable covariate vector (which can vary according on the sample size) with dimension \( q \times 1 \); the function \( g(\cdot) \) is an invertible link function; \( V(\cdot) \) is a variance function and \( \phi > 0 \) is the dispersion parameter.

**Remark 2.1.** Note that although \( \{ \alpha_t \}_{t \in \mathbb{N}} \) is assumed to be a stationary process, our SemiParTS processes \( \{ Y_t \}_{t \in \mathbb{N}} \) are non-stationary (both unconditionally and conditionally on \( \{ \alpha_t \} \)) since we are allowing for the inclusion of covariates.

**Remark 2.2.** A remarkable feature of the SemiParTS models is its broad range of applicability. Unlike many other existing models, it can accommodate many types of time series data. Due to the flexibility of the mean link and variance functions \( g \) and \( V \), the SemiParTS model can be used for counts, positive continuous, bounded, binary, and \( \mathbb{R} \)-valued time series.

**Remark 2.3.** The assumption of stationarity and strong mixing on the latent process is important to obtain consistency and asymptotic normality of the generalized linear model estimators as discussed with details by Davis and Wu (2009). The latent processes considered in this present paper satisfy these properties.
In Davis et al. (2000) and Davis and Wu (2009), parameters related to the mean are estimated by fitting a generalized linear model (GLM) but the latent process is ignored. Under some conditions involving the latent process and the covariates (see Assumptions 1 and 2 and convergences (4)-(8) from that paper), they proved that the GLM estimators are consistent and asymptotically normally distributed. Here, the information matrix based on the GLM approach does not provide the correct standard errors of the parameter estimates. The authors derived the correct information matrix for those models and proposed an alternative to performing Monte Carlo simulation to obtain standard errors of the estimates. We will adopt this Monte Carlo simulation strategy for the SemiParTS models with some adjustments since we need a time series generator from the estimated model and we do not have an explicit conditional distribution for the time series given the latent process. This is discussed in Section 3 and empirically illustrated in Section 5.

Below we state the conditions required to establish consistency for estimating the parameter vector \( \beta \), as discussed by Zeger (1988), Davis et al. (2000) and Davis and Wu (2009).

**Assumption 2.4.** Let \( \{Y_t\}_{t \in \mathbb{N}} \) be as in Definition 2.1. We assume that the latent process \( \{\alpha_t\}_{t \in \mathbb{N}} \) is such that

\[
E(Y_t) = E\left(h(x_{nt}^\top \beta + \alpha_t)\right) = h(x_{nt}^\top \beta)
\]

for all \( t \in \mathbb{N} \), where \( h(\cdot) \) is the inverse of the link function \( g(\cdot) \).

**Remark 2.5.** Further details on Assumption 2.4 are given in Subsection 3.1 of Zeger (1988).

We now present the latent processes that are an integral part of the proposed SemiParTS model. Following Zeger (1988), Davis et al. (2000) and Davis and Wu (2009), we assume a latent Gaussian AR(1) model for the count, positive continuous and real-valued cases. More explicitly, we have that

\[
\alpha_t = c + \rho \alpha_{t-1} + \eta_t, \quad t \in \mathbb{N},
\]

where \( \{\eta_t\}_{t \in \mathbb{N}} \) i.i.d. \( \sim N(0, (1 - \rho^2)\sigma^2) \), \( |\rho| < 1, \sigma^2 > 0 \) and \( c \in \mathbb{R} \) is an intercept chosen according Assumption 2.4. In this case, the process is well-known to be stationary and strongly mixing with marginals \( \alpha_t \sim N\left(\frac{c}{1-\rho}, \sigma^2\right) \), for all \( t \in \mathbb{N} \).

**Remark 2.6.** Alzahrani et al. (2018) discussed a latent AR(p) Gaussian process for the count time series model by Zeger (1988) and observed problems in identifying the order of the latent process so supporting the use of a first-order model \((p = 1)\).

We now discuss another latent process that will be used for the bounded and binary cases. This will be based on the first-order gamma autoregression (with mean 1) proposed by Sim (1990). This will permit us to obtain explicit expressions for the moments and control the range of the conditional mean, which will be complicated based on the Gaussian process (1). We say that a sequence \( \{Z_t\}_{t \in \mathbb{N}} \) follows a first-order gamma autoregression (denoted by GAR(1)) if satisfies

\[
Z_t = \kappa \odot Z_{t-1} + \eta_t, \quad t \in \mathbb{N}, \quad Z_0 \sim G(1/\sigma^2, 1/\sigma^2),
\]
Semiparametric time series models

where the operator $\odot$ is defined by $\kappa \odot Z_{t-1} \overset{d}{=} \sum_{i=1}^{N_{t-1}} W_i$, with $N_{t-1} \mid Z_{t-1} = z \sim \text{Poisson}(\kappa \rho z)$. \{W_i\}^\infty_{i=1} \overset{iid}{\sim} \text{Exponential}(\kappa)$ and \{\eta_i\}^\infty_{i=1} \overset{iid}{\sim} \mathcal{G}(\sigma^2, \kappa)$ are assumed to be independent and $\kappa = \frac{1}{\sigma^2(1-\rho)}$, for $\sigma^2 > 0$ and $\rho \in (0, 1)$. Here, $\mathcal{G}(\sigma^2, \kappa)$ denotes a gamma distribution with shape and scale parameters $\sigma^2 > 0$ and $\kappa > 0$, respectively.

The GAR process depends on the parameters $\sigma^2$ and $\rho$. The parameter $\rho$ controls the dependence of this process and in fact $\text{corr}(Z_{t+k}, Z_t) = \rho^k$ for $t, k \in \mathbb{N}$. The marginals of this model are gamma distributed with mean 1 and variance $\sigma^2$, therefore the model is stationary; see Sim (1990). The strong mixing property of this process was established recently by Barreto-Souza and Ombao (2019). Therefore, we define our latent process $\{\alpha_i\}_{i \in \mathbb{N}}$ in the bounded and binary cases by

$$\alpha_t = Z_t + \log E \left( \exp(-Z_t) \right) = Z_t - \frac{1}{\sigma^2} \log(1 + \sigma^2), \quad t \in \mathbb{N}. \quad (2)$$

The shifted gamma process $\{\alpha_i\}_{i \in \mathbb{N}}$ above is necessary to satisfy Assumption 2.4. This will be clear when we deal with the bounded/binary case in Subsection 2.3. A similar approach was considered by Davis and Wu (2009) for dealing with binary data. In that paper, the authors assumed a kind of shifted exponential process. Based on our approach, the shifted term is very simple in contrast with the term of the exponential process in Davis and Wu (2009) (see Experiment 2, page 743).

We now define our semiparametric time series models for various types of data and with specific link and variance functions. Here, we also derive the marginal moments and the autocorrelation function.

2.1. Non-negative-valued time series

Let $\{Y_t\}_{t \in \mathbb{N}}$ be a time series with support $S \subset [0, \infty)$. We will derive a model for this wide class of time series data which includes count and positive continuous time series. We propose a logarithm link function and a polynomial variance function $V(\mu) = \mu^p$, with $\mu > 0$ and $p > 0$. In this way, we define the model for non-negative time series as

$$\log \tilde{\mu}_t = x_{n,t}^T \beta + \alpha_t,$$
$$E(Y_t \mid \alpha_t) = \tilde{\mu}_t = \exp(x_{n,t}^\top \beta + \alpha_t) = \exp(x_{n,t}^\top \beta) \epsilon_t,$$
$$\text{Var}(Y_t \mid \alpha_t) = \phi V(\tilde{\mu}_t) = \phi \tilde{\mu}_t^p,$$

where $\phi > 0$, $\{\alpha_t\}_{t \in \mathbb{N}}$ is Gaussian AR(1) process defined in Equation (1) and $\epsilon_t = \exp\{\alpha_t\}$, for $t \in \mathbb{N}$. In order to ensure Assumption 2.4 is satisfied, we need to take $E(\epsilon_t) = 1$. We have that $E(\epsilon_t) = \exp \left\{ E(\alpha_t) + 0.5 \text{Var}(\alpha_t) \right\} = 1$ implies $c = -\sigma^2(1-\rho)/2$. Consequently, $\alpha_t \sim N \left( -\sigma^2/2, \sigma^2 \right)$, for all $t \in \mathbb{N}$. Under this setting, the sequence $\{\epsilon_t\}_{t \in \mathbb{N}}$ is a strictly stationary log-normal autoregressive model with mean 1 and variance equal to $\sigma^2 = \exp(\sigma^2) - 1$. The autocovariance and autocorrelation functions are derived, respectively,

$$\gamma_\epsilon(k) \equiv \text{cov}(\epsilon_{t+k}, \epsilon_t) = \exp(\gamma(k)) - 1$$

and

$$\rho_\epsilon(k) \equiv \text{corr}(\epsilon_{t+k}, \epsilon_t) = \frac{\exp(\rho(k)) - 1}{\exp(\sigma^2) - 1},$$
where \( \gamma(k) = \sigma^2 \phi^k \) and \( \rho(k) = \phi^k \) are the autocovariance and autocorrelation functions at lag \( k \in \mathbb{N} \) of the process \( \{ \alpha_t \}_{t \in \mathbb{N}} \). The usage of this log-normal process on the Poisson regression is discussed in Davis et al. (2000). Under the above specifications and using basic properties of conditional expectation, we obtain that the first two marginal cumulants are necessary to ensure Assumption 2.4 is in force. In this case, \( \{ E(Y_t | \alpha_t) \}_{t \in \mathbb{N}} \) is discussed in Davis et al. (2000). Under the above specifications and using basic properties of conditional expectation, we obtain that \( \mu_t \equiv E(Y_t) = E(E(Y_t | \alpha_t)) = \exp(x_{n1}^\top \beta) E(e_t) \) and \( \text{Var}(Y_t) = E \left( \text{Var}(Y_t | \alpha_t) \right) + \text{Var} \left( E(Y_t | \alpha_t) \right) = \phi E \left( \tilde{\mu}_t^2 \right) + \mu_t^2 \text{Var}(e_t) \). Hence, it follows that

\[
\mu_t = \exp(x_{n1}^\top \beta)
\]

and

\[
\text{Var}(Y_t) = \phi \mu_t^p \left( \sigma_e^2 + 1 \right)^{p-1} + \mu_t^2 \sigma_e^2. \tag{3}
\]

The autocovariance function of \( \{ Y_t \}_{t \in \mathbb{N}} \) is given by \( \text{Cov}(Y_{t+k}, Y_t) = \text{Cov} \left( E(Y_{t+k} | \alpha_{t+k}), E(Y_t | \alpha_t) \right) \), for \( k > 0 \). Therefore,

\[
\text{Cov}(Y_{t+k}, Y_t) = \mu_{t+k} \mu_t \gamma_e(k) \tag{4}
\]

and the autocorrelation function is given by

\[
\text{Corr}(Y_{t+k}, Y_t) = \frac{\mu_{t+k} \mu_t \gamma_e(k)}{\sqrt{\phi \mu_{t+k}^p \left( \sigma_e^2 + 1 \right)^{p-1} + \mu_{t+k}^2 \sigma_e^2} \sqrt{\phi \mu_t^p \left( \sigma_e^2 + 1 \right)^{p-1} + \mu_t^2 \sigma_e^2}}
\]

\[
= \frac{\mu_{t+k} \mu_t \gamma_e(k)}{\sqrt{\phi \sigma_e^{-2} \mu_{t+k}^p \left( \sigma_e^2 + 1 \right)^{p-1} + 1} \sqrt{\phi \sigma_e^{-2} \mu_t^p \left( \sigma_e^2 + 1 \right)^{p-1} + 1}}.
\]

**Remark 2.7.** The model by Zeger (1988) is a particular case of the class discussed in this subsection by taking \( V(\mu) = \mu \) \((p = 1)\) and \( \phi = 1 \).

### 2.2. \( \mathbb{R} \)-valued time series

Here we assume that the support of the sequence \( \{ Y_t \}_{t \in \mathbb{N}} \) is \( \mathbb{R} \). We set an identity link function and variance function \( V(\tilde{\mu}_t) = 1 \), so mimicking the first two moments of a normal distribution. More specifically, we assume that

\[
\tilde{\mu}_t = x_{n1}^\top \beta + \alpha_t,
\]

\[
E(Y_t | \alpha_t) = \tilde{\mu}_t = x_{n1}^\top \beta + \alpha_t,
\]

\[
\text{Var}(Y_t | \alpha_t) = \phi V(\tilde{\mu}_t) = \phi,
\]

where \( \{ \alpha_t \}_{t \in \mathbb{N}} \) is the Gaussian AR(1) process given in Equation (1) with \( c = 0 \), being this last condition necessary to ensure Assumption 2.4 is in force. In this case, \( \alpha_t \sim N(0, \sigma^2) \) for \( t \in \mathbb{N} \). By using basic properties of conditional expectation, we obtain that the first two marginal cumulants of \( Y_t \) are derived to be

\[
\mu_t \equiv E(Y_t) = E(E(Y_t | \alpha_t)) = E(x_{n1}^\top \beta + \alpha_t) = x_{n1}^\top \beta
\]

and

\[
\text{Var}(Y_t) = E \left( \text{Var}(Y_t | \alpha_t) \right) + \text{Var} \left( E(Y_t | \alpha_t) \right) = \phi + \text{Var} \left( \mu_t + \alpha_t \right) = \phi + \sigma^2. \tag{5}
\]
The marginal variance of \( x \)

\[
\text{Var}(Y_{t+k}, Y_t) = \text{Cov} \left( E(Y_{t+k} | \alpha_{t+k}), E(Y_t | \alpha_t) \right) + 0 = \text{Cov} \left( \mu_{t+k} + \alpha_{t+k}, \mu_t + \alpha_t \right) = \sigma^2 \rho(k) = \sigma^2 \rho^k
\]

and

\[
\text{Corr}(Y_{t+k}, Y_t) = \frac{\text{Cov}(Y_{t+k}, Y_t)}{\sqrt{\text{Var}(Y_{t+k}) \text{Var}(Y_t)}} = \frac{\rho(k) \sigma^2}{\sqrt{(\phi + \sigma^2)^2}} = \frac{\rho(k)}{\phi / \sigma^2 + 1} = \frac{\rho^k}{\phi / \sigma^2 + 1}.
\]

A Gaussian time series model driven by a latent AR(1) process was considered by \citet{Davis2009} (see Example 3). In that model, the variance of the time series, given the latent process, is assumed to be known. Thus, this model is constrained to belong to the one-parameter exponential family. In contrast, our proposed SemiParTS model does not impose this assumption on the distribution of the time series. Thus, the proposed SemiParTS model is more data-adaptive because the variance is estimated from the data.

### 2.3. Bounded and binary-valued time series

Let \( \{Y_t\}_{t \in \mathbb{N}} \) be a process having one of the following supports: \((0, 1)\), \(\{0, 1\}\) or \(\{0, 1, \ldots, m\}\), with \(m \in \mathbb{Z}^+\). Therefore, in this setting, the goal is to develop a model for proportions/rates (bounded continuous), binary and binomial time series data. We specify the link function to be \(g(z) = -\log z\) and the variance function to be \(V(z) = z(1 - z)\), for \(z \in (0, 1)\).

Consider \(\{\alpha_t\}_{t \in \mathbb{N}}\) be the shifted gamma process given in (2). Our model here is defined by the following equations:

\[
\begin{align*}
- \log \tilde{\mu}_t &= x_{nt}^\top \beta + \alpha_t, \\
E(Y_t | \alpha_t) &= \tilde{\mu}_t = \exp(-x_{nt}^\top \beta) \epsilon_t, \\
\text{Var}(Y_t | \alpha_t) &= \phi V(\tilde{\mu}_t) = \phi \tilde{\mu}_t(1 - \tilde{\mu}_t),
\end{align*}
\]

where \(\epsilon_t = \exp(-\alpha_t)\) and \(\phi = 1\) and \(\phi = m\) for the binary and binomial cases, respectively. For the bounded continuous case, we have that \(0 < \phi < 1\). Here, the vector \(\beta\) and \(\sigma^2\) are such that \(x_{nt}^\top \beta > \sigma^{-2} \log(1 + \sigma^2)\), since that \(\tilde{\mu}_t \in (0, 1)\) for all \(t \in \mathbb{N}\).

**Remark 2.8.** The model discussed here can be used for time series with support on any interval \((a, b)\) \((a < b)\) which, by appropriate shifting and rescaling, can be transformed into \((0, 1)\) for simplicity. So, without loss of generality, we consider the unit interval \((0, 1)\).

We now discuss the shifted term in the gamma autoregressive process. This is necessary to ensure that Assumption 2.4 is satisfied. The marginal mean of \(Y_t\) is now derived to be

\[
\mu_t \equiv E(Y_t) = E(E(Y_t | \alpha_t)) = \exp(-x_{nt}^\top \beta) E(\epsilon_t) = \exp(-x_{nt}^\top \beta),
\]

since \(E(\epsilon_t) = E(\exp(-\alpha_t)) = E(\exp(-Z_t)) / E(\exp(-Z_t)) = 1\). After some algebra, we obtain that the marginal variance of \(Y_t\) is

\[
\text{Var}(Y_t) = \phi \mu_t + \mu_t^2 \left\{ (1 - \phi) \left( \frac{(1 + \sigma^2)^2}{1 + 2\sigma^2} \right)^{1/\sigma^2} - 1 \right\}.
\]

The autocovariance of the process \(\{Y_t\}_{t \in \mathbb{N}}\) is

\[
\text{Cov}(Y_{t+k}, Y_t) = \text{Cov} \left( E(Y_{t+k} | \alpha_{t+k}), E(Y_t | \alpha_t) \right) + 0 = \mu_{t+k} \mu_t (1 + \sigma^2)^{2/\sigma^2} \text{Cov} \left( \exp(-Z_{t+k}), \exp(-Z_t) \right),
\]
for $k > 0$. From Eq. (2.6) from Sim (1990), we have an explicit expression for the joint Laplace function of $(Z_{t+k}, Z_t)$. Using that expression, we obtain that

$$
\text{Cov}(Y_{t+k}, Y_t) = \mu_{t+k} \mu_t \left\{ \frac{(1 + \sigma^2)^2}{1 + 2\sigma^2 + (\sigma^2)^2(1 - \rho^k)} \right\}^{1/\sigma^2} - 1. \quad (8)
$$

An expression for the autocorrelation function is immediately obtained by using (7) and (8). The time series model for binary data proposed here is an alternative to the model discussed by Davis and Wu (2009) since we are using a different latent process. We again call the attention that the shifted term considered here is simpler compared to Davis and Wu (2009) which involves multiplication of an infinite number of terms. Further, our proposed methodology enables us to deal with continuous bounded time series data.

3. Quasi-likelihood approach and method of moments

In this section we discuss estimation of the parameters by combining quasi-likelihood approach and method of moments. Let $Y_1, \ldots, Y_n$ be a random trajectory of a time series process as in Definition 2 and $\theta = (\beta, \phi, \sigma^2, \rho)^T$ be the parameter vector. The parameter vector $\beta$ will be estimated using the quasi-likelihood method (which is a well-established general statistical procedure developed in Wedderburn (1974)). The log-quasi-likelihood function is given by

$$
Q(\beta) = \sum_{t=1}^{n} Q(y_t; \mu_t),
$$

where $Q(y; \mu) = \int_{\mu}^{y} \frac{y - \omega}{V(\omega)} d\omega$ and the marginal mean $\mu_t$ depends on the linear predictor $x_{nt}^T \beta$, for $t = 1, \ldots, n$. Depending on the choices for the variance function, the quasi-likelihood models have corresponding cases in the generalized linear models which will be discussed in the following subsections. The quasi-likelihood estimator for $\beta$ is given by

$$
\hat{\beta} = \text{argmax}_\beta Q(\beta).
$$

To estimate the remaining (nuisance) parameters, we use the population moments and autocovariance function obtained in the previous section and then propose a variant of the method of moments estimators. This strategy has been used for instance by Zeger (1988), Davis et al. (2000), Davis and Wu (2009) and Christou and Fokianos (2014).

We obtain the standard errors for the quasi-likelihood estimates of $\beta$ through a Monte Carlo simulation. Note that we do not assume a specific conditional distribution for the time series given by the latent process. This should not be a problem because in each Monte Carlo replicate, it will be sufficient to assume a specific parametric model as a time series generator (for the Monte Carlo procedure) having the same mean structure as our semiparametric model. This works well even under an incorrect specification of the variance function as argued by Zeger (1988). On the other hand, we are also interested in obtaining the standard errors for the nuisance parameter estimates. Thus, a correct specification of the variance function is required for this purpose. This procedure will be illustrated in the applications to real-time series in Section 5.

In the following subsections, we discuss the estimation of the parameters with more details for the non-negative, real-valued, and bounded/binary time series models.
3.1. Non-negative-valued time series

Consider the non-negative time series model discussed in Subsection 2.1. Then, we have the variance function given by $V(\mu) = \mu^p$ for $\mu, p > 0$ and the marginal mean of $Y_t$ given by $\mu_t = \exp(x_{nt}^T \hat{\beta})$, for $t = 1, \ldots, n$. For $p \neq 1, 2$, we have that

$$Q(y; \mu) = \int_y^{\mu} \frac{y - \omega}{\omega^p} d\omega = \frac{y}{1 - p} (\mu^{-p+1} - y^{-p+1}) - \frac{1}{2 - p} (\mu^{-p+2} - y^{-p+2}).$$

For $p = 1$ and $p = 2$, we obtain respectively $Q(y; \mu) = y(\log \mu - \log y) + y - \mu$ and $Q(y; \mu) = \log(y/\mu) - y/\mu + 1$. The quasi-likelihood models with $p = 1$, $p = 2$ and $p = 3$ have the Poisson, gamma and inverse-Gaussian generalized linear models as corresponding cases.

To estimate $\phi$, $\sigma^2$ and $\rho$ through method of moments, we use the expressions of $\text{Var}(Y_t)$ and $\text{Cov}(Y_{t+k}, Y_t)$ (for $k = 1, 2$) given respectively in (3) and (4) and thus

$$\hat{\phi} = \frac{\sum_{t=1}^{n}(Y_t - \hat{\mu}_t)^2 - (\hat{\sigma}^2 - 1) \sum_{t=1}^{n} \hat{\mu}_t^2}{\hat{\sigma}^2 \rho(2)^{p-1}/2 \sum_{t=1}^{n} \hat{\mu}_t^p}$$

and

$$\exp(\hat{\sigma}^2 \hat{\rho}^k) = \frac{\sum_{t=1}^{n-k}(Y_t - \hat{\mu}_t)(Y_{t+k} - \hat{\mu}_{t+k})}{\sum_{t=1}^{n-k} \hat{\mu}_t \hat{\mu}_{t+k}} + 1,$$

for $k = 1, 2$, where $\hat{\mu}_t = \exp(x_{nt}^T \hat{\beta})$ for $t = 1, \ldots, n$ with $\hat{\beta}$ being the quasi-likelihood estimator of $\beta$.

Define $M_k \equiv \log \left( \frac{\sum_{t=1}^{n-k}(Y_t - \hat{\mu}_t)(Y_{t+k} - \hat{\mu}_{t+k})}{\sum_{t=1}^{n-k} \hat{\mu}_t \hat{\mu}_{t+k} + 1} \right)$, for $k = 1, 2$. After some algebra, we obtain an explicit solution from the Equations given in (10), that is $\hat{\rho} = M_2/M_1$ and $\hat{\sigma}^2 = M_2^2/M_1$. Consequently, we also obtain an explicit estimator for $\phi$ given in (9).

3.2. $\mathbb{R}$-valued time series

For real-valued time series, we will assume that $V(\mu) = 1$. As discussed in Subsection 2.2, we choose the latent factor having null mean so that the marginal mean of $Y_t$ is $\mu_t = x_{nt}^T \beta$. The $Q$-function is this case is given by

$$Q(y; \mu) = \int_y^{\mu} (y - \omega) d\omega = \frac{(y - \mu)^2}{2},$$

for $y, \mu \in \mathbb{R}$. By maximizing the logarithm of the quasi-likelihood function, we obtain the estimators for the regression coefficients, say $\hat{\beta}$. From expressions (5) and (6), we obtain the following method of moments estimators for $\phi$, $\rho$ and $\sigma^2$:

$$\hat{\phi} = \frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{\mu}_t)^2 - \hat{\sigma}^2,$$
Semiparametric time series models

\[ \hat{\rho} = \frac{\sum_{t=1}^{n-2} (Y_t - \hat{\mu}_t)(Y_{t+2} - \hat{\mu}_{t+2})}{\sum_{t=1}^{n-1} (Y_t - \hat{\mu}_t)(Y_{t+1} - \hat{\mu}_{t+1})} \]

and

\[ \hat{\sigma}^2 = \frac{\left(\sum_{t=1}^{n-1} (Y_t - \hat{\mu}_t)(Y_{t+1} - \hat{\mu}_{t+1})\right)^2}{n \sum_{t=1}^{n-2} (Y_t - \hat{\mu}_t)(Y_{t+2} - \hat{\mu}_{t+2})}. \]

**Remark 3.1.** As can be noted, the estimators proposed here for the parameters \( \phi, \rho, \) and \( \sigma^2 \) assume simple forms as the conditional least squares estimators for the AR(1) Gaussian model.

### 3.3. Bounded and binary-valued time series

In this case, for \( y \in (0, 1) \), the \( Q \)-function assumes the form

\[ Q(y; \mu) = \int_{y}^{\mu} \frac{y - \omega}{V(\omega)} d\omega \]

\[ = y \left\{ \log(\mu(1 - \mu)) - \log \left( \frac{y}{1 - y} \right) \right\} + \log(1 - \mu) - \log(1 - y), \]

For \( y = 0 \) and \( y = 1 \) we obtain respectively \( Q(0, \mu) = \log(1 - \mu) \) and \( Q(1, \mu) = \log \mu \). Let \( \hat{\mu}_t = \exp(-x^T_{nt} \hat{\beta}) \), for \( t = 1, \ldots, n \), with \( \hat{\beta} \) denoting the quasi-likelihood estimator obtained based on the above \( Q \)-function. Assume \( \phi \) is a unknown parameter to be estimated. From Expressions (7) and (8), we obtain that the method of moments estimator of \( \phi \) is

\[ \hat{\phi} = \frac{\sum_{t=1}^{n} (Y_t - \hat{\mu}_t)^2 - (\sigma^2 \sum_{t=1}^{n} \hat{\mu}_t^2)}{\sum_{t=1}^{n} \hat{\mu}_t - \sigma^2 \sum_{t=1}^{n} \hat{\mu}_t^2} \]

and the estimators for \( \sigma^2 \) and \( \rho \) are obtained by solving the system of non-linear equations

\[ v(\hat{\sigma}, \hat{\rho}^k) = \frac{\sum_{t=1}^{n-k} (Y_t - \hat{\mu}_t)(Y_{t+k} - \hat{\mu}_{t+k})}{\sum_{t=1}^{n-k} \hat{\mu}_t \hat{\mu}_{t+k}} + 1, \quad \text{for } k = 1, 2, \quad (11) \]

where \( w(x) = \left( \frac{(1 + x)^2}{1 + 2x} \right)^{1/x} \) and \( v(x, y) = \left( \frac{(1 + x)^2}{1 + 2x + x^2(1 - y)} \right)^{1/x} \), for \( x > 0 \) and \( y \in (0, 1) \).

Since there is not closed form for the method of moments estimators of \( \sigma^2 \) and \( \rho \), some numerical optimization is needed. For the case where \( \phi \) is known, as in the Bernoulli and binomial cases where \( \phi = 1 \) and \( \phi = m \), respectively, just use (11) to get estimators for \( \sigma^2 \) and \( \rho \).

We conclude this subsection by recommending the paper by Cavaliere and Xu (2014) for those interested in testing non-stationarity in bounded time series.
4. Simulated results and forecasting

4.1. In-sample estimation performance

We perform three simulation studies to evaluate the methodology presented for estimating the model parameters based on the quasi-likelihood approach combined with the method of moments. All the implementations in this paper were conducted through the R Core Team (2019) software. We here illustrate the positive continuous, real-valued, and bounded cases. For all cases considered in these simulated studies, we take 1000 Monte Carlo replicates and sample sizes \( n = 500, 1000, 2000 \).

For the first case, we consider the semiparametric time series (SemiParTS) model for positive continuous data defined in Subsection 2.1 driven by the Gaussian AR(1) process. More specifically, we take the variance function to be quadratic, \( V(t) = \mu^2 \), so mimicking the GLM gamma model. In this simulation, we set the covariate vector

\[
x_{nt} = \{1, \cos(2\pi t/12), \sin(2\pi t/12)\}, \quad t = 1, \ldots, n,
\]

with regression coefficients \( \beta = (5, -0.2, 0.4)^T \), \( \phi = 0.1 \), \( \sigma^2 = 0.5 \) and \( \rho = 0.6 \). For generating the simulated time series in each Monte Carlo replica, we assume a conditional gamma distribution (given the latent process) with mean \( \tilde{\mu}_t \) and variance \( \phi V(\tilde{\mu}_t) \), for \( t = 1, \ldots, n \). Figure 1 provides simulated trajectories of a positive continuous time series and the associated latent process under first scenario. Estimation of the parameters is performed as proposed in Subsection 3.1.

In Table 1, we present the empirical means and standard errors of the quasi-likelihood estimates of the \( \beta \)'s and the method of moments (MMs) estimates of \( \phi, \sigma^2 \) and \( \rho \) with their respective standard errors. We call the attention that MM estimators can produce estimates out of the parameter space. In these cases, the samples were discarded and a new Monte Carlo replica was considered. This is a well-known problem of this kind of estimator and it is attenuated when working with moderate or large sample sizes.

From Table 1, we observe that the quasi-likelihood estimators yielded almost unbiased estimates for \( \beta \) for all sample sizes considered. The MM estimators also provided satisfactory results for estimating \( \phi, \sigma^2 \), and \( \rho \). These results indicate generally good performance and consistency of the proposed estimators as the sample size increases.

![Simulated trajectories of a positive continuous time series and the associated latent process under the first scenario.](image)
Table 1
Empirical means and standard errors of the quasi-likelihood estimates of $\beta$ and method of moments estimates of $\phi$, $\sigma^2$ and $\rho$ based on the SemiParTS for positive continuous data.

| parameter | true value | $n = 500$           |       | $n = 1000$           |       | $n = 2000$           |       |
|-----------|------------|---------------------|------|---------------------|------|---------------------|------|
|           |            | mean | stand. err. | mean | stand. err. | mean | stand. err. |
| $\beta_0$ | 5          | 4.997 | 0.070       | 4.998 | 0.049       | 4.997 | 0.035     |
| $\beta_1$ | $-0.2$     | $-0.199$ | 0.076       | $-0.202$ | 0.054       | $-0.200$ | 0.037     |
| $\beta_2$ | 0.4        | 0.394 | 0.074       | 0.398 | 0.053       | 0.401 | 0.039     |
| $\phi$    | 0.1        | 0.131 | 0.089       | 0.115 | 0.071       | 0.107 | 0.059     |
| $\sigma^2$| 0.5        | 0.448 | 0.107       | 0.475 | 0.086       | 0.487 | 0.058     |
| $\rho$    | 0.6        | 0.626 | 0.101       | 0.615 | 0.075       | 0.603 | 0.102     |

Figure 2: Simulated trajectories of a $\mathbb{R}$-valued time series (to the left) and associated latent process (to the right) under the second scenario.

We now consider a second scenario involving real-valued time series with the semiparametric model given in Subsection 2.2 ($V(\mu) = 1$) based on the null mean Gaussian AR(1) process.

We assume the covariate vector

$$x_{nt} = \{1, t/n, \cos(2\pi t/6)\}$$

with $\beta = (0.1, 0.5, 0.7)^T$. We also set $\phi = 3$, $\sigma^2 = 1$ and $\rho = 0.5$. To generate the $\mathbb{R}$-valued time series, we take the conditional distribution of $Y_t$ given the latent process to be normal distributed, for $t = 1, \ldots, n$. Simulated trajectories of a time series and its associated latent process under this scenario are displayed in Figure 2.

The empirical means and standard errors of the model parameters based on the estimation procedure discussed in Subsection 3.2 are presented in Table 2. From these results, we can observe a good performance of the proposed estimators based on the quasi-likelihood approach combined with the method of moments for the considered real-valued time series.

Our last scenario is about bounded time series on the interval $(0, 1)$. We illustrate the finite-sample behaviour of the estimators given in Subsection 3.3 for the bounded/binary SemiParTS model (presented in Subsection 2.3) driven by the shifted gamma AR(1) process (2).
Semiparametric time series models

Table 2
Empirical means and standard errors of the quasi-likelihood estimates of $\beta$ and method of moments estimates of $\phi$, $\sigma^2$ and $\rho$ based on the SemiParTS for real-valued data.

| parameter | true value | $n = 500$ | $n = 1000$ | $n = 2000$ |
|-----------|------------|-----------|-------------|-------------|
| $\beta_0$ | 0.1        | 0.106     | 0.100       | 0.096       |
| $\beta_1$ | 0.5        | 0.496     | 0.501       | 0.502       |
| $\beta_2$ | 0.7        | 0.696     | 0.697       | 0.699       |
| $\phi$    | 3          | 2.700     | 2.813       | 2.832       |
| $\sigma^2$| 1          | 1.280     | 1.184       | 1.157       |
| $\rho$    | 0.5        | 0.519     | 0.516       | 0.499       |

Figure 3: Simulated trajectories of a bounded time series (to the left) and associated latent process (to the right) under the third scenario.

Here, we set the parameters of the latent process as $\sigma^2 = 0.3$ and $\rho = 0.8$, the dispersion parameter $\phi = 0.1$, and the regression covariates including trend components:

$$x_{nt} = \left\{ 1, t/n, (t/n)^2 \right\}^T, \quad t = 1, \ldots, n,$$

with associated regression coefficients $\beta = (1, 0.3, 0.5)^T$. The time series generator here assumes that $Y_t$ given $\alpha_t$ follows a beta distribution with mean $\mu_t = \exp(-x_{nt}^T\beta - \alpha_t)$ and variance $\phi\mu_t(1 - \mu_t)$, for $t = 1, \ldots, n$. Figure 3 provides simulated trajectories of a bounded time series and its associated latent process under the third scenario.

In Table 3 we present the empirical means and standard errors of the quasi-likelihood estimates, as well as the estimates by the method of moments. In this case, we can observe a considerable bias in the quasi-likelihood estimates for the $\beta$’s, especially for $n = 500$. On the other hand, we see a good performance of the method of moments estimators for the parameters $\phi$, $\sigma^2$, and $\rho$. This difficulty in estimating the regression coefficients was reported by Davis and Wu (2009) in a similar setting. The authors considered a binary time series model driven by an exponential latent process. In the simulated results of that paper, it is only assumed an intercept for the mean and a GLM approach is considered for estimating it, which yielded estimates with considerable bias.
Anyway, we have empirical evidence that the proposed estimators are consistent for the scenario considered here even for the quasi-likelihood estimators of the $\beta$’s as the sample size increases.

Table 3

Empirical means and standard errors of the quasi-likelihood estimates of $\beta$ and method of moments estimates of $\phi$, $\sigma^2$ and $\rho$ based on the SemiParTS for bounded data.

| parameter | true value | $n = 500$ | $n = 1000$ | $n = 2000$ |
|-----------|------------|------------|------------|------------|
| $\beta_0$ | 1          | 0.932      | 0.965      | 0.989      |
| $\beta_1$ | 0.3        | 0.616      | 0.437      | 0.349      |
| $\beta_2$ | 0.5        | 0.228      | 0.384      | 0.459      |
| $\phi$    | 0.1        | 0.096      | 0.099      | 0.099      |
| $\sigma^2$| 0.3        | 0.333      | 0.301      | 0.306      |
| $\rho$    | 0.8        | 0.773      | 0.788      | 0.792      |

4.2. Out-of-sample forecasting

We here propose two methods to perform forecasting and present some pseudo-out-of-sample prediction exercises based on simulations. The first method is based on simulations, which is a common approach for latent factor-based time series models. Assume that we have observed a time series $Y_1, \ldots, Y_n$ and are interested in forecasting $Y_{n+1}$. We estimate the parameters based on $Y_1, \ldots, Y_n$ and use them to generate $M$ trajectories of sample size $n+1$, say $\{\tilde{Y}_{ij}; i = 1, \ldots, n+1, j = 1, \ldots, M\}$. This generation is done by using the same strategy to obtain the standard errors discussed in Section 3 (before Subsection 3.1) and illustrated in Subsection 4.1. Hence, we use some empirical quantity from the generated sample $\{\tilde{Y}_{n+1j}; j = 1, \ldots, M\}$, such as mean, median or other quantiles, for predicting $Y_{n+1}$. Forecasting based on simulated trajectories has been considered, for example, by Muniai and Ziel (2020). As argued by the authors, the simulated paths, also called an ensemble, can be seen as probabilistic forecasts which approximate well the underlying predictive distribution. Forecast for a longer horizon $h > 1$ follows in a similar fashion just by generating trajectories until time $n + h$. In our simulated results, we have experienced that this approach does not provide accurate predictions and just gives a big picture of how the series will behave in the future. This is already expected since the method is based on simulated trajectories. The simulated-based forecasting procedure above will be illustrated in the real data applications.

We now propose an analytical forecasting method, which has a low computational cost and also performs more accurately in contrast with the above simulated method. Following as in proof of Proposition 2.7 from Barreto-Souza and Ombao (2019), we have that

$$P_{n+h|n} \equiv E(Y_{n+h}|Y_n) = E \left( E(Y_{n+h}|\alpha_n)|Y_n \right),$$

(12)

where

$$E(Y_{n+h}|\alpha_n) = E \left( E(Y_{n+h}|\alpha_{n+h})|\alpha_n \right) = E \left( h(x_{n+h}^T \beta + \alpha_{n+h})|\alpha_n \right),$$

(13)

for $n, h \in \mathbb{N}$. The proposed analytical prediction of $Y_{n+h}$ given $Y_n$ will be given by $P_{n+h|n}$ in (12) by replacing the parameters by their estimates, say $\hat{P}_{n+h|n}$. Note that we are not using the whole past,
just the “present” observation, which is adequate for Markov Chains. In our case, our processes are hidden Markov chains (note, these are not Markov chains). On the other hand, we do not use this fact in our inferential procedure. Moreover, as it will be illustrated in both simulated and real data applications, such analytical procedure works quite satisfactorily. Let us now provide explicit expressions for (12) and (13) for the non-negative case discussed in this paper, which will be considered in the numerical experiments. The remaining cases follow in a very similar fashion and therefore are omitted. Note that we need to specify some conditional distribution for computing the standard errors.

Consider \( h(x) = e^x \), and \( \mu_n \) and \( \hat{\mu}_n \) as defined previously in Subsection 2.1. We have that

\[
E(Y_{n+h}|\alpha_t) = e^{x_{n+h}^T} E(e^{\alpha_t + h}) = \exp\left(x_{n+h}^T \beta + \rho^h \sigma^2 (1 - \rho^h)/2 + \rho^h \alpha_t\right),
\]

where we have used the fact that \( \alpha_t \), for all \( t \in \mathbb{N} \). We will consider the same conditional distribution discussed for computing the above expression for \( \sigma^2 \) and variance \( \phi \tilde{\mu}_n \) for all \( n \). Hence, the conditional expectation in (14) is

\[
E\left(\exp\{\rho^h \alpha_t\}|Y_n\right) = \frac{\int_{-\infty}^{\infty} \exp\left\{ -\frac{\alpha}{\phi} - \frac{\phi y_n e^{-\alpha}}{\phi \mu_n} - \frac{(\alpha + \sigma^2/2)^2}{2\sigma^2} \right\} d\alpha}{\int_{-\infty}^{\infty} \exp\left\{ -\frac{\alpha}{\phi} - \frac{\phi y_n e^{-\alpha}}{\phi \mu_n} - \frac{(\alpha + \sigma^2/2)^2}{2\sigma^2} \right\} d\alpha}.
\]

which is computed numerically. For the count case, by assuming that the conditional distribution of \( Y_n \) given \( \alpha_n \) is Poisson with mean \( \hat{\mu}_n \), we obtain that

\[
E\left(\exp\{\rho^h \alpha_t\}|Y_n\right) = \frac{\int_{-\infty}^{\infty} \exp\left\{ -\mu_n e^\alpha + \alpha y_n - \frac{(\alpha + \sigma^2/2)^2}{2\sigma^2} \right\} d\alpha}{\int_{-\infty}^{\infty} \exp\left\{ -\mu_n e^\alpha + \alpha y_n - \frac{(\alpha + \sigma^2/2)^2}{2\sigma^2} \right\} d\alpha}.
\]

In what follows, we present a rolling estimation window approach for performing forecasting, which mimics a practitioner in real data application updating estimates as soon as new observations arrive and then predicting future observations. This approach has been considered, for instance, by Jung and Tremayne (2006), Agosto et al. (2016), Muniain and Ziel (2020), and Zhang et al. (2020). This will be important to check the predictive performance of our proposed procedure under different scenarios to be specified in the sequence. Suppose we have a time series \( \{Y_t\} \) with sample size \( n \). We then split it in two, say \( (Y_1, \ldots, Y_{n_0}) \) and \( (Y_{n_0+1}, \ldots, Y_{n}) \), for some \( n_0 < n \). Parameters are estimated based on \( Y_1, \ldots, Y_{n_0} \) and \( Y_{n_0+1} \) is predicted, say \( \hat{Y}_{n_0+1} \), as described above. Next step consists in moving the estimation window in 1 unit of time. Now, estimation is done by using \( Y_1, \ldots, Y_{n_0+1} \) (re-estimation by also considering the “new” observation \( Y_{n_0+1} \)) and \( Y_{n_0+2} \) is predicted, say \( \hat{Y}_{n_0+2} \). This procedure is executed until reaching the last observation, so yielding the forecasts \( \hat{Y}_{n_0+1}, \ldots, \hat{Y}_{n} \). The prediction here will be based on the analytical form given in Eq. (14).
To evaluate the predictive power of the proposed methodology, we consider the loss function given by the mean-square forecasting error (in short MSFE):

$$\text{MSFE}_t = \frac{1}{t - n_0} \sum_{s=n_0+1}^{t} (Y_s - \hat{Y}_s)^2, \quad t = n_0 + 1, \ldots, n.$$ 

This loss function has been considered for instance in Agosto et al. (2016) for checking and comparing the predictive power of the Poisson INGARCH model by including or not covariates in a real data application.

As suggested by a referee, we now evaluate the forecasting performance under low and large variance (via $\sigma^2$ and $\phi$), low and large sample size, and also by considering $\rho$ (parameter controlling dependence) close or not to the boundary of the parameter space. We consider the positive continuous case with (i) $\sigma^2 = 0.05, 0.6$, (ii) initial sample size $n_0 = 200, 400$, (iii) $\rho = 0.5, 0.95$, and (iv) $\phi = 0.05, 0.8$. In each of these configurations, the remaining settings not explicitly mentioned here are exactly specified as in the simulations presented in Subsection 4.1. The total sample size for all cases is $n = 600$. To obtain an empirical probabilistic behavior of the MSFE quantity, we run a Monte Carlo simulation with 1000 replications.

Figure 4: Mean MSFE curves for the simulated positive continuous case under the settings (i) $\sigma^2 = 0.05, 0.6$, (ii) initial sample size $n_0 = 200, 400$, (iii) $\rho = 0.5, 0.95$, and (iv) $\phi = 0.05, 0.8$.

Figure 4 presents the mean MSFE curves for the simulated positive continuous case. As expected, the forecasting is more precise under the settings where we have more information or less
Semiparametric time series models

uncertainty, which are \( \sigma^2 = 0.05, n_0 = 400, \rho = 0.95, \) and \( \phi = 0.08 \) in the Scenarios (i), (ii), (iii), and (iv), respectively.

5. Time series data applications

In this section we apply the SemiParTS models for analyzing time series on the number of asthma cases and total insolation, so illustrating the performance of the proposed methodology for different types of time series data (count and positive continuous). Further, we compare the predictive power of the SemiParTS models with existing models through the rolling estimation window approach discussed in Subsection 4.2.

5.1. Number of asthma cases

This first application is devoted to the daily presentations of asthma from a single hospital at Campbelltown, South West region of Sydney, Australia (Campbelltown). This count time series was observed from 1 January 1990 to 1 January 1999, with total number of observations \( n = 1461 \), which can be obtained from the R package glarma; for instance, see https://CRAN.R-project.org/package=glarma. This dataset has been analyzed by Davis et al. (2003a) under the Generalized Linear Autoregressive Moving Average Models (GLARMA) approach, which will be considered for comparison purposes at the end of this subsection. We here consider the non-negative (more specifically count) time series model presented in Subsection 2.1 for this application with \( p = 1 \) and \( \phi = 1 \), which mimics the first two moments of a Poisson distribution for the conditional distribution of the response given the latent process. The aim is to study the relationship between atmospheric pollution and the number of admissions due to asthma cases. For more details, we refer the reader to Davis et al. (2003a). The following covariates are considered here: indicator variables for measuring the effect of Sunday and Monday, terms of annual seasonality, scaled lagged and smoothed humidity, and maximum nitrogen dioxide.

Figure 5: Plots of the daily number of asthma cases in the Campbelltown Hospital from 1 January 1990 to 1 January 1999 (to the left) and its associated ACF (to the right).

Figure 5 displays plots of the daily number of asthma cases and its associated ACF, which indicates that the series is non-stationary. The non-stationarity will be taken into account in our SemiParTS model through the incorporation of covariates.
Table 4
Parameter estimates and respective standard errors of the semiparametric count time series model for the number of asthma cases based on quasi-likelihood method (combined with the method of moments) and on Monte Carlo simulations.

| covariates/par. | Quasi+MM estimates | stand.err. | Simulated estimates | stand.err. |
|-----------------|---------------------|------------|---------------------|------------|
| Intercept       | 0.680               | 0.058      | 0.676               | 0.073      |
| Sunday          | 0.206               | 0.056      | 0.206               | 0.056      |
| Monday          | 0.227               | 0.055      | 0.228               | 0.056      |
| $\cos(2\pi t/365)$ | $-0.198$           | 0.035      | $-0.197$            | 0.050      |
| $\sin(2\pi t/365)$ | 0.375               | 0.029      | 0.377               | 0.047      |
| Humidity        | 0.191               | 0.053      | 0.190               | 0.077      |
| NO$_{2}$max     | $-0.083$            | 0.032      | $-0.083$            | 0.039      |
| $\sigma^2$      | 0.089               | –          | 0.108               | 0.043      |
| $\rho$          | 0.838               | –          | 0.738               | 0.165      |

Table 4 provides the estimates of the parameters with their respective standard errors obtained through Monte Carlo simulation (fourth column). For the simulated results, we take a Poisson distribution for generating time series as discussed in Section 4. We also present the standard errors obtained from the quasi-likelihood (QL) estimation by ignoring the dependence among the observations due to the latent process (second column). As can be seen, there is a difference between the standard errors based on the Monte Carlo simulation (considering the presence of the latent process) and those from the quasi-likelihood approach for some parameter estimates. This is also nicely discussed in the papers by Davis et al. (2000) and Davis and Wu (2009), where a generalized linear model approach is considered. Anyway, we obtain that all the covariates are significant by using a significance level at 5% based on both approaches. In the next application, we will see that different conclusions might be drawn depending on the approach. From Table 4, it is also possible to note a suitable agreement between quasi-likelihood and method of moments estimates and those caught from Monte Carlo simulation.

Figures 6 and 7 respectively show the histograms and qq-plots of the standardized Monte Carlo estimates of the $\beta$’s, which indicate satisfactory normal approximations.

We now focus our attention on analyzing the predictive power of our fitted SemiParTS model. For sake of comparison, we consider the GLARMA model (moving average component at lag 7) from the empirical illustration given in Davis et al. (2003a). We split the asthma series with $n_0 = 1200$ being the initial sample size. The rolling estimation window described in Subsection 4.2 is performed with one-step-ahead predictions (in each step, the parameters are re-estimated). Figure 8 present the count time series data with one-step-ahead forecasting based on the SemiParTS approach (via conditional expectation and simulated trajectories) and GLARMA model. The MSFE curves are presented in Figure 9. These plots reveal that the forecasting procedure based on the SemiParTS model via the conditional expectation given in (12) provides the best result.
5.2. Total insolation data analysis

We now consider the monthly total insolation data (in hours) of the city of Belo Horizonte in the state of Minas Gerais, Brazil, from January 1961 to January 2019. These data consist of $n = 616$ observations and can be obtained from the Meteorological Database for Teaching and Research – INMET, Brazil; please see http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep.

Figure 10 presents the plots of the total insolation time series and its associated ACF. As expected, we can see a seasonal behaviour of this time series. The model for non-negative time series data given in Subsection 2.1 with $p = 2$ ($V(\mu) = \mu^2$) is applied here. After a preliminary analysis,
we consider the following covariates:

$$x_{nt} = \begin{bmatrix} 1, \cos(2\pi t/12), \cos(2\pi t/6), \cos(2\pi t/3) \end{bmatrix}^\top,$$

for $t = 1, \ldots, 616$.

Table 5 shows the parameter estimates and respective standard errors of the SemiParTS model for the total insolation data. For getting the Monte Carlo results, we followed the strategy discussed in Section 4 and considered a conditional gamma distribution for $Y_t$ given $\alpha_t$, for $t = 1, \ldots, 616$.

The quasi-likelihood and method of moments procedures provide similar estimates than the Monte Carlo method, especially for estimating the $\beta$’s. By using a significance level at 5% and
Figure 8: One-step ahead predictions of the asthma count data from \( t = 1201 \) to \( t = 1461 \) under our SemiParTS approach given in Eq. (14) (top to the left) and via simulated trajectories (top to the right), and under GLARMA model (bottom).

Figure 9: MSFE curves based on simulated trajectories and analytical predictions of our SemiParTS model and GLARMA approach.

taking into account the latent process, the covariates \( \cos(2\pi t/12) \) and \( \cos(2\pi t/3) \) were significant. These covariates correspond to annual and quarterly seasonality. On the other hand, by ignoring the presence of the latent process, the quarterly seasonalities are not significant. This shows the impor-
Semiparametric time series models

Figure 10: Plots of the monthly total insolation in the city of Belo Horizonte from January 1961 to January 2019 (to the left) and its associated ACF (to the right).

Table 5
Parameter estimates and respective standard errors of the semiparametric non-negative time series model for the total insolation data based on quasi-likelihood method (combined with method of moments) and on Monte Carlo simulations.

| covariates/par. | Quasi+MM estimates | stand. err. | Simulated estimates | stand. err. |
|-----------------|--------------------|-------------|---------------------|-------------|
| Intercept       | 5.335              | 0.008       | 5.335               | 0.001       |
| cos(2πt/12)     | 0.046              | 0.012       | 0.046               | 0.015       |
| cos(2πt/6)      | −0.008             | 0.012       | −0.007              | 0.011       |
| cos(2πt/3)      | −0.021             | 0.012       | −0.021              | 0.010       |
| φ               | 0.021              | –           | 0.020               | 0.006       |
| σ²              | 0.022              | –           | 0.022               | 0.006       |
| ρ               | 0.600              | –           | 0.606               | 0.121       |

In Figure 11 and 12, we present the histograms and qq-plots of the standardized quasi-likelihood estimates of the β’s, respectively. These plots again indicate a satisfactory normal approximation for the distribution of the quasi-likelihood estimators. This is in line with our simulated results provided in Section 4.

For analyzing the predictive power of our fitted SemiParTS model in this insolation time series data, we split the series with \( n_0 = 400 \). The rolling estimation window is then performed with one-step-ahead predictions. We here compare our approach with the ARIMAX(1, 1, 1) model, where \( X \) stands for the inclusion of covariates (we use the same set of covariates as in our model). Figure 13 provides the insolation data with one-step-ahead forecasting based on the SemiParTS approach (via conditional expectation and simulated trajectories) and ARIMAX model. Figure 14 gives the MSFE curves. Once again, we obtain the best forecasting performance from the SemiParTS
6. Concluding remarks

A flexible class of semiparametric time series (SemiParTS) models was proposed by assuming a quasi-likelihood model driven by a latent factor process. Our proposed methodology has the advantage of generality: it can model a variety of time series outcomes including positive continuous, count, bounded, binary, and real-valued. Inference on the model parameters was discussed and Monte Carlo simulations were addressed for checking estimation performance. Applications on the number of asthma cases and insolation time series data illustrated the usefulness of the proposed methodology in practical situations.

A challenging point seems to be the estimation of the parameters related to the mean for the model via the conditional expectation given in (12). Note that the prediction based on simulated trajectories does not produce satisfactory results in this particular empirical illustration.
Figure 13: One-step ahead predictions of the insolation data from $t = 401$ to $t = 616$ under our SemiParTS approach given in Eq. (14) (top to the left) and via simulated trajectories (top to the right), and under ARIMAX model (bottom).

Figure 14: MSFE curves based on simulated trajectories and analytical predictions of our SemiParTS model and ARIMAX approach.

bounded case, where a considerable bias was observed, which was also indicated in Davis and Wu (2009) in a binary time series model. A possible solution may be to use a bootstrap procedure (Efron and Tibshirani, 1994) for obtaining the bias and then correct the quasi-likelihood estimates.
Semiparametric time series models

We do not pursue this idea here because this will present another layer of computational complexity and our goal, for now, is to point out this problem.

We would like to call attention to the further generality and flexibility of the proposed Semi-ParTS method. Indeed, other forms for the variance function can be considered and the results discussed in this paper can be easily adapted. For example, in the bounded case, one might be interested in considering the variance function $V(\mu) = \mu^3(1 - \mu)^3$, with $\mu \in (0, 1)$. The marginal moments and autocorrelation function for this case are obtained following the same steps given in Subsection 2.3. If higher-order moments are important and informative (here we have worked with the two first conditional moments specified), a different approach is necessary and we believe this is worth investigating. Other points we believe that deserve to be investigated in future research are: (i) deeper study on forecasting, (ii) diagnostic tools, and (iii) multivariate extension.

Acknowledgments

We would like to thank the Editors of the Special Issue 30 Years of Cointegration and Dynamic Factor Models Forecasting and its Future with Big Data and the anonymous Referee for the insightful and valuable comments and suggestions which leads to a substantial improvement of the paper. G. Maia and W. Barreto-Souza would like to thank the financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil, grant number 305543/2018-0). W. Barreto-Souza and H. Ombao would like to acknowledge support for their work by KAUST Research Fund and NIH 1R01EB028753-01.

References

Agosto, A., Cavaliere, G., Kristensen, D., Rahbek, A., 2016. Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). Journal of Empirical Finance 38, 640–663.

Alzahrani, N., Neal, P., Spencer, S., McKinley, T., Touloupou, P., 2018. Model selection for time series of count data. Computational Statistics and Data Analysis 122, 33–44.

Barreto-Souza, W., Ombao, H., 2019. Negative binomial process: A tractable model with composite likelihood-based inference. Submitted for publication.

Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M., 2003. Generalized autoregressive moving average models. Journal of the American Statistical Association 98, 214–223.

Berkes, I., Horváth, L., 2003. The rate of consistency of the quasi-maximum likelihood estimator. Statistics and Probability Letters 61, 133–143.

Cavaliere, G., Xu, F., 2014. Testing for unit roots in bounded time series. Journal of Econometrics 178, 259–272.

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.

Cox, D.R., 1981. Statistical analysis of time series: some recent developments. Scandinavian Journal of Statistics 93–115.

Davis, R., Dunsmuir, W., Streett, S., 2003a. Observation-driven models for poisson counts. Biometrika 90, 777–790.

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

Davis, R.A., Dunsmuir, W.T., Streett, S.B., 2003b. Observation-driven models for Poisson counts. Biometrika 90, 777–790.

Davis, R.A., Dunsmuir, W.T., Wang, Y., 2000. On autocorrelation in a Poisson regression model. Biometrika 87, 491–505.

Davis, R.A., Wu, R., 2009. A negative binomial model for time series of counts. Biometrika 96, 735–749.

Efron, B., Tibshirani, R., 1994. An Introduction to the Bootstrap. Chapman & Hall, CRC Press.

Françco, C., Zakoïan, J.M., 2004. Maximum likelihood estimation of pure garch and arma-garch processes. Bernoulli 10, 605–637.

Heyde, C.C., 1997. Quasi-likelihood and its Application: A General Approach to Optimal Parameter Estimation. Springer Science & Business Media.

Jung, R., Tremayne, A., 2006. Coherent forecasting in integer time series models. International Journal of Forecasting 22, 223–238.

Muniain, P., Ziel, F., 2020. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. International Journal of Forecasting 36, 1193–1210.

Ombao, H., Fiecas, M., Ting, C., Low, Y., 2018. Statistical models for brain signals with properties that evolve across trials. NeuroImage 180, 609–618.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. Test 18, 529–545.
Semiparametric time series models

Sim, C., 1990. First-order autoregressive models for gamma and exponential processes. Journal of Applied Probability 27, 325–332.

Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. Annals of Statistics 34, 2449–2495.

Wedderburn, R.W., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 61, 439–447.

Zeger, S.L., 1988. A regression model for time series of counts. Biometrika 75, 621–629.

Zeger, S.L., Qaqish, B., 1988. Markov regression models for time series: a quasi-likelihood approach. Biometrics 44, 1019–1031.

Zhang, Y., Ma, F., Liao, Y., 2020. Forecasting global equity market volatilities. International Journal of Forecasting 36, 1454–1475.