Journal of Econometrics 246 (2024) 105894

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Pseudo-variance quasi-maximum likelihood estimation of semi-parametric time series models

Mirko Armillotta [a], Paolo Gorgi [b,c,*]

[a] University of Rome Tor Vergata, Rome, Italy
[b] Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[c] Tinbergen Institute, Amsterdam, The Netherlands

ARTICLE INFO

Dataset link: https://github.com/mirkoarmillotta/Pseudo_variance

JEL classification:
C32
C52
C58

Keywords:
Double-bounded time series
Integer-valued autoregressions
Quasi-maximum likelihood

ABSTRACT

We propose a novel estimation approach for a general class of semi-parametric time series models where the conditional expectation is modeled through a parametric function. The proposed class of estimators is based on a Gaussian quasi-likelihood function and it relies on the specification of a parametric pseudo-variance that can contain parametric restrictions with respect to the conditional expectation. The specification of the pseudo-variance and the parametric restrictions follow naturally in observation-driven models with bounds in the support of the observable process, such as count processes and double-bounded time series. We derive the asymptotic properties of the estimators and a validity test for the parameter restrictions. We show that the results remain valid irrespective of the correct specification of the pseudo-variance. The key advantage of the restricted estimators is that they can achieve higher efficiency compared to alternative quasi-likelihood methods that are available in the literature. Furthermore, the testing approach can be used to build specification tests for parametric time series models. We illustrate the practical use of the methodology in a simulation study and two empirical applications featuring integer-valued autoregressive processes, where assumptions on the dispersion of the thinning operator are formally tested, and autoregressions for double-bounded data with application to a realized correlation time series.

1. Introduction

A wide range of time series models have been proposed in the literature to model the conditional mean of time series data. Their specification often depends on the nature of the time series variable of interest. For example, AutoRegressive Moving Average (ARMA) models (Box et al., 1970) are typically employed for time series variables that are continuous and take values on the real line. INteger-valued AutoRegressive (INAR) models (Al-Osh and Alzaid, 1987; McKenzie, 1988) and INteger-valued GARCH models (INGARCH) (Heinen, 2003; Ferland et al., 2006) are designed to account for the discrete and non-negative nature of count processes. Autoregressive Conditional Duration (ACD) models (Engle and Russell, 1998) are used for modeling non-negative continuous processes. Beta autoregressive models (Rocha and Cribari-Neto, 2009) are employed for modeling double-bounded time series data lying in a specified interval domain. The estimation of such models can be carried out by the Maximum Likelihood Estimator (MLE), which constitutes the gold standard approach for the estimation of unknown parameters in parametric models. However, the MLE requires parametric assumptions on the entire conditional distribution of the time series process. This feature is not appealing when the interest of the study is only on modeling the conditional mean instead of the entire conditional distribution. Furthermore, the

* Corresponding author at: Tinbergen Institute, Amsterdam, The Netherlands.
E-mail address: p.gorgi@vu.nl (P. Gorgi).

https://doi.org/10.1016/j.jeconom.2024.105894
Received 25 October 2023; Received in revised form 3 September 2024; Accepted 7 November 2024

likelihood function can sometimes present a complex form and the implementation of the MLE can become unfeasible. For instance, maximum likelihood inference of INAR(p) models is well-known to be cumbersome and numerically difficult when the order of the model p is large (Pedeli et al., 2015). In such situations, the use of quasi-likelihood methods becomes attractive.

The Quasi-MLE (QMLE), introduced by Wedderburn (1974), is a likelihood-based estimator where there is a quasi-likelihood that is not necessarily the true distribution of the data. Quasi-likelihoods are typically a member of the one-parameter exponential family. Gourieroux et al. (1984) show that the QMLE is consistent for the true unknown parameters of the model. Nevertheless, QMLEs can be inefficient because, given a parametric definition for the conditional mean of the process, the conditional variance is implicitly constrained to be a function of the conditional mean as determined by the exponential family of distributions that is considered. In order to improve the estimation efficiency for the parameters of the conditional mean in time series models, Aknouche and Francq (2023) propose a two-stage Weighted Least Squares Estimator (WLSE) where in the first step the conditional variance of the process is estimated and it is then used in the second step as weighting sequence for the solution of the weighted least squares problem. It is shown that this WLSE leads to improved efficiency with respect to QMLE if the variance function is correctly specified. A similar estimator has been more recently proposed in the context of estimating functions approach leading to the same type of efficiency improvement (Francq and Zakoian, 2023).

In this paper, we propose a novel class of QMLEs for the estimation of the conditional expectation of semi-parametric time series models. The estimators are based on a Gaussian quasi-likelihood and a pseudo-variance specification, which can contain restrictions with the parameters of the conditional expectation. The Pseudo-Variance QMLEs (PVQMLEs) only require parametric assumptions on the conditional expectation as the pseudo-variance function does not need to be correctly specified. We establish strong consistency and asymptotic normality of the PVQMLEs under very general conditions. The case in which the pseudo-variance formulation corresponds to the true conditional variance of the process is obtained as a special case. We show that when no restrictions are imposed between the mean and pseudo-variance, the resulting unrestricted PVQMLE has the same asymptotic efficiency of a particular WLSE. Furthermore, if the pseudo-variance is correctly specified it achieves the same asymptotic efficiency as the efficient WLSE. On the other hand, when parameter restrictions are considered, the resulting restricted PVQMLEs can achieve higher efficiency compared to the efficient WLSE and alternative QMLEs. This result is theoretically shown in some special cases and empirically verified for INAR models through an extensive numerical exercise. We discuss how the specification of the pseudo-variance and the parameter restrictions naturally arise for time series processes with bounded support. We obtain that the restricted PVQMLEs retain the desired asymptotic properties when the imposed restrictions are valid with respect to the true parameter of the mean and a pseudo-true parameter of the conditional variance. The validity of such restrictions can be tested without requiring correct specification of the conditional variance. We derive a test for this purpose that can be used as a consistency test for restricted PVQMLEs. When the evidence-based parameter constraints are identified and validated, they constitute a restriction set where an higher-efficiency restricted PVQMLE can be obtained. Furthermore, under correct specification of the pseudo-variance, the test can be used as a specification test on the underlying process generating the data.

Finally, the practical usefulness of PVQMLE approach is illustrated by means of two real data applications. One is concerned with INAR models and one with a Beta autoregression for double-bounded data. INAR processes depend on the distribution assumed for the innovation and the thinning specification (Lu, 2021). Guerrero et al. (2022) consider an alternative INAR parametrization that is based on the innovation and marginal distributions that leads to an equivalent INAR specification where the thinning operator is specified implicitly. Our test allows us to test for the degree of dispersion in the thinning operator as well as the error term. There exists a vast literature of INAR models in testing innovations and marginal distributions dispersion (Schweer and Weiß, 2014; Aleksandrov and Weiß, 2020), testing for serial dependence (Sun and McCabe, 2013), and general goodness of fit tests (Weiß, 2018b). However, to the best of our knowledge, specification tests are not available for the thinning dispersion. The thinning operator is typically assumed to be binomial, which implies underdispersion in the thinning. Once appropriate thinning and innovation restrictions are identified through the specification test, the corresponding PVQMLE is used to estimate the parameters of the INAR model. The second application concerns the analysis of daily realized correlations between a pair of stock returns, which forms a double-bounded time series as the realized correlation takes values between minus one and one. We consider a pseudo-variance specification based on the implied variance from Beta-distributed variables for the definition of PVQMLEs. We then test the validity of parametric restrictions between the mean and pseudo-variance to validate the use of restricted PVQMLEs.

The remainder of the paper is organized as follows. Section 2 introduces the general mean and pseudo-variance framework and the PVQMLEs, together with some examples. Section 3 presents the main theoretical results of the paper on the asymptotic properties of the PVQMLE and some special cases. Section 4 discusses the efficiency of the PVQMLE. Section 5 introduces the specification test for the validity of the constraints with an extensive simulation study in the case of INAR models. Section 6 presents empirical applications. Section 7 concludes the paper. The proofs of the main results are deferred to Appendix A. Finally, Appendix B includes additional numerical results.

2. Specification and estimation

2.1. PVQML estimators

Consider a stationary and ergodic time series process $\{Y_t\}_{t\in\mathbb{Z}}$ with elements taking values in the sample space $\mathcal{Y} \subseteq \mathbb{R}$ and with conditional mean given by

$$\mathrm{E}(Y_t|\mathcal{F}_{t-1}) = \lambda(Y_{t-1}, Y_{t-2}, \dots; \psi_0) = \lambda_t(\psi_0), \quad t \in \mathbb{Z}, \tag{1}$$

where $\mathcal{F}_t$ denotes the σ-field generated by $\{Y_s,\ s \le t\}$, $\lambda : \mathbb{R}^\infty \times \Psi \to \mathbb{R}$ is a known measurable function, and $\psi_0 \in \Psi \subset \mathbb{R}^p$ is the true unknown p-dimensional parameter vector. We denote with v_t the conditional variance of the process, i.e. $\mathrm{V}(Y_t|\mathcal{F}_{t-1}) = v_t$, which is considered to have an unknown specification. The model is a semi-parametric model as the quantity of interest is the parameter vector of the conditional mean ψ_0 and other distributional properties are left unspecified and treated as an infinite dimensional nuisance parameter. The general specification of the model in (1) includes a wide range of time series models as special case. For instance, it includes linear and non-linear ARMA models when $\mathcal{Y} = \mathbb{R}$, INGARCH and INAR models when $\mathcal{Y} = \mathbb{N}$, ACD models when $\mathcal{Y} = (0, \infty)$, and Beta autoregressive models for bounded data when $\mathcal{Y} = (0, 1)$.

The main objective is to estimate the parameter vector ψ_0 of the conditional expectation. For this purpose, we consider the specification of a pseudo-variance

$$v_t^*(\gamma) = v^*(Y_{t-1}, Y_{t-2}, \dots ; \gamma), \quad t \in \mathbb{Z}, \tag{2}$$

where $v^* : \mathbb{R}^\infty \times \Gamma \to [0, +\infty)$ is a known function that is indexed by the k-dimensional parameter $\gamma \in \Gamma \subset \mathbb{R}^k$. We refer to this as a pseudo-variance as it is not necessarily correctly specified, i.e. there may be no value $\gamma \in \Gamma$ such that $v_t^*(\gamma) = v_t$. The idea is to use the pseudo-variance $v_t^*(\gamma)$ to enhance the efficiency of the estimation of ψ_0 by means of a Gaussian QMLE. We denote the whole parameter vector that contains both the parameter of the mean and pseudo-variance with $\theta = (\psi', \gamma')'$ and $\theta \in \Theta = \Psi \times \Gamma \subset \mathbb{R}^m$, $m = p + k$.

We introduce the class of PVQMLEs that relies on a Gaussian quasi-likelihood for the mean equation with the pseudo-variance as scale of the Gaussian density. We consider estimators based on both unrestricted and restricted quasi-likelihood functions. Assume that we have an observed sample of size T from the process defined in (1), given by $\{Y_t\}_{t=1}^T$. Since $\lambda_t(\psi)$ and $v_t^*(\gamma)$ can depend on the infinite past of Y_t, we define their approximations of $\tilde{\lambda}_t(\psi)$ and $\tilde{v}_t^*(\gamma)$ based on the available finite sample $\{Y_t\}_{t=1}^T$,

$$\tilde{\lambda}_t(\psi) = \lambda(Y_{t-1}, \dots, Y_1, \tilde{Y}_0, \tilde{Y}_{-1}, \dots ; \psi), \quad \tilde{v}_t^*(\gamma) = v^*(Y_{t-1}, \dots, Y_1, \tilde{Y}_0, \tilde{Y}_{-1}, \dots ; \gamma), \tag{3}$$

where $\tilde{Y}_0, \tilde{Y}_{-1}, \dots$ are given initial values. The Gaussian quasi-likelihood for ψ with the pseudo-variance scaling is defined as

$$\tilde{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{l}_t(\theta), \quad \tilde{l}_t(\theta) = -\frac{1}{2} \log \tilde{v}_t^*(\gamma) - \frac{[Y_t - \tilde{\lambda}_t(\psi)]^2}{2\tilde{v}_t^*(\gamma)}. \tag{4}$$

Based on the quasi-likelihood function in (4), we define the unrestricted and restricted PVQMLE. The unrestricted PVQMLE is based on the unconstrained maximization of the pseudo-likelihood without imposing any constrains between ψ and γ. The unrestricted PVQMLE $\hat{\theta}$ is defined as

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} \tilde{L}_T(\theta), \tag{5}$$

where $\hat{\theta} = (\hat{\psi}', \hat{\gamma}')'$ and $\hat{\psi}$ is the unrestricted PVQMLE of ψ_0. In Section 3, we shall see that the unrestricted PVQMLE $\hat{\psi}$ is a consistent estimator of ψ_0 and, in fact, it is asymptotically equivalent to a specific WLSE. If the pseudo-variance is correctly specified, i.e. there is $\gamma_0 \in \Gamma$ such that $v_t^*(\gamma_0) = v_t$, then $\hat{\psi}$ is asymptotically equivalent to the efficient WLSE.

In models where the sample space $\mathcal{Y}$ is bounded, such as count-time series models, there can be a natural relationship between the conditional mean and variance of the process. For example, in a count time series process we have that if the mean goes to zero, then also the variance goes to zero as, in fact, the limit case is the mean being exactly zero. Such relationship between mean and variance, as given by parametric models, provide a natural way to introduce restrictions between the mean and pseudo-variance parameters ψ and γ. Several examples are presented at the end of this section.

To specify the restricted PVQMLE, we consider the constrained parameter set Θ_R that imposes r restrictions on the pseudo-variance parameters

$$\Theta_R = \{\theta \in \Theta : S\gamma = g(\psi)\},$$

where S is a $r \times k$ selection matrix and $g : \Psi \to \mathbb{R}^r$. The estimator derived from the maximization of (4) over the set Θ_R is the restricted PVQMLE,

$$\hat{\theta}_R = \operatorname*{argmax}_{\theta \in \Theta_R} \tilde{L}_T(\theta) \tag{6}$$

where $\hat{\theta}_R = (\hat{\psi}_R', \hat{\gamma}_R')'$ and $\hat{\psi}_R$ is the restricted PVQMLE of ψ_0. In Section 3, we shall see that the restricted PVQMLE $\hat{\psi}_R$ is a consistent estimator of ψ_0 if the constrains in Θ_R hold with respect to a pseudo-true parameter γ^*. The advantage of the restricted PVQMLE $\hat{\psi}_R$ is that it can achieve higher efficiency than the unrestricted one. Furthermore, as it shall be presented in Section 5, the validity of the restrictions can be tested under both misspecification and correct specification of the pseudo-variance. The test can be interpreted as a consistency test for the restricted estimator when the pseudo-variance is misspecified. Instead, it can be employed as a specification test if we assume correct specification of the pseudo-variance. For instance, it shall be employed to test for underdispersion, equidispersion or overdispersion in the thinning operator of INAR models.

In practice, both the unrestricted and the restricted PVQMLE (5)–(6) do not have a closed form solution. Therefore, the estimation of model parameters is carried out by numerical optimization. This is done by employing standard optimization functions of the R software using the BFGS algorithm (Nocedal and Wright, 1999).

2.2. Examples

The model specification in (1) is very general and it covers a wide range of semi-parametric observation-driven time series model. The unrestricted and restricted QMLE based on the pseudo-variance in (2) can be employed for such general class of models. However, PVQMLEs are particularly suited for time series processes where the support of the conditional mean is bounded and a natural relationship with the conditional variance can be assumed. In Section 4.2 it will be shown that in models where conditional mean and pseudo-variance share some parameter restrictions, a more efficient estimator may be obtained with respect to alternative estimation approaches available in the literature. The specification of the pseudo-variance and the parameter restrictions with the conditional mean can be based on well known model specifications. The validity of such restrictions is testable and the asymptotic properties do not require correct specification of the pseudo-variance. This means that no assumptions on the true conditional variance are needed and the consistency of the restricted PVQMLE can also be tested without relying on such assumptions. Below we present some examples of models that are encompassed in the framework defined in Eqs. (1) and (2), and provide a general way to specify the pseudo-variance and the parameter restrictions with the conditional mean.

Example 1 (*INAR Models*)**.** INAR models are widely used in the literature to model count time series. The INAR(1) model is given by

$$Y_t = a \circ Y_{t-1} + \varepsilon_t , \quad t \in \mathbb{Z}, \tag{7}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an iid sequence of non-negative integer-valued random variables with mean $\omega_1 > 0$ and variance $\omega_2 > 0$, and '$\circ$' is the thinning operator of Steutel and Van Harn (1979). For a given $N \in \mathbb{N}$, a general formulation of the thinning operator is $a \circ N = \sum_{j=1}^{N} X_j$ when $N > 0$, and 0 otherwise, where X_j is a sequence of iid non-negative integer-valued random variables following a distribution with finite mean a and variance b, say $X_j \sim D_X(a, b)$. The most common formulation (Steutel and Van Harn, 1979) is the binomial thinning where X_j is a sequence of independent Bernoulli random variables with success probability $a \in (0, 1)$, therefore $a \circ N$ is a binomial random variable with N trials and success probability a. The conditional mean of the INAR(1) is

$$\lambda_t = aY_{t-1} + \omega_1 . \tag{8}$$

The form of the variance for INAR models is known to be linear in the observations, therefore the pseudo-variance can be specified as

$$v_t^* = bY_{t-1} + \omega_2 . \tag{9}$$

Several restrictions can be considered for the PVQMLE. For instance, the restriction $b = a(1-a)$ is implied by a binomial thinning and $\omega_1 = \omega_2$ is implied by a Poisson error. The same estimation framework applies to INAR models with general lag order p, called INAR(p).

$$Y_t = a_1 \circ Y_{t-1} + \cdots + a_p \circ Y_{t-p} + \varepsilon_t , \quad t \in \mathbb{Z},$$

$$\lambda_t = \sum_{h=1}^{p} a_h Y_{t-h} + \omega_1 , \qquad v_t^* = \sum_{h=1}^{p} b_h Y_{t-h} + \omega_2 .$$

Further results on INAR models are discussed in Sections 5 and 3.1. An application to real data is presented in Section 6.

Example 2 (*INGARCH Models*)**.** Another popular model for time series of counts is the INGARCH model. The conditional mean of the INGARCH(1, 1) model takes the form

$$\lambda_t = \omega_1 + \alpha_1 Y_{t-1} + \beta_1 \lambda_{t-1} , \tag{10}$$

where $\omega_1, \alpha_1, \beta_1 \geq 0$. The pseudo-variance can be specified as

$$v_t^* = \omega_2 + \alpha_2 Y_{t-1} + \beta_2 \lambda_{t-1} . \tag{11}$$

Also in this case, several restrictions can be considered for the PVQMLE. For instance, the restrictions $\omega_2 = \omega_1$, $\alpha_2 = \alpha_1$ and $\beta_2 = \beta_1$ are implied by an equidispersion assumption $v_t^* = \lambda_t$, which follows assuming a conditional Poisson distribution for example. Alternatively, the restrictions $\omega_2 = c\omega_1$, $\alpha_2 = c\alpha_1$ and $\beta_2 = c\beta_1$ with $c > 0$ are implied by a proportional variance assumption $v_t^* = c\lambda_t$.

Example 3 (*ACD Models*)**.** ACD models are typically used to model non-negative continuous time series variables, like durations or volumes. These models take the form $Y_t = \lambda_t \varepsilon_t$ where ε_t is a sequence of positive variables with mean equal to 1. The conditional expectation λ_t may take the form as in Eq. (10). The pseudo-variance can be specified in several ways and restrictions can be imposed. For instance, the restriction $v_t^* = \lambda_t^2$ follows by assuming an exponential error distribution. An alternative restriction is given by $v_t^* = c\lambda_t^2$, $c > 0$.

Example 4 (*Double-bounded Autoregressions*)**.** For double-bounded time series data the conditional mean λ_t can be specified as in Eq. (10), see Gorgi and Koopman (2023) for instance. Several specifications and restrictions for the pseudo-variance can be considered. For instance, the restriction $v^*_t = \lambda_t(1-\lambda_t)/(1+\phi)$ is implied by a beta conditional distribution with dispersion parameter $\phi > 0$. Intermediate restrictions on the pseudo-variance are discussed in the corresponding application in Section 6. See also Section 3.2 for further results established on this class of models.

We note that the examples presented in this section are focused on a linear mean equation for simplicity of exposition. Several other non-linear model specifications are encompassed in the general framework in (1) and (2), see for example Creal et al. (2013) and Christou and Fokianos (2015).

3. Asymptotic theory

In this section, the asymptotic properties of the PVQMLEs in (5) and (6) are formally derived. Although asymptotic results related to quasi-maximum likelihood estimators of observation-driven models are well-established in the literature, the associated theory for PVQMLEs differs as it relies on simultaneous estimation of mean and pseudo-variance parameters, where the latter can be misspecified and present parameter restrictions with the mean. Since the pseudo-variance can be misspecified, the estimator of the pseudo-variance parameter $\hat{\gamma}$ will be consistent with respect to a pseudo-true value γ^*, which is given by

$$\gamma^* = \underset{\gamma\in\Gamma}{\operatorname{argmax}} -\frac{1}{2}\mathrm{E}\left(\log v^*_t(\gamma) + \frac{[Y_t-\lambda_t(\psi_0)]^2}{v^*_t(\gamma)}\right). \tag{12}$$

We define the vector $\theta_0 = (\psi_0', \gamma^{*\prime})'$ that contains both true and pseudo-true parameters. The estimator of the mean parameters preserves the consistency and asymptotic normality results to the true parameter vector ψ_0 irrespective of the correct specification of the conditional variance. We show that such result holds for both unrestricted (5) and restricted (6) estimators, where the restricted estimator requires the validity of the imposed restrictions with respect to the pseudo-true parameter, i.e. $S\gamma^* = g(\psi_0)$. We note that the validity of such restriction is a weaker condition than the correct specification of the pseudo-variance. In fact, the test proposed in Section 5 is a restriction test and, under the null hypothesis of valid restrictions, the pseudo-variance can still be misspecified.

We start by showing consistency and asymptotic normality of the unrestricted PVQMLE in (5). We first obtain the score function related to (4)

$$\tilde{S}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\tilde{s}_t(\theta), \quad \tilde{s}_t(\theta) = \frac{Y_t-\tilde{\lambda}_t(\psi)}{\tilde{v}^*_t(\gamma)}\frac{\partial\tilde{\lambda}_t(\psi)}{\partial\theta} + \frac{[Y_t-\tilde{\lambda}_t(\psi)]^2-\tilde{v}^*_t(\gamma)}{2\tilde{v}^{*2}_t(\gamma)}\frac{\partial\tilde{v}^*_t(\gamma)}{\partial\theta}. \tag{13}$$

Then, define $L_T(\theta)$, $l_t(\theta)$, $S_T(\theta)$ and $s_t(\theta)$ as the random functions obtained from $\tilde{L}_T(\theta)$, $\tilde{l}_t(\theta)$, $\tilde{S}_T(\theta)$ and $\tilde{s}_t(\theta)$ by substituting $\tilde{\lambda}_t(\psi)$ and $\tilde{v}^*_t(\gamma)$ with $\lambda_t(\psi)$ and $v^*_t(\gamma)$, respectively. Furthermore, let $H(\theta_0) = \mathrm{E}[-\partial^2 l_t(\theta_0)/\partial\theta\partial\theta']$ and $I(\theta_0) = \mathrm{E}[s_t(\theta_0)s_t(\theta_0)']$. Consider the following assumptions.

A1 The process $\{Y_t, \lambda_t\}_{t\in Z}$ is strictly stationary and ergodic.

A2 $\lambda_t(\cdot)$ is continuous in Ψ, $v^*_t(\cdot)$ is continuous in Γ and the set Θ is compact. Moreover,

$$\mathrm{E}\sup_{\gamma\in\Gamma}\left|\log v^*_t(\gamma)\right| < \infty, \quad \mathrm{E}\sup_{\theta\in\Theta}\frac{[Y_t-\lambda_t(\psi)]^2}{v^*_t(\gamma)} < \infty.$$

A3 $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. if and only if $\psi = \psi_0$.

A4 There is a constant $\underline{v}^* > 0$ such that $v^*_t(\gamma), \tilde{v}^*_t(\gamma) \geq \underline{v}^*$ for any $t \geq 1$ and any $\gamma \in \Gamma$.

A5 Define $a_t = \sup_{\psi\in\Psi}|\tilde{\lambda}_t(\psi) - \lambda_t(\psi)|$ and $b_t = \sup_{\gamma\in\Gamma}\left|\tilde{v}^*_t(\gamma) - v^*_t(\gamma)\right|$, it holds that

$$\lim_{t\to\infty}\left(1+|Y_t| + \sup_{\psi\in\Psi}|\lambda_t(\psi)|\right)a_t = 0, \quad \lim_{t\to\infty}\left(1+Y_t^2 + \sup_{\psi\in\Psi}\lambda_t^2(\psi)\right)b_t = 0 \quad a.s.$$

A6 The pseudo-true parameter $\gamma^* \in \Gamma$ defined in (12) is unique.

A7 Define $c_t = \sup_{\theta\in\Theta}\|\partial\tilde{\lambda}_t(\psi)/\partial\theta - \partial\lambda_t(\psi)/\partial\theta\|$, $d_t = \sup_{\theta\in\Theta}\left\|\partial\tilde{v}^*_t(\gamma)/\partial\theta - \partial v^*_t(\gamma)/\partial\theta\right\|$. The following quantities are of order $\mathcal{O}(t^{-\delta})$ a.s. for some $\delta > 1/2$

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\psi)}{\partial\theta}\right\|a_t, \quad \sup_{\theta\in\Theta}\left\|\frac{\partial v^*_t(\gamma)}{\partial\theta}\right\|\left(1+|Y_t|+\sup_{\psi\in\Psi}|\lambda_t(\psi)|\right)a_t, \quad \sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\psi)}{\partial\theta}\right\|\left(|Y_t|+\sup_{\psi\in\Psi}|\lambda_t(\psi)|\right)b_t,$$

$$\sup_{\theta\in\Theta}\left\|\frac{\partial v^*_t(\gamma)}{\partial\theta}\right\|\left(1+Y_t^2+\sup_{\psi\in\Psi}\lambda_t^2(\psi)\right)b_t, \quad \left(1+|Y_t|+\sup_{\psi\in\Psi}|\lambda_t(\psi)|\right)c_t, \quad \left(1+Y_t^2+\sup_{\psi\in\Psi}\lambda_t^2(\psi)\right)d_t.$$

A8 $\lambda_t(\cdot)$ and $v^*_t(\cdot)$ have continuous second-order derivatives in their spaces. Moreover,

$$\mathrm{E}\sup_{\theta\in\Theta}\frac{[Y_t-\lambda_t(\psi)]^4}{v^{*2}_t(\gamma)} < \infty, \quad \mathrm{E}\sup_{\theta\in\Theta}\left\|\frac{1}{\sqrt{v^*_t(\gamma)}}\frac{\partial^2\lambda_t(\psi)}{\partial\theta\partial\theta'}\right\|^2 < \infty,$$

$$\mathrm{E}\sup_{\theta\in\Theta}\left\|\frac{1}{v^*_t(\gamma)}\frac{\partial\lambda_t(\psi)}{\partial\theta}\frac{\partial\lambda_t(\psi)}{\partial\theta'}\right\| < \infty, \quad \mathrm{E}\sup_{\theta\in\Theta}\left\|\frac{1}{v^*_t(\gamma)}\frac{\partial\lambda_t(\psi)}{\partial\theta}\frac{\partial v^*_t(\gamma)}{\partial\theta'}\right\|^2 < \infty,$$

$$\mathrm{E}\ \sup_{\theta\in\Theta}\left\|\frac{1}{v_t^{*2}(\gamma)}\frac{\partial v_t^*(\gamma)}{\partial\theta}\frac{\partial v_t^*(\gamma)}{\partial\theta'}\right\|^2<\infty\,,\quad \mathrm{E}\ \sup_{\theta\in\Theta}\left\|\frac{1}{v_t^*(\gamma)}\frac{\partial^2 v_t^*(\gamma)}{\partial\theta\partial\theta'}\right\|^2<\infty\,,$$
$$\mathrm{E}\ \frac{[Y_t-\lambda_t(\psi_0)]^8}{v_t^{*4}(\gamma^*)}<\infty\,,\quad \mathrm{E}\left\|\frac{1}{v_t^*(\gamma^*)}\frac{\partial\lambda_t(\psi_0)}{\partial\theta}\frac{\partial\lambda_t(\psi_0)}{\partial\theta'}\right\|^2<\infty\,.$$

A9 The matrices $H(\theta_0)$ and $I(\theta_0)$ are positive definite.
A10 $\theta_0\in\dot{\Theta}$, where $\dot{\Theta}$ is the interior of Θ.
A11 The sequence $\sqrt{T}S_T(\theta_0)$ obeys the central limit theorem.

The strict stationarity and ergodicity in assumption **A1** depends upon the model formulation in (1) and (2) and it can be established by means of different probabilistic approaches, see for instance Straumann and Mikosch (2006) and Debaly and Truquet (2021). Assumption **A2** is a standard moment condition. Assumption **A3** is required for the identification of the true parameter ψ_0. Assumptions **A5** and **A7** are needed to guarantee that the initialization of filters in (3) is asymptotically irrelevant. Assumption **A6** imposes the uniqueness of the pseudo-true parameter for the variance equation. In Corollary 3 below, we show that this assumption can be dropped if the researcher is not interested in the asymptotic normality of the estimator but only in the consistency. Assumption **A8** imposes moments on the second derivatives of the log-quasi-likelihood that are required for asymptotic normality to apply. Assumption **A9** is required to obtain the positive definiteness of the asymptotic covariance matrix of the estimators. This condition is left high-level for generality purposes. However, in Lemma 3 in Appendix A.2, we introduce some special cases and sufficient low-level conditions that verify the assumption. Assumption **A10** is the standard condition for asymptotic normality that the pseudo-true parameter value is in the interior of the parameter set. Finally, assumption **A11** is an high-level condition that a central limit theorem applies to the score. This condition is also left high-level for generality purposes since the score function $s_t(\theta_0)$ is not a martingale difference sequence, see Eq. (13). There are several alternative Central Limit Theorems (CLT) for non-martingale sequences and the choice of the most appropriate one is strongly dependent on the specific mean–variance model formulation. For example, CLTs appealing the concept of mixing processes or mixingales are widely available, see the surveys in Doukhan (1994), Bradley (2005) and White (1994). See also the proof of Theorem 4 below for an example in which the assumption is satisfied by appealing the CLT for α-mixing processes. Finally, in case of correct conditional variance specification then assumption **A11** can be dropped, see Corollary 2. Theorem 1 delivers the consistency and asymptotic normality of the unrestricted PVQMLE of the true parameter ψ_0.

Theorem 1. *Consider the unrestricted PVQMLE in (5). Under conditions **A1**–**A6***

$$\hat{\psi}\to\psi_0\,,\quad a.s.\quad T\to\infty\,. \tag{14}$$

*Moreover, if also **A7**–**A11** hold, as $T\to\infty$*

$$\sqrt{T}\left(\hat{\psi}-\psi_0\right)\xrightarrow{d}N(0,\Sigma_\psi)\,,\qquad \Sigma_\psi=H_\psi^{-1}(\theta_0)I_\psi(\theta_0)H_\psi^{-1}(\theta_0)\,, \tag{15}$$

where

$$H_\psi(\theta_0)=\mathrm{E}\left[\frac{1}{v_t^*(\gamma^*)}\frac{\partial\lambda_t(\psi_0)}{\partial\psi}\frac{\partial\lambda_t(\psi_0)}{\partial\psi'}\right],\ I_\psi(\theta_0)=\mathrm{E}\left[\frac{v_t}{v_t^{*2}(\gamma^*)}\frac{\partial\lambda_t(\psi_0)}{\partial\psi}\frac{\partial\lambda_t(\psi_0)}{\partial\psi'}\right]. \tag{16}$$

In addition, Σ_ψ is positive definite.

The asymptotic properties of the estimator of the pseudo-variance parameters γ are obtained from Theorem 1 as a byproduct. We make the result explicit in Corollary 1 below. Let $s_t(\theta_0)=[s_t^{(\psi)}(\theta_0)',s_t^{(\gamma)}(\theta_0)']'$ be the partition of the score with respect to the mean and (pseudo-)variance parameters. Define the partitions $H_\gamma(\theta_0)=\mathrm{E}[-\partial^2 l_t(\theta_0)/\partial\gamma\partial\gamma']$ and $I_\gamma(\theta_0)=\mathrm{E}[s_t^{(\gamma)}(\theta_0)s_t^{(\gamma)}(\theta_0)']$.

Corollary 1. *Under the assumptions of Theorem 1 we have that as $T\to\infty$, a.s. $\hat{\gamma}\to\gamma^*$ and $\sqrt{T}(\hat{\gamma}-\gamma^*)\xrightarrow{d}N(0,\Sigma_\gamma)$, where $\Sigma_\gamma=H_\gamma^{-1}(\theta_0)I_\gamma(\theta_0)H_\gamma^{-1}(\theta_0)$. In addition, Σ_γ is positive definite.*

Theorem 1 determines the asymptotic distribution of the unrestricted PVQMLE of ψ_0 without requiring correct specification of the pseudo-variance. The following result shows that in the special case in which the variance is well-specified then the estimator $\hat{\psi}$ gains in efficiency.

Corollary 2. *Consider the assumptions of Theorem 1. If, in addition, the variance (2) is correctly specified, i.e. $v_t^*(\gamma^*)=v_t$, then **A1**–**A10** entail (14) and*

$$\sqrt{T}\left(\hat{\psi}-\psi_0\right)\xrightarrow{d}N(0,I_\psi^{-1})\,,\qquad I_\psi=\mathrm{E}\left[\frac{1}{v_t}\frac{\partial\lambda_t(\psi_0)}{\partial\psi}\frac{\partial\lambda_t(\psi_0)}{\partial\psi'}\right], \tag{17}$$

where $\Sigma_\psi-I_\psi^{-1}$ is positive semi-definite.

We also note that in Corollary 2 the uniqueness of the variance parameter in assumption **A6** is implied by the condition $v_t^*(\gamma)=v_t^*(\gamma^*)$ a.s. if and only if $\gamma=\gamma^*$. This follows immediately from the correct specification of the pseudo-variance. Corollary 3 below shows that even if the pseudo-true parameter γ^* is not unique, i.e. assumption **A6** does not hold, the consistency of the unrestricted estimator $\hat{\psi}$ is retained without any additional assumption. The overall estimator $\hat{\theta}$ will instead be set consistent over the set of values that maximize the limit of the quasi-likelihood, Θ_0, since the pseudo-true parameter γ^* is not uniquely identified.

Corollary 3. *Consider the unrestricted PVQMLE (5) and assume conditions **A1**–**A5** hold. Then, as $T \to \infty$, $\inf_{\theta_0 \in \Theta_0} \|\hat{\theta} - \theta_0\| \to 0$ a.s. and $\hat{\psi} \to \psi_0$ a.s.*

We now treat the case in which the conditional mean and pseudo-variance parameters are constrained. We study the asymptotic properties of the restricted PVQMLE $\hat{\psi}_R$ defined in (6).

A12 The equality $S\gamma^* = g(\psi_0)$ holds and $g(\cdot)$ is continuous.

Assumption **A12** is required to ensure that $\theta_0 \in \Theta_R$, i.e. the imposed restrictions are valid with respect to the true parameter ψ_0 and the pseudo-true parameter γ^*. The continuity of $g(\cdot)$ guarantees that Θ_R remains compact. Define $\gamma = (\gamma_1', \gamma_2')'$ where $\gamma_1 = S\gamma = g(\psi)$ is the sub-vector of pseudo-variance parameters that are restricted to mean parameters and γ_2 constitutes the sub-vector of remaining free parameters. For $\theta \in \Theta_R$, we have $\theta = (\psi', \gamma_1', \gamma_2')' = (\psi', g(\psi)', \gamma_2')'$ so the m-dimensional vector of parameters to estimate is reduced to $\theta = (\psi', \gamma_2')'$, with some abuse of notation. The new parameter vector has dimension $m_R = p + k_2$ where k_2 is the length of the extra nuisance parameters γ_2. Recall that $H_x(\theta_0) = \mathrm{E}\left[-\partial^2 l_t(\theta_0)/\partial x \partial x'\right]$ and $I_x(\theta_0) = \mathrm{E}[s_t^{(x)}(\theta_0) s_t^{(x)}(\theta_0)']$. Moreover, define $H_{x,z}(\theta_0) = \mathrm{E}\left[-\partial^2 l_t(\theta_0)/\partial x \partial y'\right]$, $I_{x,z}(\theta_0) = \mathrm{E}[s_t^{(x)}(\theta_0) s_t^{(z)}(\theta_0)']$ and $I_{z,x}(\theta_0) = I_{x,z}'(\theta_0)$. Analogously, set $D(\theta_0) = H^{-1}(\theta_0)$ and $D_{x,y}(\theta_0)$ being the corresponding partition related to rows x and columns y of $D(\theta_0)$. Theorem 2 delivers the asymptotic distribution of the restricted PVQMLE.

Theorem 2. *Consider the restricted PVQMLE in (6). Under conditions **A1**–**A6** and **A12***

$$\hat{\psi}_R \to \psi_0 , \quad a.s. \quad T \to \infty . \tag{18}$$

*Moreover, if also **A7**–**A11** hold, as $T \to \infty$*

$$\sqrt{T}\left(\hat{\psi}_R - \psi_0\right) \xrightarrow{d} N(0, \Sigma_R) , \tag{19}$$

where

$$\Sigma_R = D_{\psi}(\theta_0) I_{\psi}(\theta_0) D_{\psi}(\theta_0) + D_{\psi,\gamma_2}(\theta_0) I_{\gamma_2,\psi}(\theta_0) D_{\psi}(\theta_0) + D_{\psi}(\theta_0) I_{\psi,\gamma_2}(\theta_0) D_{\gamma_2,\psi}(\theta_0) + D_{\psi,\gamma_2}(\theta_0) I_{\gamma_2}(\theta_0) D_{\gamma_2,\psi}(\theta_0) . \tag{20}$$

In addition, Σ_R is positive definite.

We note that Corollaries 1–3 can easily be adapted to hold also for $\hat{\theta}_R$. In Section 4.2 below, we shall see that the restricted PVQMLE can lead to substantial gains in efficiency with respect to the unrestricted PVQMLE. The consistency of the restricted PVQMLE requires the additional assumption **A12**. However, as discussed in Section 5, this assumption can be tested and the correct specification of the pseudo-variance is not required. Clearly, when ψ and γ do not have parameter restrictions, i.e. $\hat{\psi}_R = \hat{\psi}$, it can be noted that Theorem 1 is equivalent to Theorem 2 with $\Sigma_R = \Sigma_{\psi}$, since $H_{\psi,\gamma_2}(\theta_0) = 0$, $H(\theta_0)$ becomes block diagonal, its inverse has block elements $D_x(\theta_0) = H_x^{-1}(\theta_0)$ and $D_{x,y}(\theta_0) = D_{y,x}(\theta_0) = 0$, implying that $\Sigma_R = \Sigma_{\psi}$.

To illustrate the relevance of the theoretical results, in the remainder of the section we provide an application of the asymptotic results to two specific models of interest introduced in Example 1.

3.1. Integer-valued autoregressive models

We consider the class of INAR models specified in Eq. (7) with the corresponding conditional mean given in (8). Recall that for INAR models the thinning operator is defined as $a \circ N = \sum_{j=1}^{N} X_j$ when $N > 0$, and 0 otherwise, where $X_j \sim D_X(a, b)$ are iid with finite mean a and variance b. We start by studying the stochastic properties of the general class of INAR processes. Theorem 3 below provides conditions for strict stationarity and mixing properties of the INAR process.

Theorem 3. *Let the INAR process (7) satisfy $a < 1$. Then, the process admits a strictly stationary and ergodic solution with finite second moment $\mathrm{E}(Y_t^2) < \infty$. Moreover, the process is β-mixing with coefficients decaying geometrically fast.*

Next, we derive the strong consistency and asymptotic normality of several PVQML estimators of INAR models by appealing to Theorems 1–2. We assume that the observations are generated from an INAR(1) model with thinning and innovation distributions following some unspecified equidispersed distributions (i.e. mean equal to the variance). We consider PVQMLEs for the parameter vector based on the pseudo-variance specified in (9). We study the asymptotic properties of the unrestricted PVQMLE and the restricted PVQMLE with restrictions $\omega_2 = \omega_1$ and $b = a$. In this case, the restrictions hold but no assumptions on the shape of the distribution of the data generating process are imposed for the asymptotic results of the PVQMLE.

Theorem 4. *Let $\{Y_1, \ldots, Y_T\}$ be generated by the INAR(1) process in (7) with an equidispersed error, $\mathrm{E}(\varepsilon_t) = \mathrm{V}(\varepsilon_t) = \omega_1$, and an equidispersed thinning operator, $\mathrm{E}(a \circ N) = \mathrm{V}(a \circ N) = aN$, with $a < 1$. Consider PVQMLEs for the parameter vector $\theta = (\omega_1, a, \omega_2, b)'$ based on the pseudo-variance v_t^* specified in (9). Furthermore, assume that $\theta_0 \in \Theta$, where Θ is a compact parameter set such that $\omega_1 > 0$, $\omega_2 > 0$, $a \geq 0$, $b \geq 0$. Then, the unrestricted PVQMLE (5) and the following restricted PVQMLEs (6) with restrictions (i) $\omega_2 = \omega_1$, (ii) $b = a$ and (iii) $(\omega_2 = \omega_1, b = a)$ are strongly consistent. Assume further that $\theta_0 \in \dot{\Theta}$ and $\mathrm{E}(Y_t^8) < \infty$. Then, all the PVQMLEs are also asymptotically normally distributed with asymptotic covariance matrix given in Theorems 1–2.*

For the INAR model, the existence of the r-moments with $r > 2$ depends on the specific discrete distribution for the errors ε_t and the thinning. Since we keep such distributions unspecified, the existence of higher-order moments is required. However, the moment condition is satisfied for several INAR models. For example, when the thinning and the error distributions are Poisson the observation Y_t are Poisson marginally distributed and then the moments of any order are finite (Christou and Fokianos, 2014, Lem. A.1). For details on more general INAR modeling see Weiß (2018a). The result can straightforwardly be extended to INAR models with a general order p.

3.2. Double-bounded auto-regressive model

As a second illustration, we consider an application of the asymptotic results to double-bounded time series processes. We study a process that takes values in the unit interval $[0, 1]$, however, we note that the same results apply to the generic bounds $[L, U]$ as the observable process can be transformed to lie in the unit interval. We consider the following specification for the conditional mean and pseudo-variance of the PVQMLE

$$\begin{aligned} \lambda_t &= \omega_1 + \alpha_1 Y_{t-1} + \beta_1 \lambda_{t-1}, \\ v_t^* &= \frac{\mu_t(1-\mu_t)}{1+\phi}, \qquad \mu_t = \omega_2 + \alpha_2 Y_{t-1} + \beta_2 \mu_{t-1}, \end{aligned} \tag{21}$$

where the double-bounded nature of the data requires $0 < \omega_i + \alpha_i + \beta_i < 1$ for $i = 1, 2$ and $\phi > 0$. If the observable variable Y_t follows a conditional beta distribution with mean λ_t and dispersion parameter ϕ, $Beta(\lambda_t, \phi)$, then the conditional variance will take the form defined in (21) with $\mu_t = \lambda_t$. We assume this beta process as data generating process.

Theorem 5. *Assume the process $\{Y_t, \lambda_t\}_{t\in\mathbb{Z}}$ is generated by $Y_t|\mathcal{F}_{t-1} \sim Beta(\lambda_t, \phi)$ with conditional mean specified as in (21) and $\omega_1 + \alpha_1 + \beta_1 < 1$. Then, the process admits a strictly stationary and ergodic solution with finite moments of any order $\mathrm{E}(Y_t^r) < \infty$ for all $r \geq 1$.*

The results follows immediately by Gorgi and Koopman (2023, Thm. 2.1) and all moments exist since the time series is bounded. Next, we derive the strong consistency and asymptotic normality of the PVQML estimators of double-bounded autoregressions. We consider the case where the observations are generated from $Beta(\lambda_t, \phi)$ with mean specified as in (21). We study the asymptotic properties of the unrestricted PVQMLE and the restricted PVQMLE with restrictions $\omega_2 = \omega_1$, $\alpha_2 = \alpha_1$ and $\beta_2 = \beta_1$.

Theorem 6. *Let $\{Y_1, \ldots, Y_T\}$ be generated by a beta autoregressive process with conditional distribution $Y_t|\mathcal{F}_{t-1} \sim Beta(\lambda_t, \phi)$ where λ_t follows the recursive equation in (21). Consider PVQMLEs for the parameter vector $\theta = (\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2)'$ based on the pseudo-variance v_t^* specified in (21). Furthermore, assume that $\theta_0 \in \Theta$, where Θ is a compact parameter set such that $\omega_i > 0$, $\alpha_i > 0$, $\beta_i \geq 0$, $\omega_i + \alpha_i + \beta_i < 1$, $\phi > 0$, for $i = 1, 2$ and for any $\theta \in \Theta$. Then, the unrestricted PVQMLE (5) and the restricted PVQMLE (6) with restrictions $\omega_2 = \omega_1$, $\alpha_2 = \alpha_1$ and $\beta_2 = \beta_1$ are strongly consistent. Assume further that $\theta_0 \in \dot{\Theta}$. Then, both PVQMLEs are asymptotically normally distributed with asymptotic covariance matrix given in Theorems 1–2.*

4. Efficiency of the PVQMLE

4.1. Comparison to alternative estimators

In this section, we show that the unrestricted PVQMLE achieves the same asymptotic variance of existing estimators. Consider the unrestricted PVQMLE depicted in Theorem 1. The partition of the score related to the mean parameter ψ is

$$\tilde{s}_t^{(\psi)}(\theta) = \frac{Y_t - \tilde{\lambda}_t(\psi)}{\tilde{v}_t^*(\gamma)} \frac{\partial \tilde{\lambda}_t(\psi)}{\partial \psi}. \tag{22}$$

We compare (22) with some alternative semi-parametric estimators presented in the literature.

Consider the two-stage Weighted Least Squares (WLSE) of Aknouche and Francq (2023) defined as

$$\hat{\psi}_W = \operatorname*{argmax}_{\psi\in\Psi} \frac{1}{T} \sum_{t=1}^{T} \tilde{l}s_t(\psi, \hat{w}_t), \quad \tilde{l}s_t(\psi, \hat{w}_t) = -\frac{[Y_t - \tilde{\lambda}_t(\psi)]^2}{\hat{w}_t},$$

where $\hat{w}_t$ is a first-step estimator of the set of weights w_t. The resulting score of the WLSE is

$$\tilde{s}_t(\psi, \hat{w}_t) = \frac{Y_t - \tilde{\lambda}_t(\psi)}{\hat{w}_t} \frac{\partial \tilde{\lambda}_t(\psi)}{\partial \psi}. \tag{23}$$

Since it is well-known that the conditional variance is the optimal weight for the WLSE, the same authors set $w_t = v_t^*(\xi) = v^*(Y_{t-1}, Y_{t-2}, \ldots; \xi)$ by defining a functional form for a pseudo-variance, where the parameters ξ may also contain ψ_0 or parts of it. The corresponding first-step estimated weights are $\hat{w}_t = \tilde{v}_t^*(\hat{\xi})$, where $\hat{\xi}$ represents the first-step estimate of the parameter ξ.

Another related estimator is the Estimating Function (EF) approach for dynamic models that has been recently introduced by Francq and Zakoian (2023). In the case where only the conditional mean is correctly specified, the EF equation can be written as a slightly modified version of (23), by setting $\hat{w}_t = \tilde{v}_t^*(\tilde{\xi})$ and $\tilde{\xi} = (\psi, \hat{\zeta})$ where $\hat{\zeta}$ are first-step estimates of parameters not in common with the mean equation. The estimator is then defined as the solution of the following system of equation $\sum_{t=1}^{T} \tilde{s}_t(\psi, \hat{w}_t) = 0$.

Consider the general QMLE of Wedderburn (1974) and Gourieroux et al. (1984) based on the exponential family of quasi-likelihoods defined as

$$\hat{\psi}_Q = \underset{\psi \in \Psi}{\operatorname{argmax}}\, \tilde{l}_T(\psi),$$

where the log-quasi-likelihood $\tilde{l}_T(\psi)$ is a member of the one-parameter exponential family with respect to $\tilde{\lambda}_t(\psi)$. The corresponding score is given by

$$\tilde{s}_t(\psi) = \frac{Y_t - \tilde{\lambda}_t(\psi)}{\tilde{v}_t(\psi)} \frac{\partial \tilde{\lambda}_t(\psi)}{\partial \psi}, \tag{24}$$

where the conditional variance $\tilde{v}_t(\psi)$ is typically a function of the mean, i.e. $\tilde{v}_t(\psi) = h(\tilde{\lambda}_t(\psi))$ for some function $h(\cdot)$. For example, selecting the Poisson quasi-likelihood yields $\tilde{v}_t(\psi) = \tilde{\lambda}_t(\psi)$ (Ahmad and Francq, 2016), see Aknouche and Francq (2023, Sec. 2.2) for other examples.

The expressions of the scores in (22)–(24) highlight how the unrestricted PVQMLE is related to WLSE, EF estimator and the QMLE based on the exponential family. The main difference between the unrestricted PVQMLE and the QMLE with score in (24) is that the QMLE only considers the specification of the conditional mean and the conditional variance is a function of the conditional mean that is implied by the selected distribution in the exponential family. On the other hand, the unrestricted PVQMLE differs from the WLSE as the parameters are estimated jointly instead of a multi-step estimation. A similar difference applies between the unrestricted PVQMLE and the EF approach, which also estimates some of the variance parameters in a first stage. The unrestricted PVQMLE, the QMLE, the WLSE and the EF estimator enjoy the same consistency property for the mean parameters ψ_0 irrespective of the correct specification of the conditional variance. Furthermore, when they have the same specification of the conditional pseudo-variance, these estimators are asymptotically equivalent.

Corollary 4. *Assume Theorem 1 holds. Moreover, suppose the WLSE (23) with $w_t = v_t^*(\gamma^*)$ is consistent and asymptotically normal with limiting variance Σ_W. Then the unrestricted PVQMLE in (5) is asymptotically as efficient as the WLSE, meaning that $\Sigma_\psi = \Sigma_W$. In addition, if $v_t^*(\cdot) = v_t(\cdot)$, then $\Sigma_W = \Sigma_\psi = I_\psi^{-1}$.*

The result in Corollary 4 follows immediately from Theorem 1 and Corollary 2. Since the EF estimator still involves a two-step procedure, it is not surprising to see that the EF estimator has the same efficiency as the WLSE (Francq and Zakoian, 2023). Therefore, the results of Corollary 4 also hold for the EF approach. We also note that if Corollary 4 holds then also Corollaries 2.1–2.3 in Aknouche and Francq (2023) hold for the unrestricted PVQMLE. This has two direct consequences: (i) if the variance is well-specified, the unrestricted PVQMLE is asymptotically more efficient than the QMLE of ψ_0, if the variance implied by the exponential family is not the true one, and (ii) if the conditional distribution of Y_t comes from the exponential family, then the well-specified PVQMLE is asymptotically as efficient as the MLE of ψ_0.

We note that the comparison discussed so far only concerns the unrestricted PVQMLE. This asymptotic equivalence of the PVQMLE with respect to the WLSE, the EF estimator and the QMLE does not hold for the restricted PVQMLE. This can be noted from the form of the score function given in Eq. (13) and the fact that the partial derivative of $\tilde{v}_t^*(\gamma)$ with respect to ψ is no longer equal to zero. This partial derivative is non-zero also in the EF approach for the special case of correctly specified conditional variance. However, even in this special case of correct specification of the conditional variance, PVQMLEs differ from the EF approach as the latter assumes that $\tilde{v}_t^*$ only depends on the parameter ψ, i.e. no additional free parameters are allowed in the conditional variance equation, which is instead included in our approach. Below we discuss how the restricted PVQMLE can achieve higher efficiency compared to the unrestricted PVQMLE.

4.2. Some results on the efficiency of PVQMLE

Given that the PVQMLE with distinct parameters on mean and pseudo-variance is asymptotically equivalent to the WLSE for the mean parameters ψ_0 (Corollary 4), it may be expected that if the mean and pseudo-variance equations share common parameters in θ, i.e. ψ_0 and γ^* are not completely distinct so that $\theta_0 \in \Theta_R$, then the restricted PVQMLE in (6) could show improved efficiency over the unrestricted PVQMLE and the WLSE. It is not straightforward to prove this result in general but for the following special cases it is verified.

A13 $E(Y_t^4|\mathcal{F}_{t-1}) < \infty$ almost surely.

A14 Set $p = k = 1$ and $Y_t|\mathcal{F}_{t-1} \sim q(\lambda_t, v_t)$ where $q(\cdot)$ has kurtosis ≤ 3. One of the following conditions holds:

- **A14.a** $q(\cdot)$ is symmetric.
- **A14.b** The first derivatives of the functions $\lambda_t(\psi_0)$ and $v_t(\gamma_0)$ have the (opposite) same sign and $q(\cdot)$ is (positive) negative skewed.

Proposition 1. *Assume that Assumptions **A1**–**A14** hold with $v_t^*(\gamma^*) = v_t$. Moreover, suppose that the WLSE in (23) with $w_t = v_t$ is consistent and asymptotically normal with asymptotic variance I_ψ^{-1}. Then, the restricted PVQMLE in (6) is asymptotically more efficient than the unrestricted PVQMLE and the WLSE, i.e. $I_\psi^{-1} - \Sigma_R$ is positive semi-definite.*

The conditions stated in Assumptions A14 can be somewhat restrictive, however, we note that they are only sufficient conditions. In general, it is not straightforward to derive sharper theoretical conditions under which the restricted PVQMLE is more efficient than the unrestricted PVQMLE. However, for specific models, we can appeal to numerical methods to obtain the asymptotic covariance matrix of the two estimators and evaluate their relative efficiency.

We consider the INAR(1) model in (7) with binomial thinning and Poisson error distribution as an example. The unrestricted PVQMLE $\hat{\psi}$ is based on the following conditional mean and pseudo-variance equations

$$\lambda_t(\psi) = aY_{t-1} + \omega_1 \,, \qquad v_t^*(\gamma) = bY_{t-1} + \omega_2 \,, \tag{25}$$

where $\psi' = (a, \omega_1)$ and $\gamma' = (b, \omega_2)$. Instead, the restricted PVQMLE $\hat{\psi}_R$ imposes the restrictions $b = a(1-a)$ and $\omega_2 = \omega_1 = \omega$.

We focus on the analysis of the asymptotic variances of these estimators. To this aim, we simulate a long time series (T = 10,000) from the INAR(1) process (binomial thinning and Poisson errors) for different values of the parameters a and ω_1 over a grid. The asymptotic covariance matrices of the two estimators are computed by approximating their expectations with the corresponding sample means. Fig. 1 reports an heatmap plot of the ratio (in $\log_{10}$ scale) between the asymptotic variance of the unrestricted and the restricted PVQMLEs for the parameter estimates of a and ω_1. The regions of the parameter set where the $\log_{10}$-variance ratio is greater than zero, i.e. variance ratio is greater than one, indicate the parameter values for which the restricted estimator is more efficient of the unrestricted one, and vice versa. The pictures suggest that the restricted estimator $\hat{\psi}_R$ is more efficient than the unrestricted estimator $\hat{\psi}$ in most cases, except when a and ω_1 are close to zero. Furthermore, the lack of efficiency of the restricted PVQMLE in the green areas is showed to be minimal. For example, a $\log_{10}$-variance ratio around -0.05 indicates a variance ratio around 0.9. Therefore, for small values of a and ω_1 the two estimators are essentially equivalent. Instead, for larger values of a and ω_1, the variance ratio gets substantially larger with the unrestricted PVQMLE estimator having up to 30 times larger variance of the restricted one. This is further illustrated in Fig. 2, which displays a graph of cross-section of the $\log_{10}$-variance ratio for some fixed values of a and ω_1.

Another way to grasp the intuition behind the improved efficiency of the restricted PVQMLE comes from the literature on saddlepoint approximations (Daniels, 1954). Saddlepoint approximations are used to approximate a density function with a function

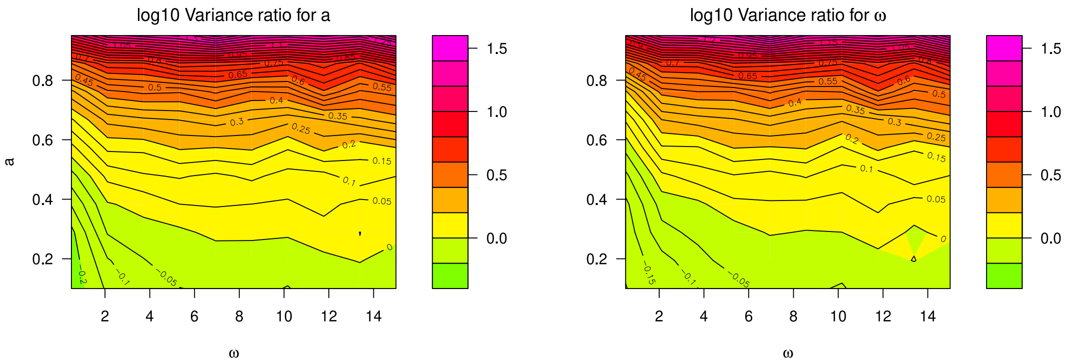

Fig. 1. Contour plots of $\log_{10}$-variance ratios for the INAR coefficients. Left: ratio $\log_{10}[Var(\hat{a})/Var(\hat{a}_R)]$ plotted for several values of a and ω. Right: ratio $\log_{10}[Var(\hat{\omega})/Var(\hat{\omega}_R)]$ plotted for several values of a and ω. The green area indicates a variance ratio smaller than one.

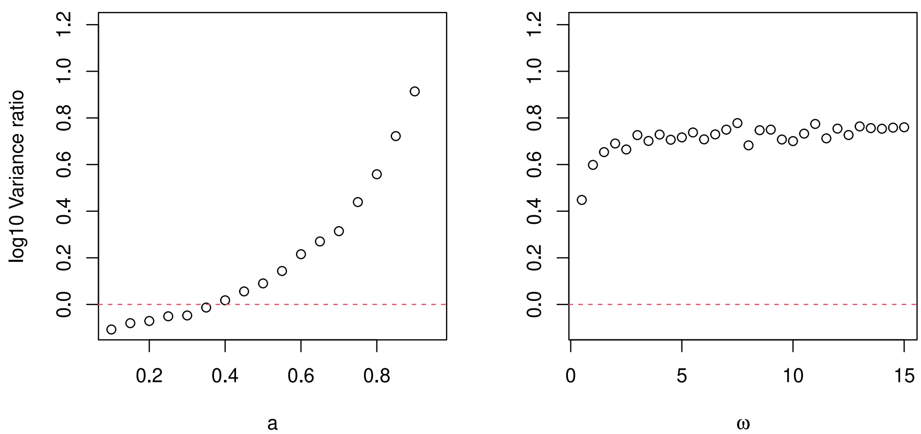

Fig. 2. $\log_{10}$-variance ratios plots for the INAR coefficients. Dashed red line: y-axis=0. Left: ratio $\log_{10}[Var(\hat{a})/Var(\hat{a}_R)]$ plotted for several values of a and $\omega = 3$. Right: ratio $\log_{10}[Var(\hat{\omega})/Var(\hat{\omega}_R)]$ plotted for several values of ω and $a = 0.85$.

Table 1
Bias and RMSE of estimators of the mean parameters when the data generating process is an INAR(1) with $a = 0.85$ and $\omega = 3$, and sample size $T = \{100, 500, 2000\}$.

	$T = 100$				$T = 500$				$T = 2000$			
	ω_1		a		ω_1		a		ω_1		a	
Est.	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\hat{\psi}_Q$	0.7462	1.4348	−0.0389	0.0736	0.1496	0.5218	−0.0077	0.0262	0.0323	0.2425	−0.0016	0.0121
$\hat{\psi}_{LS}$	0.7453	1.4391	−0.0389	0.0739	0.1503	0.5217	−0.0077	0.0262	0.0290	0.2401	−0.0014	0.0119
$\hat{\psi}_W$	0.7382	1.4319	−0.0385	0.0735	0.1475	0.5175	−0.0076	0.0260	0.0301	0.2392	−0.0015	0.0119
$\hat{\psi}_{WUN}$	0.7377	1.4318	−0.0385	0.0735	0.1474	0.5173	−0.0076	0.0260	0.0300	0.2393	−0.0015	0.0119
$\hat{\psi}$	0.7203	1.4235	−0.0376	0.0731	0.1452	0.5166	−0.0075	0.0260	0.0295	0.2388	−0.0015	0.0119
$\hat{\psi}_{R_1}$	0.7050	1.3837	−0.0368	0.0710	0.1417	0.4985	−0.0073	0.0250	0.0305	0.2314	−0.0015	0.0115
$\hat{\psi}_{R_2}$	0.5913	1.1980	−0.0313	0.0620	0.1316	0.5051	−0.0068	0.0255	0.0246	0.2332	−0.0012	0.0115
$\hat{\psi}_{R_3}$	0.0311	0.4586	−0.0028	0.0235	−0.0010	0.2036	−0.0002	0.0101	0.0027	0.1011	−0.0001	0.0049
$\hat{\psi}_{ML}$	0.0317	0.4551	−0.0028	0.0234	−0.0002	0.2018	−0.0002	0.0100	0.0029	0.1009	−0.0001	0.0049

that is based on the cumulant generating function of the data, which is typically called saddlepoint density. Pedeli et al. (2015) show that the conditional saddlepoint density can approximate the conditional density of the INAR(p) model in (7) to a certain degree of accuracy. It is not hard to see that the conditional saddlepoint density is approximately equal to the pseudo-variance quasi-likelihood in (4) with correctly specified variance (Pedeli et al., 2015, Sec. 3.4). Therefore, when the variance is correctly specified, the restricted PVQMLE of the INAR(p) model is close to the maximizer of the log-likelihood obtained by the saddlepoint density, which in turn is expected to get closer to the MLE as $\lambda_t \to \infty$. This is confirmed empirically from the results in Figs. 1 and 2, where the efficiency of restricted PVQMLE over the unrestricted PVQMLE grows as $a, w \to \infty$ i.e. where restricted PVQMLE approximates more accurately the MLE. We conjecture that similar results may apply also to other models. For the case of independent observations, Goodman (2022) has recently shown that the approximation error in using saddlepoint approximation is negligible compared to the inferential uncertainty inherent in the MLE. Although the literature is still under development, these arguments provide reliable evidence on the higher asymptotic performance of restricted PVQMLEs compared to the unrestricted one and other quasi-likelihood methods presented in Section 4.1.

Finally, we consider a simulation study to assess the small sample properties of PVQMLEs in comparison with several other alternative estimators. The study consists of 1000 Monte Carlo replications where we generate data from the Poisson INAR(1) process and estimate the mean parameter vector ψ. We consider several PVQMLEs based on different restrictions of the variance parameter vector γ. The unrestricted PVQMLE $\hat{\psi}$ is based on the mean and pseudo-variance equations in (25). The first restricted PVQMLE $\hat{\psi}_{R_1}$ imposes the restriction $R_1 : b = a(1-a)$, the second restricted PVQMLE $\hat{\psi}_{R_2}$ imposes the restriction $R_2 : \omega_2 = \omega_1$, and the third restricted PVQMLE $\hat{\psi}_{R_3}$ imposes the restriction $R_3 : b = a(1-a),\ \omega_2 = \omega_1$. Furthermore, we consider the QMLE based on the Poisson quasi-likelihood $\hat{\psi}_Q$, the conditional least squares estimator (CLSE) $\hat{\psi}_{LS}$, the WLSE in (23) with weights $\hat{w}_t = \hat{a}_{LS}(1-\hat{a}_{LS})Y_{t-1} + \hat{\omega}_{LS}$ where $(\hat{\omega}_{LS}, \hat{a}_{LS})' = \hat{\psi}_{LS}$ are first-step estimates obtained from the CLSE, and the unfeasible WLSE $\hat{\psi}_{WUN}$ with weights given by the true conditional variance. Finally, we also include the MLE $\hat{\psi}_{ML}$ for comparison purposes. The results of the simulation study are reported in Table 1.

Since the PVQMLE without constraints on the first two moments is asymptotically equivalent to the WLSE, it can be expected that the restricted PVQMLE where suitable constraints corresponding to the true model are imposed should show improved performances over the other quasi-likelihood-type estimators. Indeed, from Table 1 it can be seen that QMLE, CLSE, WLSE and unrestricted PVQMLE of model (25) share similar performances both in terms of bias and RMSE. Instead, a partial specification of the true constraints underlying the model in $\hat{\psi}_{R_1}$ and $\hat{\psi}_{R_2}$ already leads to an improvement with respect to the other estimation techniques; such improvement becomes substantial in $\hat{\psi}_{R_3}$ where all the correct constraints are considered. Moreover, this last restricted PVQMLE has comparable performance to the MLE. This is important since when $p \gg 1$ the MLE can become hard to compute and therefore our approach is a valid alternative.

5. Testing restrictions

In Section 3, we have seen that correctly identified constraints on mean and pseudo-variance equations can deliver a restricted PVQMLE with improved efficiency. In this section, we develop a test based on the unrestricted estimator in (5) which allows us to test the validity of the restriction $S\gamma = g(\psi)$. We define $r(\theta) = S\gamma - g(\psi)$ and we denote with $\Sigma(\theta_0) = H^{-1}(\theta_0)I(\theta_0)H^{-1}(\theta_0)$ the asymptotic covariance matrix of the entire unrestricted estimator vector $\hat{\theta}$. Moreover, consider the following plug-in estimators of $H(\theta_0)$ and $I(\theta_0)$ given by $\tilde{H}_T(\hat{\theta}) = T^{-1}\sum_{t=1}^T -\partial^2 \tilde{l}_t(\hat{\theta})/\partial\theta\partial\theta'$ and $\tilde{I}_T(\hat{\theta}) = T^{-1}\sum_{t=1}^T \tilde{s}_t(\hat{\theta})\tilde{s}_t'(\hat{\theta})$, respectively. The following result holds.

Proposition 2. *Assume that the assumptions of Theorem 1 hold. Consider the test $H_0 : r(\theta_0) = 0$ versus $H_1 : r(\theta_0) \neq 0$ where the function $r(\cdot)$ is continuously differentiable. Let $R(\theta) = \partial r(\theta)/\partial\theta'$. Then, under H_0, as $T \to \infty$*

$$W_T = T r'(\hat{\theta})\left[R(\hat{\theta})\Sigma(\theta_0)R'(\hat{\theta})\right]^{-1} r(\hat{\theta}) \xrightarrow{d} \chi^2_r,$$

where we can estimate $\Sigma(\theta_0)$ by $\tilde{\Sigma}_T(\hat{\theta}) = \tilde{H}_T^{-1}(\hat{\theta})\tilde{I}_T(\hat{\theta})\tilde{H}_T^{-1}(\hat{\theta})$.

The result follows immediately by the multivariate delta method, the continuous mapping theorem and standard asymptotic convergence arguments. Proposition 2 provides us a testing procedure for $H_0 : \theta_0 \in \Theta_R$ versus $H_1 : \theta_0 \notin \Theta_R$. It is worth nothing that the hypothesis test depicted in Proposition 2 does not require the variance of the model to be correctly specified. In the special case in which the pseudo-variance is correctly specified, then the test can be interpreted as a test of correct specification.

For example, consider the INAR(1) model in (7) with conditional mean and pseudo-variance equations as defined in Eqs. (8)–(9). We may consider the following test

$$H_0 : b = a(1-a) \quad \text{vs} \quad H_1 : b \neq a(1-a)\,, \tag{26}$$

which is a test for the assumption of a binomial thinning operator '$\circ$'. This follows from the definition of the INAR model in (7) as the autoregressive coefficient of the variance takes the form $b = a(1-a)$ under the assumption of binomial thinning. Alternative thinning specifications can be tested leading to a different form of the autoregressive variance parameter b, see Latour (1998) for the properties of INAR models with a general thinning specification. For instance, if we have a Poisson distribution for the thinning operator we have the restriction $b = a$. The corresponding test is

$$H_0 : b = a \quad \text{vs} \quad H_1 : b \neq a, \tag{27}$$

which assesses the validity of the assumption of equidispersion in the thinning operator versus either overdispersion or underdispersion.

We carry out a simulation study with 5000 Monte Carlo replications to assess the empirical size and power of the test of the parameter restrictions for the INAR(1) model. We consider the hypothesis in (27). To assess the size of the test we simulate under H_0 from a model with Poisson thinning operator and a Poisson distribution of the error term. Table 2 reports the results on the empirical size of the test. We can see that the test is slightly oversized for the smallest sample size, though still close to the nominal level, and it quickly becomes correctly sized as the sample size increases. Next, we evaluate the power of the test by simulating under the alternative. We consider a negative binomial thinning specification such that $a \circ N$ has a negative binomial distribution with mean aN and variance bN, $b = a + a^2/v$, where v is the dispersion parameter of the negative binomial. We note that this generates overdipersion in the thinning as $b = a + a^2/v > a$ and the smaller the parameter v the more the overdispersion. Fig. 3 shows the power of the test in (27) to reject the null hypothesis. As expected, we see that the power increases as the relative overdispersion $1 - a/b$ increases (v decreases) and as the sample size increases. Overall, the results show how the test has appropriate size and it has power against alternative hypotheses.

We also report additional simulation results in Appendix B that consider alternative data generating processes. First, we evaluate the robustness of the described test statistic by repeating the same simulation study with the inclusion of an outlier defined as 3 times the standard deviation of the observations plus their sample mean. The results show that the test is slightly oversized but the empirical size is still in line with nominal values. Moreover, the power of the test converges to 1 at a slightly slower rate

Table 2
Empirical size for test in (27). The model considered under H_0 is an INAR(1) model with Poisson thinning as well as Poisson error with parameter values $a = 0.75$ and $\omega = 1$.

Nominal size	T				
	100	250	500	1000	2000
0.1000	0.1222	0.1222	0.1140	0.1060	0.0986
0.0500	0.0720	0.0642	0.0582	0.0514	0.0518
0.0100	0.0202	0.0142	0.0114	0.0122	0.0128

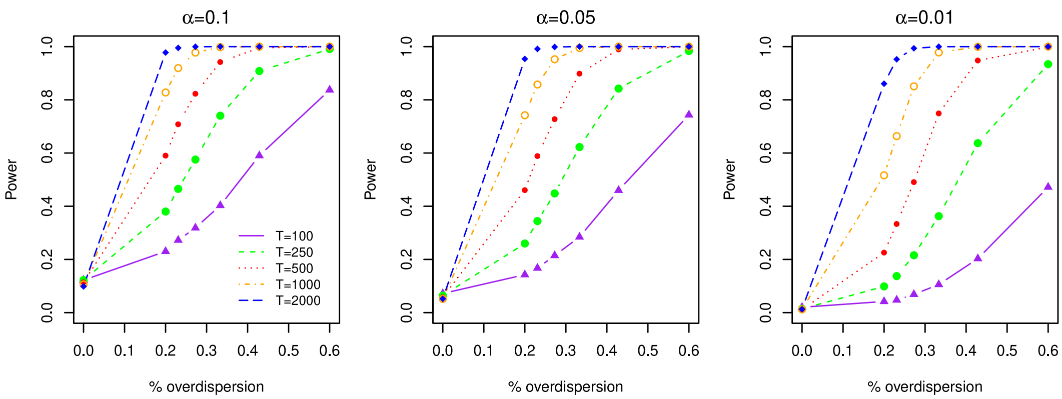

Fig. 3. Empirical power for test in (27). The true parameter values of the INAR(1) model with negative binomial thinning and Poisson error are $a = 0.75$ and $\omega = 1$. The value of the dispersion parameter v changes as indicated in the horizontal axis through the % of overdispersion: $1 - a/(a + a^2/v)$.

with the increasing sample size but it still performs satisfactorily. Second, we evaluate the test in case of near-unit root. The test seems conservative in this case, which can be due to the finite sample distribution of the estimators being different from the normal distribution near the unit-root boundary. The results on power indicate an adequate rejection rate when the sample size is large enough. Third, we evaluate the power of the equidispersion test under a different thinning specification. There are several thinning specifications available in the literature, see Ristić et al. (2013), Miletić Ilić (2016), Borges et al. (2016), Nastić et al. (2017), Borges et al. (2017), and Bourguignon et al. (2018), amongst others. We consider the Binomial-Negative Binomial (BiNB) thinning as in Bourguignon and Weiß (2017). The results show that the power increases as the relative overdispersion of the thinning increases. Finally, we consider the case of testing equidispersion of the thinning operator in an INAR(2) model. As expected, the results are comparable to the INAR(1) results with slower convergence rate towards the correct nominal size and power as the sample size grows. This is due to a more complex testing problem and larger set of parameters to be estimated.

6. Real data applications

In this section, we present two empirical applications where we employ PVQMLEs. We consider the test described in Section 5 to select appropriate parameter restrictions and compare different PVQMLEs. The first application concerns a dataset of crime counts, where the INAR model is considered for the specification of the conditional mean and the pseudo-variance. The second application concerns the realized correlation between two financial assets that forms a double-bounded time series, where we consider a beta autoregression for the specification of the conditional mean and the pseudo-variance.

6.1. INAR model for crime counts

We consider an empirical application to the monthly number of offensive conduct reports in the city of Blacktown, Australia, from January 1995 to December 2014. This dataset has been employed in several articles featuring the INAR(1) model (Gorgi, 2018; Leisen et al., 2019). The time series is displayed in Fig. 4. In the literature, the distributional structure of the INAR innovation term ε_t is typically allowed to be flexible or left unspecified but the thinning operator is typically considered to be binomial. We consider the test proposed in the previous section to formally test the validity of binomial thinning assumption as well as the dispersion of the error term. We obtain the unrestricted PVQMLE for the INAR conditional mean and pseudo-variance equations in (25) and test several restrictions based on the test in Proposition 2. We test for equidispersion in the error $H_0 : \omega_1 = \omega_2$, binomial thinning $H_0 : b = a(1 - a)$, Poisson thinning $H_0 : a = b$ and geometric thinning $H_0 : b = a + a^2$. As discussed in Latour (1998), INAR(p) models have the same autocorrelation structure as continuous-valued AR(p) models. In this case, we can focus on INAR(1) model as the residuals obtained from the one-lag unrestricted model appear uncorrelated.

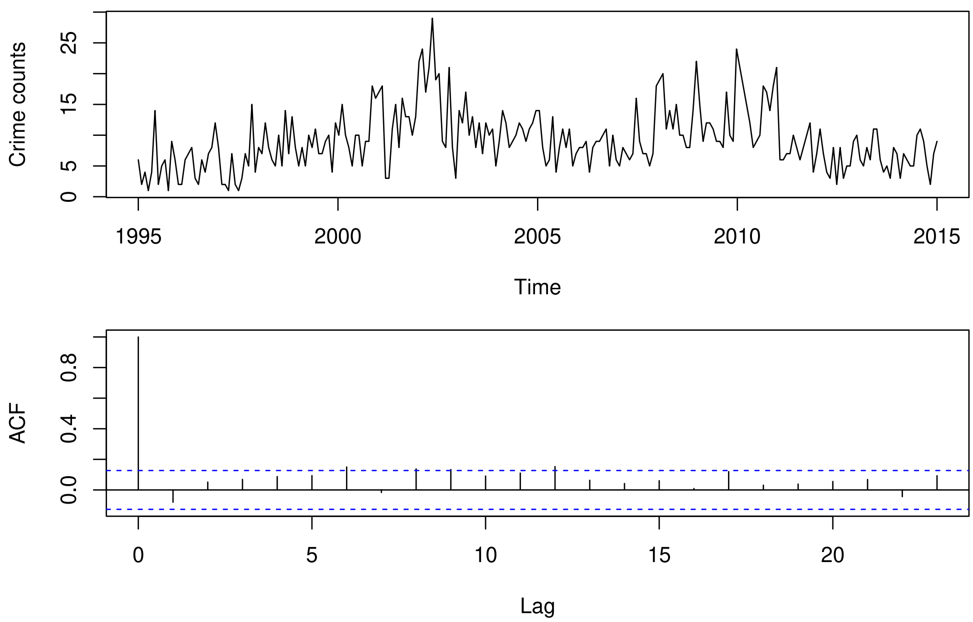

Fig. 4. Monthly number of offensive conduct reports in Blacktown, Australia, from January 1995 to December 2014. The second plot represents the sample autocorrelation function of the residuals obtained from the unrestricted estimator with 95% confidence bounds.

Table 3
p-values of the restriction tests for the INAR(1) model.

$\omega_1 = \omega_2$	Thinning		
	binomial	Poisson	geometric
0.372	0.005	0.043	0.229

Table 4
PVQMLEs of the INAR(1) model for the crime time series dataset. Standard errors in brackets.

	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{a}$	$\hat{b}$
Unrestricted	4.559	6.644	0.509	1.170
	(0.520)	(2.374)	(0.058)	(0.330)
Binomial thinning	6.280	–	0.371	–
	(0.434)		(0.040)	
Poisson thinning	4.820	–	0.524	–
	(0.523)	–	(0.058)	–
Geometric thinning	4.129	–	0.592	–
	(0.500)	–	(0.059)	–

The results of the tests are summarized in Table 3. We can see that the test does not reject the hypothesis of equidispersion in the error $\omega_1 = \omega_2$. As it concerns the tests on the thinning, the binomial and Poisson thinning are rejected at 5% significance level, instead, the geometric thinning is not rejected. This indicates that there is overdispersion in the thinning and the geometric one may be appropriate to describe the degree of overdispersion. Table 4 reports the estimation results for several PVQMLEs that are based on the different restrictions on the thinning operator. The standard errors are computed from the empirical counterparts of the asymptotic covariance matrices Eqs. (15) and (20) for the unrestricted and the restricted estimators, respectively. We can see that restricting to a binomial thinning leads to substantially biased estimates with respect to the unrestricted PVQMLE. Instead, from the geometric thinning we do not have such bias and the estimator can be expected to have an higher efficiency. We have also considered a BiNB thinning and it yields to equivalent estimation results as the geometric thinning. This follows as the BiNB thinning nests the geometric thinning and the estimated Bernoulli probability of the BiNB thinning is equal to zero, leading to a geometric thinning.

6.2. Double-bounded autoregression for realized correlation

The second application we present concerns the modeling of daily realized correlations between Boeing and Honeywell stocks as considered in Gorgi and Koopman (2023). Fig. 5 reports the plot of the time series. The sample size is $T = 2515$. Realized correlation measures take values in the interval $[-1, 1]$ and the transformation $Y_t/2 + 1/2$ is applied to rescale the realized correlation in the unit interval $[0, 1]$. We refer to Gorgi and Koopman (2023) for a discussion on how models on the unit interval can be extended to a general interval with known bounds.

We consider the specification for the conditional mean and pseudo-variance defined in (21). Besides the unrestricted PVQMLE, we consider a restricted PVQMLE with $\omega_2 = \omega_1$, $\alpha_2 = \alpha_1$, $\beta_1 = \beta_2$, which implies $\mu_t = \lambda_t$. These restrictions impose that the pseudo-variance is equal to the conditional variance implied by a beta distribution with mean parameter λ_t and precision parameter ϕ. In

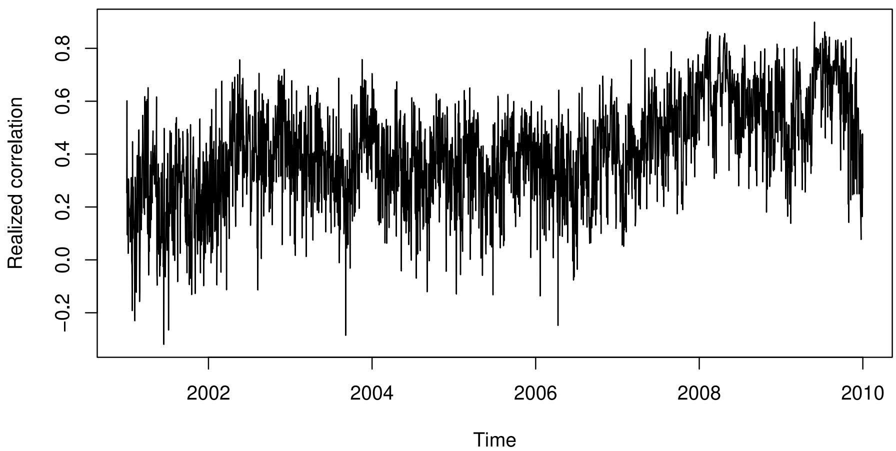

Fig. 5. Daily time series of realized correlations between Boeing (BA) and Honeywell (HON) asset returns, from January 2001 to December 2010.

Table 5
Estimation results for the realized correlation series. Standard errors in brackets. The bottom of the table reports the p-values of the tests on the restrictions.

	$\hat{\omega}_1$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\phi}$	$\hat{\omega}_2$	$\hat{\alpha}_2$	$\hat{\beta}_2$
Unrestricted	0.01	0.163	0.822	22.226	0.055	0.045	0.898
	(0.003)	(0.013)	(0.015)	(2.745)	(0.019)	(0.007)	(0.022)
Restricted	0.01	0.161	0.826	36.963	–	–	–
	(0.003)	(0.013)	(0.015)	(1.073)			
H_0	$\omega_1 = \omega_2$	$\alpha_1 = \alpha_2$	$\beta_1 = \beta_2$	joint test			
p-value	0.02	<0.001	0.01	<0.001			

this way, we can also test the adequacy of the beta autoregression for modeling the analyzed data through the specification test on the restriction. Table 5 reports the estimation results together with the restriction tests. We can see that the specification test rejects the null hypothesis of equality for the estimated α coefficients. For the same reason also the null assumption of the combined joint test is rejected. However, the null hypothesis is instead not rejected for ω and β coefficients at 1% level. This leans in favor of the restricted PVQMLE. We also notice that the estimated coefficients and the corresponding standard errors of the restricted PVQMLE are fairly close to the ones obtained from the beta autoregression reported in Table 1 of Gorgi and Koopman (2023).

7. Conclusions

We have introduced a novel methodology for the estimation of a broad range of semi-parametric time series models, where only the conditional mean is correctly specified by a parametric function. Our proposed PVQMLE is based on a Gaussian quasi-likelihood function and relies on the specification of a parametric pseudo-variance, which does not need to be the true conditional variance of the process and it may include restrictions on parameters related to the conditional expectation. We have established the asymptotic properties of the PVQMLE estimator with and without restrictions on the parameter space, and derived a test to validate the parameter restrictions. Importantly, our findings hold regardless of the correct specification of the pseudo-variance. A significant advantage of our restricted estimators is their potential to achieve greater efficiency compared to other quasi-likelihood methods found in existing literature. Additionally, our testing approach enables the development of specification tests for parametric time series models. We have demonstrated the practical application of our methodology through simulation studies and empirical cases.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We greatly appreciate comments made by the Editor, Associate Editor and two reviewers on an earlier version of the manuscript. Mirko Armillotta acknowledges financial support from the EU Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No. 101108797.

Appendix A. Proofs of results

A.1. Proofs

Proof of Theorem 1. Let $L(\theta) = \mathrm{E}[l_t(\theta)]$ be the limit log-quasi-likelihood. In what follows we show the following intermediate results.

(i) Uniform convergence: $\sup_{\theta\in\Theta} |\tilde{L}_T(\theta) - L(\theta)| \to 0$ almost surely, as $T \to \infty$.
(ii) Identifiability: the pseudo-true parameter value θ_0 is the unique maximizer of $L(\theta)$, i.e. $\mathrm{E}\left[l_t(\theta)\right] < \mathrm{E}\left[l_t(\theta_0)\right]$ for all $\theta \in \Theta, \theta \neq \theta_0$.

In order to prove (i) the uniform convergence of the two summands of (A.1) should be shown.

$$|\tilde{L}_T(\theta) - L(\theta)| \leq |\tilde{L}_T(\theta) - L_T(\theta)| + |L_T(\theta) - L(\theta)| \,. \tag{A.1}$$

The first term converges uniformly by Lemma 1 in Appendix A.2, under A4–A5, implying that the starting value of the process is asymptotically unimportant for the quasi-likelihood contribution. By assumption A1 the log-quasi-likelihood contribution $l_t(\theta)$ is stationary and ergodic. Moreover, it is uniformly bounded

$$\mathrm{E}\sup_{\theta\in\Theta}\left|l_t(\theta)\right| \le \frac{1}{2}\mathrm{E}\sup_{\gamma\in\Gamma}\left|\log v_t^*(\gamma)\right| + \frac{1}{2}\mathrm{E}\sup_{\theta\in\Theta}\left(\frac{\left[Y_t-\lambda_t(\psi)\right]^2}{v_t^*(\gamma)}\right) < \infty$$

by assumption A2. For the continuity of the quasi-likelihood and the compactness of Θ, Straumann and Mikosch (2006, Thm. 2.7) applies providing the uniform convergence of the second term in (A.1); in symbols $\sup_{\theta\in\Theta}|L_T(\theta)-L(\theta)| \to 0$ almost surely, as $T\to\infty$. This concludes the proof of (i).

We now prove (ii). First note that by the uniform limit theorem $L(\theta)=\mathrm{E}[l_t(\theta)]$ is a continuous function and it attains at least a maximum in Θ since Θ is compact. We now prove that such maximum is unique so that it can be univocally identified. Recall that $\theta=(\psi',\gamma')'$, assumption A2 provides $\mathrm{E}\sup_{\psi\in\Psi}\left|l_t(\psi,\gamma)\right| < \infty$ and $\mathrm{E}\sup_{\gamma\in\Gamma}\left|l_t(\psi_0,\gamma)\right| < \infty$ so also the function $l_t(\psi,\gamma)$ has at least a maximum for $\psi\in\Psi$, and $l_t(\psi_0,\gamma)$ has at least a maximum for $\gamma\in\Gamma$. Consider now $\mathrm{E}\left\{l_t(\theta)-l_t(\theta_0)\right\} = \mathrm{E}\left\{l_t(\theta)-l_t(\psi_0,\gamma)\right\}+\mathrm{E}\left\{l_t(\psi_0,\gamma)-l_t(\theta_0)\right\}$. The first summand is bounded as follows,

$$\mathrm{E}\left\{l_t(\theta)-l_t(\psi_0,\gamma)\right\} = \mathrm{E}\left\{-\frac{\mathrm{E}[\left(Y_t-\lambda_t(\psi)\right)^2|\mathcal{F}_{t-1}]}{2v_t^*(\gamma)}+\frac{v_t}{2v_t^*(\gamma)}\right\} \le \mathrm{E}\left\{-\frac{v_t}{2v_t^*(\gamma)}+\frac{v_t}{2v_t^*(\gamma)}\right\}=0$$

with equality if and only if $\psi=\psi_0$ by assumption A3. Moreover, $\mathrm{E}\left\{l_t(\psi_0,\gamma)-l_t(\theta_0)\right\}=\mathrm{E}\left[l_t(\psi_0,\gamma)\right]-\mathrm{E}\left[l_t(\psi_0,\gamma^*)\right]\le 0$ by assumption A6. This concludes the proof of (ii). The consistency of the whole estimator $\hat\theta$ follows from (i), (ii) and the compactness of Θ by Pötscher and Prucha (1997, Lemma 3.1). This implies (14).

To prove the asymptotic normality of the estimator we establish additional intermediate results.

(a) $\sqrt{T}\sup_{\theta\in\Theta}\|S_T(\theta)-\tilde S_T(\theta)\|\to 0$ almost surely, as $T\to\infty$.
(b) Define $H_T(\theta)=T^{-1}\sum_{t=1}^T -\partial^2 l_t(\theta)/\partial\theta\partial\theta'$. $H_T(\theta)\to H(\theta)$ almost surely uniformly over $\theta\in\Theta$, as $T\to\infty$.
(c) $\mathrm{E}[s_t(\theta_0)]=0$.

The condition (a) is satisfied by Lemma 2 in Appendix A.2, under A4–A5 and A7 implying that initial values of the process do not affect the asymptotic distribution of the PVQMLE.

Consider the second derivative of the log-quasi-likelihood contribution.

$$\begin{aligned}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} &= \left(\frac{1}{2v_t^{*2}(\gamma)}-\frac{[Y_t-\lambda_t(\psi)]^2}{v_t^{*3}(\gamma)}\right)\frac{\partial v_t^*(\gamma)}{\partial\theta}\frac{\partial v_t^*(\gamma)}{\partial\theta'}-\frac{Y_t-\lambda_t(\psi)}{v_t^{*2}(\gamma)}\left(\frac{\partial\lambda_t(\psi)}{\partial\theta}\frac{\partial v_t^*(\gamma)}{\partial\theta'}-\frac{\partial v_t^*(\gamma)}{\partial\theta}\frac{\partial\lambda_t(\psi)}{\partial\theta'}\right)\\ &\quad -\frac{1}{v_t^*(\gamma)}\frac{\partial\lambda_t(\psi)}{\partial\theta}\frac{\partial\lambda_t(\psi)}{\partial\theta'}+\frac{Y_t-\lambda_t(\psi)}{v_t^*(\gamma)}\frac{\partial^2\lambda_t(\psi)}{\partial\theta\partial\theta'}+\left(\frac{[Y_t-\lambda_t(\psi)]^2}{2v_t^{*2}(\gamma)}-\frac{1}{2v_t^*(\gamma)}\right)\frac{\partial^2 v_t^*(\gamma)}{\partial\theta\partial\theta'}.\end{aligned}\tag{A.2}$$

Assumptions A8 and the Cauchy–Schwarz inequality yield $\mathrm{E}\sup_{\theta\in\Theta}\left|\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j\right| < \infty$ for all $i,j=1,\dots,m$. Furthermore, the second derivative is a continuous, stationary and ergodic sequence. Then, an application of Straumann and Mikosch (2006, Thm. 2.7) provides condition (b). Note that since in this case $\partial\lambda_t(\psi)/\partial\gamma=\partial v_t^*(\gamma)/\partial\psi=0$ the matrix $H(\theta_0)$ is block diagonal with diagonal block matrices $H_\psi(\theta_0)=\mathrm{E}\left[-\partial^2 l_t(\theta_0)/\partial\psi\partial\psi'\right]$ and $H_\gamma(\theta_0)=\mathrm{E}\left[-\partial^2 l_t(\theta_0)/\partial\gamma\partial\gamma'\right]$. The former is defined in (16).

For establishing the asymptotic normality of the estimator $\hat\theta$ the proof of (c) is needed. Let $s_t(\theta_0)=[s_t^{(\psi)}(\theta_0)',s_t^{(\gamma)}(\theta_0)']'$ be the partition of the score between mean and pseudo-variance parameters. Observe that $\mathrm{E}(s_t^{(\psi)}(\theta_0)|\mathcal{F}_{t-1})=0$ but $\mathrm{E}\left(s_t(\theta_0)|\mathcal{F}_{t-1}\right)\neq 0$. Note that $\sup_{\theta\in\Theta}\left|\partial l_t(\theta)/\partial\theta_i\right| \le 2\left[\sup_{\theta\in\Theta}\left|l_t(\theta)\right|\right]^{1/2}\left[\sup_{\theta\in\Theta}\left|\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_i\right|\right]^{1/2}$, by Rudin (1976, p. 115). Moreover, $\mathrm{E}\sup_{\theta\in\Theta}\left|l_t(\theta)\right|<\infty$, and $\mathrm{E}\sup_{\theta\in\Theta}\left|\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j\right|<\infty$. Then an application of Cauchy–Schwarz inequality entails $\mathrm{E}\sup_{\theta\in\Theta}\left|\partial l_t(\theta)/\partial\theta_i\right|<\infty$. Finally, $\left\|\partial l_t(\theta)/\partial\theta\right\|\le\sup_{\theta\in\Theta}\left\|\partial l_t(\theta)/\partial\theta\right\|$ and an application of the dominated convergence theorem leads to $\mathrm{E}\left[\partial l_t(\theta)/\partial\theta\right]=\partial\mathrm{E}\left[l_t(\theta)\right]/\partial\theta$. By noting that θ_0 is the unique maximizer of $\mathrm{E}\left[l_t(\theta)\right]$ the result (c) follows.

Using the formula (13) some tedious algebra allows to show that $\mathrm{E}\left\|s_t(\theta_0)s_t(\theta_0)'\right\| < \infty$, by A8 and an application of Cauchy–Schwarz inequality. Therefore $I(\theta_0)=\mathrm{E}\left[s_t(\theta_0)s_t(\theta_0)'\right]$ is finite.

For T large enough $\hat\theta\in\dot\Theta$ by A10, so the following derivatives exist almost surely

$$0=\sqrt{T}\tilde S_T(\hat\theta)=\sqrt{T}S_T(\hat\theta)+o_p(1)=\sqrt{T}S_T(\theta_0)-H_T(\bar\theta)\sqrt{T}(\hat\theta-\theta_0)+o_p(1),$$

where the first equality comes from the definition (4), the second equality holds by (a), and the third equality is obtained by Taylor expansion at θ_0 with $\bar\theta$ lying between $\hat\theta$ and θ_0. By assumption A11 and (c) we have $\sqrt{T}S_T(\theta_0)\xrightarrow{d}N(0,I(\theta_0))$. This fact and (b) establish the asymptotic normality of the estimator $\hat\theta$ with covariance matrix $\Sigma(\theta_0)=H^{-1}(\theta_0)I(\theta_0)H^{-1}(\theta_0)$ by assumption A9, where

$$H(\theta_0)=\begin{pmatrix}H_\psi(\theta_0) & 0\\ 0 & H_\gamma(\theta_0)\end{pmatrix},\quad I(\theta_0)=\begin{pmatrix}I_\psi(\theta_0) & I_{\psi,\gamma}(\theta_0)\\ I_{\psi,\gamma}(\theta_0)' & I_\gamma(\theta_0)\end{pmatrix},\tag{A.3}$$

with $H_x(\theta_0)=\mathrm{E}\left[-\partial^2 l_t(\theta_0)/\partial x\partial x'\right]$, $I_x(\theta_0)=\mathrm{E}[s_t^{(x)}(\theta_0)s_t^{(x)}(\theta_0)']$ and $I_{x,z}(\theta_0)=\mathrm{E}[s_t^{(x)}(\theta_0)s_t^{(z)}(\theta_0)']$. In particular, standard algebra shows that $I_\psi(\theta_0)$ equals (16). See also Eq. (22). A suitable block matrix multiplication of (A.3) provides

$$\Sigma(\theta_0)=\begin{pmatrix}\Sigma_\psi(\theta_0) & \Sigma_{\psi,\gamma}(\theta_0)\\ \Sigma_{\psi,\gamma}(\theta_0)' & \Sigma_\gamma(\theta_0)\end{pmatrix},$$

where $\Sigma_\psi(\theta_0)$ takes the form defined in (15). In addition, note that for the marginal property of the multivariate Gaussian distribution result (15) holds with covariance matrix Σ_ψ being the partition of $\Sigma(\theta_0)$ for the mean parameters ψ.

The positive definiteness of the matrix $\Sigma(\theta_0)$ follows since for all $\delta \in \mathbb{R}^m$, with $\delta \neq 0$, we have $H(\theta_0)^{-1}\delta \neq 0$ as $H(\theta_0)^{-1}$ is full rank by **A9**. Now by setting $\eta = H(\theta_0)^{-1}\delta$ we have that $\eta' I(\theta_0)\eta > 0$ by **A9**. Therefore, it follows that $\delta' H(\theta_0)^{-1} I(\theta_0) H(\theta_0)^{-1}\delta > 0$. The principal submatrices of $\Sigma(\theta_0)$ are also positive definite. □

Proof of Corollary 2. Condition **A11** is not required since in this case is easily showed by (13) that $\mathrm{E}\left(s_t(\theta_0)|\mathcal{F}_{t-1}\right) = 0$. Recall that $\sqrt{T} s_T(\theta_0) = T^{-1/2}\sum_{t=1}^T U_t$ where $U_t = s_t(\theta_0)$. Note that $\{U_t, \mathcal{F}_t\}$ is a stationary martingale difference, and due to **A8**–**A9** it has a finite and positive definite second moments matrix. Then **A11** follows by the central limit theorem for martingales (Billingsley, 1961) and the Cramér–Wold device. The consistency and asymptotic normality of $\hat{\theta}$ follow as above. Finally, in view of (22) and $\mathrm{E}(s_t^{(\psi)}(\theta_0)|\mathcal{F}_{t-1}) = 0$

$$\mathrm{Var}\left[H_\psi^{-1}(\theta_0)s_t^{(\psi)}(\theta_0) - I_\psi^{-1}(\theta_0)s_t^{(\psi)}(\theta_0)\right] = \Sigma_\psi - I_\psi$$

being necessarily positive semi-definite. □

Proof of Corollary 3. Analogously to the proof of Theorem 1, **A1**–**A5** guarantee that $L_t(\theta)$ is continuous and a.s. uniformly convergent to $\mathrm{E}[l_t(\theta)]$. By recalling that Θ is compact the result follows by Pötscher and Prucha (1997, Lemma 4.2). □

Proof of Theorem 2. The consistency of $\hat{\theta}_R$ follows from the fact that by the proof of Theorem 1 we have that $\mathrm{E}[l_t(\psi,\gamma)] \leq \mathrm{E}[l_t(\psi_0,\gamma^*)]$ for any $\theta \in \Theta$ with equality holding only if $\theta = (\psi_0{}', \gamma^{*\prime})'$, and assumption **A12** ensures that $(\psi_0{}', \gamma^{*\prime})' \in \Theta_R$ with $\Theta_R \subseteq \Theta$. The consistency in (18) follows. The asymptotic normality of the estimator $\hat{\theta}_R$ follows as in the proof of Theorem 1 with covariance matrix $\Sigma(\theta_0) = H^{-1}(\theta_0)I(\theta_0)H^{-1}(\theta_0)$. In this case Hessian and Fisher information matrices can be written in the following block matrix form

$$H(\theta_0) = \begin{pmatrix} H_\psi(\theta_0) & H_{\psi,\gamma_2}(\theta_0) \\ H_{\psi,\gamma_2}(\theta_0)' & H_{\gamma_2}(\theta_0) \end{pmatrix}, \quad I(\theta_0) = \begin{pmatrix} I_\psi(\theta_0) & I_{\psi,\gamma_2}(\theta_0) \\ I_{\psi,\gamma_2}(\theta_0)' & I_{\gamma_2}(\theta_0) \end{pmatrix}. \tag{A.4}$$

Moreover, recall that

$$H^{-1}(\theta_0) = D(\theta_0) = \begin{pmatrix} D_\psi(\theta_0) & D_{\psi,\gamma_2}(\theta_0) \\ D_{\psi,\gamma_2}(\theta_0)' & D_{\gamma_2}(\theta_0) \end{pmatrix}. \tag{A.5}$$

By computing $\Sigma(\theta_0)$ using the block matrix multiplication as defined in (A.4) and (A.5) the partition of $\Sigma(\theta_0)$ for the mean parameters ψ equals Σ_R. This entails (19). □

Proof of Theorem 3. The result follows by a combination of Doukhan et al. (2012, Thm. 1–2) and the results of Doukhan et al. (2012, Sec. 4.1) given that $X_j \sim D_X(a,b)$ and $\mathrm{E}(X_j) = a$. Then the process is stationary, ergodic and $\mathrm{E}(Y_t) < \infty$. The same results show that the process is β-mixing with geometrically decaying coefficients. Finally, following Latour (1997, Sec. 3) we conclude that $\mathrm{E}(Y_t^2) < \infty$. □

Proof of Theorem 4. To prove the results we have to prove conditions **A1**–**A12** for the specified model. First note that since the pseudo-variance v_t^* defined in (9) is correctly specified we have that $v_t^*(\cdot) = v_t(\cdot)$. Moreover, **A12** holds. The condition **A1** holds by Theorem 3. **A4** holds since a.s. $v_t(\gamma) \geq \omega_2$. Note that a.s. $\sup_{\gamma\in\Gamma}|\log v_t(\gamma)| \leq \sup_{\gamma\in\Gamma}(v_t(\gamma)+1)/\min\{\omega_2, 1\}$ and $\sup_{\theta\in\Theta}(Y_t - \lambda_t(\psi))^2/v_t(\gamma) \leq (2Y_t^2 + 2\sup_{\psi\in\Psi}\lambda_t^2(\psi))/\omega_2$. By the c_p inequality it holds that $\mathrm{E}\sup_{\psi\in\Psi}\lambda_t^r(\psi) < \infty$ and $\mathrm{E}\sup_{\gamma\in\Gamma}v_t^r(\gamma) < \infty$ for $r \leq 2$ so the moments in **A2** are finite.

We prove **A3** by contradiction. We have that a.s. $\lambda_t(\psi) - \lambda_t(\psi_0) = \omega_1 - \omega_{1,0} + (a - a_0)Y_{t-1}$. If $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. with $\omega_1 \neq \omega_{1,0}$ then $0 \neq \omega_{1,0} - \omega_1 = (a - a_0)Y_{t-1}$ a.s. and the equality will be possible only if $(a - a_0) \neq 0$ and Y_{t-1} equals a.s. a non-zero constant. However, Y_{t-1} is non-constant. Therefore, if $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. then $\omega_1 = \omega_{1,0}$ and $0 = (a - a_0)Y_{t-1}$. Now to have $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. with $a \neq a_0$ we shall have that $Y_{t-1} = 0$ a.s. but this is impossible since Y_{t-1} is non-constant. Hence, if $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. then $a = a_0$. Analogous results hold for $v_t(\gamma)$.

Assumptions **A5**–**A7** are trivially satisfied here since λ_t and v_t are initialized using the first observation of the sample so $\lambda_t(\cdot) = \tilde{\lambda}_t(\cdot)$ and $v_t(\cdot) = \tilde{v}_t(\cdot)$. Condition **A6** is verified since γ^* is the true parameter vector of the variance, say γ_0, so $v_t(\gamma_0) = v_t$ a.s. and

$$\mathrm{E}\left[l_t(\psi_0,\gamma) - l_t(\psi_0,\gamma_0)\right] = \mathrm{E}\left[\frac{1}{2}\log\frac{v_t}{v_t(\gamma)} + \frac{1}{2} - \frac{v_t}{2v_t(\gamma)}\right] \leq \mathrm{E}\left[\frac{v_t}{2v_t(\gamma)} - \frac{1}{2} + \frac{1}{2} - \frac{v_t}{2v_t(\gamma)}\right] = 0$$

where the inequality follows by $\log(x) \leq x - 1$ for $x > 0$. So $\mathrm{E}\left[l_t(\psi_0,\gamma)\right] \leq \mathrm{E}\left[l_t(\psi_0,\gamma_0)\right]$ with equality if and only if (henceforth, iff) $v_t(\gamma) = v_t(\gamma_0)$ a.s. but by **A3** this happens iff $\gamma = \gamma_0$. Therefore γ_0 is unique maximizer of (12). Let 0_k be a $k \times 1$ vector of zeros. Recall that $\partial\lambda_t(\psi)/\partial\theta = (1, Y_{t-1}, 0_2')' = \bar{Y}_{t-1}$ and $\partial v_t(\gamma)/\partial\theta$ is a permutation of the elements of $\bar{Y}_{t-1}$. Therefore an application of Hölder's inequality and $\mathrm{E}(Y_t^8) < \infty$ provide **A8**. To prove **A9** note that the elements of $\partial\lambda_t(\psi)/\partial\psi = \partial v_t(\gamma)/\partial\gamma = (1, Y_{t-1})'$ are linearly independent and $v_t^*(\gamma^*) = v_t$ so by employing the results of Lemma 3 in Appendix A.2 the sufficient condition **A9*** holds. The same follows for the restricted estimators since $\partial v_t(\gamma)/\partial\gamma_2$ is a subvector of $(1, Y_{t-1})'$.

Finally, recall that the process $\{Y_t\}$ is β-mixing with coefficients $\beta(n) \le C\rho^n$ where C, ρ are positive constants and $\rho \in (0,1)$. Following Francq and Zakoian (2019, Sec. A.3) the score contribution $s_t(\theta_0)$ is also β-mixing with coefficients $\beta_s(n) \le \beta(n-1)$ for $n \ge 1$. By recalling that $\alpha_s(n) \le \beta_s(n)$ for $n \ge 1$ and $\alpha_s(0) \le 1/4$, we have that $\sum_{n=0}^{\infty}[\alpha(n)]^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$. Moreover, by a combination of Hölder's and c_p inequalities, $\mathrm{E}(Y_t)^8 < \infty$ is sufficient to show that $\mathrm{E}(\eta' s_t(\theta_0))^{2+\delta} < \infty$ for $\delta = 2$ and for all $\eta \in \mathbb{R}^m$ with $\eta \neq 0$. Therefore, an application of the Cramér–Wold device and the central limit theorem for α-mixing processes (Francq and Zakoian, 2019, Thm. A.4) shows that $\sqrt{T} S_T(\theta_0) \xrightarrow{d} N(0, I(\theta_0))$ as $T \to \infty$. This proves **A11**. □

Proof of Theorem 6. Since the observations are generated from a beta distribution, the results of Theorem 5 guarantee that **A1** and the restrictions of **A12** are satisfied with $v_t^*(\cdot) = v_t(\cdot)$. Define $\delta_i = \omega_i + \alpha_i + \beta_i$ for $i = 1, 2$. The restrictions on the parameter space imply that a.s. $0 < \omega_1 \le \lambda_t(\psi) \le \delta_1 < 1$, $0 < \omega_2 \le \mu_t(\gamma) \le \delta_2 < 1$ and $0 < \underline{v} \le v_t(\gamma) < \bar{v} < 1$ for any $\theta \in \Theta$ and any $t \ge 1$ where $\bar{v} = 1/(1+\phi) < 1$ since $\mu_t(1-\mu_t) < 1$ and $\underline{v} = \min\{\underline{v}_1, \underline{v}_2\}$ where $\underline{v}_1 = \omega_2(1-\omega_2)/(1+\phi)$ and $\underline{v}_2 = \delta_2(1-\delta_2)/(1+\phi)$. All these processes are a.s. bounded in the $(0,1)$ interval for any $\theta \in \Theta$ therefore all their sup-moments are bounded. Hence, **A2** and **A4** hold.

A3 is proved by contradiction. Assume that a.s. $\lambda_{t-1}(\psi) = \lambda_{t-1}(\psi_0) = \lambda_{t-1}$. Then $\lambda_t(\psi) - \lambda_t(\psi_0) = \omega_1 - \omega_{1,0} + (\alpha_1 - \alpha_{1,0})Y_{t-1} + (\beta_1 - \beta_{1,0})\lambda_{t-1}$. If $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. with $\omega_1 \neq \omega_{1,0}$ then $0 \neq \omega_{1,0} - \omega_1 = (\alpha_1 - \alpha_{1,0})Y_{t-1} + (\beta_1 - \beta_{1,0})\lambda_{t-1}$ a.s. and the equality will be possible only if $(\alpha_1 - \alpha_{1,0}) \neq 0$ and Y_{t-1} equals a non-zero constant a.s. and/or $(\beta_1 - \beta_{1,0}) \neq 0$ and λ_{t-1} equals a non-zero constant. However, Y_{t-1} is non-constant and since $\alpha_1 > 0$ this is true also for λ_{t-1}. Therefore, if $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. then $\omega_1 = \omega_{1,0}$ and $0 = (\alpha_1 - \alpha_{1,0})Y_{t-1} + (\beta_1 - \beta_{1,0})\lambda_{t-1}$. Now to have $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. with $\alpha_1 \neq \alpha_{1,0}$ and $\beta_1 \neq \beta_{1,0}$ we shall have that a.s. $Y_{t-1} = \lambda_{t-1} = 0$ but this is impossible since Y_{t-1} and λ_{t-1} are non-constant. Therefore, if $\lambda_t(\psi) = \lambda_t(\psi_0)$ a.s. then $\alpha_1 = \alpha_{1,0}$ and $\beta_1 = \beta_{1,0}$. An analogous result holds for $\mu_t(\gamma)$, consequently $v_t(\gamma) = v_t(\gamma_0)$ a.s. if and only if $\gamma = \gamma_0$. Then, **A6** holds following the same arguments provided in the proof of Theorem 4.

Recall that $\tilde{Y}_{-i} \in [0,1]$ for $i = 0, 1, \dots$ so a.s. $|\lambda_t(\psi) - \tilde{\lambda}_t(\psi)| = \beta_1^t |\lambda_0(\psi) - \tilde{\lambda}_0(\psi)| \le 2\beta_1^t$, $|\mu_t(\gamma) - \tilde{\mu}_t(\gamma)| \le 2\beta_2^t$. The variance is a function of μ_t so in simplified notation $\partial v_t(\gamma,\mu)/\partial\mu = (1-2\mu)/(1+\phi)$ and $|1-2\mu| \le c < 1$ since $0 < \mu < 1$, therefore by the mean value theorem a.s. $|v_t(\gamma) - \tilde{v}_t(\gamma)| \le 2c\beta_2^t$. This implies that, as $t \to \infty$, $a_t, b_t \to 0$ e.a.s. where *e.a.s.* means *exponentially fast a.s. convergence* (Straumann and Mikosch, 2006, Sec. 2.1). Then, the limits in **A5** converge e.a.s to 0.

Recall that 0_k is a $k \times 1$ vector of zeros. Define $Z_t(\theta) = (1, Y_t, \lambda_t(\psi), 0_4')'$, $C_1 = \|(1,1,1,0_4')'\|$ and $\sup_{\psi\in\Psi}\beta_1 = \rho_1$. Note that

$$\frac{\partial\lambda_t(\psi)}{\partial\theta} = Z_{t-1}(\theta) + \beta_1\frac{\partial\lambda_{t-1}(\psi)}{\partial\theta}, \qquad \frac{\partial v_t(\gamma)}{\partial\theta} = \frac{1-2\mu_t(\gamma)}{1+\phi}\frac{\partial\mu_t(\gamma)}{\partial\theta} + \frac{\mu_t^2(\gamma) - \mu_t(\gamma)}{(1+\phi)^2}1 = A(\theta) + B(\theta),$$

and $c_t \to 0$ e.a.s. by Gorgi and Koopman (2023, Lem. A.2). Then, for t large enough, with probability 1

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\psi)}{\partial\theta}\right\| \le \sum_{i=0}^{t-1}\rho_1^i \sup_{\theta\in\Theta}\left\|Z_{t-1-i}(\theta)\right\| + \rho_1^t \sup_{\theta\in\Theta}\frac{\partial\lambda_0(\psi)}{\partial\theta} \le C_1\sum_{i=0}^{\infty}\rho_1^i + 1 = M < \infty,$$

since $\rho_1 < 1$. By similar arguments $\sup_{\theta\in\Theta}\|\partial\mu_t(\gamma)/\partial\theta\| \le K$ and $\sup_{\theta\in\Theta}\|\partial v_t(\gamma)/\partial\theta\| \le 3K+2$, for t large enough, where K is a positive constant. By employing again the mean value theorem it follows that, for t large enough and with probability 1

$$d_t \le \sup_{\theta\in\Theta}\sup_{\mu\in(0,1)}\left\|\frac{\partial}{\partial\theta}\left(\frac{\partial v_t(\gamma,\mu)}{\partial\mu}\right)\right\| \sup_{\gamma\in\Gamma}|\mu_t(\gamma) - \tilde{\mu}_t(\gamma)| \le (3+2K)\sup_{\gamma\in\Gamma}|\mu_t(\gamma) - \tilde{\mu}_t(\gamma)|$$

converging to 0 e.a.s. as $t \to \infty$. Then, the limits in **A7** converge e.a.s to 0 and are of order $\mathcal{O}(t^{-\delta})$.

Recall that $O_{m,n}$ is a $m \times n$ matrix of zeros. The second derivative has the form

$$\frac{\partial^2\lambda_t(\psi)}{\partial\theta\partial\theta'} = \dot{Z}_{t-1}(\theta) + \beta_1\frac{\partial^2\lambda_{t-1}(\psi)}{\partial\theta\partial\theta'}, \qquad \dot{Z}_{t-1}(\theta) = \begin{pmatrix} O_{2,3} & 0_2 \\ \frac{\partial\lambda_{t-1}(\psi)}{\partial\theta'} & 0 \\ 0_3' & 0 \end{pmatrix}.$$

Following the same arguments of the first derivative, for t large enough and probability 1

$$\sup_{\theta\in\Theta}\left\|\frac{\partial^2\lambda_t(\psi)}{\partial\theta\partial\theta'}\right\| \le \sum_{i=0}^{\infty}\rho_1^i \sup_{\theta\in\Theta}\left\|\dot{Z}_{t-1-i}(\theta)\right\| + 1 \le \frac{C_2 M}{1-\rho_1} + 1 < \infty,$$

where C_2 is a positive constant depending on the type of matrix norm $\|\cdot\|$ employed. Analogously, $\sup_{\theta\in\Theta}\left\|\partial^2\mu_t(\gamma)/\partial\theta\partial\theta'\right\|$ and $\sup_{\theta\in\Theta}\left\|\partial^2 v_t(\gamma)/\partial\theta\partial\theta'\right\|$ are a.s. bounded by a constant so **A8** is verified.

Consider a deterministic vector $\eta \in \mathbb{R}^m$ with $\eta = (\eta_1', \eta_2)'$ where η_1 is of dimension 3 and η_2 is a scalar. By appealing the results of Lemma 3 in Appendix A.2, we prove **A9** by showing that $\eta_1'\partial\lambda_t(\psi_0)/\partial\psi = 0$ a.s. if and only if (henceforth, iff) $\eta_1 = 0$. The proof is by contradiction. Assume that $\eta_1'\partial\lambda_t(\psi_0)/\partial\psi = 0$ a.s. for some $\eta_1 \neq 0$. Then $\eta_1'\partial\lambda_{t-1}(\psi_0)/\partial\psi = 0$ a.s. by stationarity. Therefore from the formula of the first derivative we should have $\eta_1' Z_{t-1}(\psi_0) = \eta_1'(1, Y_{t-1}, \lambda_{t-1}(\psi_0))' = 0$ a.s. for some $\eta_1 \neq 0$. However, this is impossible since $Z_{t-1}(\psi_0)$ has linearly independent elements so it follows that $\eta_1' Z_{t-1}(\psi_0) = 0$ a.s. iff $\eta_1 = 0$. Recall that $\gamma = (\gamma_1, \phi)'$ where $\gamma_1 = (\omega_2, \alpha_2, \beta_2)'$. Note that

$$\eta'\frac{\partial v_t(\gamma_0)}{\partial\gamma} = \eta_1'\frac{\partial v_t(\gamma_0)}{\partial\gamma_1} + \eta_2\frac{\partial v_t(\gamma_0)}{\partial\phi} = \eta_1'\frac{1-2\mu_t(\gamma_0)}{1+\phi}\frac{\partial\mu_t(\gamma_0)}{\partial\gamma_1} - \eta_2\frac{\mu_t(\gamma_0)(1-\mu_t(\gamma_0))}{(1+\phi)^2} = \eta_1' m_t n_t - \eta_2 o_t$$

with obvious notation. We appeal again the proof by contradiction so assume that $\eta'\partial v_t(\gamma_0)/\partial\gamma = 0$ a.s. for some $\eta \neq 0$. We consider three cases. (i) $\eta_1 \neq 0, \eta_2 = 0$. We have that $m_t \neq 0$ a.s. since $\mu_t(\gamma_0)$ is non-degenerate, therefore it should be that $\eta_1' n_t = 0$ a.s. for some $\eta_1 \neq 0$, however n_t has linearly independent elements, following the same arguments of $\eta_1'\partial\lambda_t(\psi_0)/\partial\psi$ above, so the assumed statement cannot be true. (ii) $\eta_1 = 0, \eta_2 \neq 0$. In this case we have $o_t > 0$ a.s., by definition, so $\eta_2 o_t = 0$ a.s. cannot occur since $\eta_2 \neq 0$. (iii) $\eta_1 \neq 0, \eta_2 \neq 0$. In this case we shall have a.s. $\eta_1' n_t = \eta_2 m_t^{-1} o_t$ and therefore $\beta_2 \eta_1' n_{t-1} = \eta_2 m_t^{-1} o_t - \eta_1' Z_{t-1}(\gamma_0)$ where $Z_{t-1}(\gamma_0) = (1, Y_{t-1}, \mu_{t-1}(\gamma_0))'$. However this cannot hold because the left-hand side is $\mathcal{F}_{t-2}$-measurable whereas the right-hand side is not since it depends on Y_{t-1}. Then $\eta' \partial v_t(\gamma_0)/\partial\gamma = 0$ a.s. iff $\eta = 0$. Therefore **A9*** holds and **A9** follows. Noting that $\gamma_2 = \phi$, condition **A9*** holds by the arguments in (ii) so **A9** holds also for the restricted estimator.

Finally, **A11** holds as in the proof of Corollary 2 because the score contribution $s_t(\theta_0)$ is a martingale difference sequence and therefore $\sqrt{T} S_T(\theta_0) \xrightarrow{d} N(0, I(\theta_0))$ as $T \to \infty$. □

Proof of Proposition 1. Under the conditions of Proposition 1, $p = k = 1$ so ψ and γ are scalar. In particular, $\gamma = \gamma_1 \in \mathbb{R}$, i.e. there are no free nuisance parameters γ_2. So, under the results of Theorem 2, following the notation for restricted estimators defined below assumption **A12**, it is not hard to show that the limiting covariance of the restricted estimator is a scalar and takes the form $\Sigma_R = H_\psi^{-1}(\theta_0) I_\psi(\theta_0) H_\psi^{-1}(\theta_0)$ with

$$H_\psi(\theta_0) = \mathrm{E}\left[\frac{1}{v_t(\gamma_0)} \frac{\partial\lambda_t(\psi_0)}{\partial\psi}^2 + \frac{1}{2v_t^2(\gamma_0)} \frac{\partial v_t(\gamma_0)}{\partial\psi}^2\right], \tag{A.6}$$

$$I_\psi(\theta_0) = \mathrm{E}\left[\frac{1}{v_t(\gamma_0)} \frac{\partial\lambda_t(\psi_0)}{\partial\psi}^2 + \frac{h_t}{2v_t^3(\gamma_0)}\left(\frac{\partial\lambda_t(\psi_0)}{\partial\psi}\frac{\partial v_t(\gamma_0)}{\partial\psi} + \frac{\partial v_t(\gamma_0)}{\partial\psi}\frac{\partial\lambda_t(\psi_0)}{\partial\psi}\right)\right] + \mathrm{E}\left[\left(\frac{k_t}{v_t^2(\gamma_0)} - 1\right)\frac{1}{4v_t^2(\gamma_0)}\frac{\partial v_t(\gamma_0)}{\partial\psi}^2\right], \tag{A.7}$$

where $h_t = \mathrm{E}[(Y_t - \lambda_t(\psi_0))^3 | \mathcal{F}_{t-1}]$ and $k_t = \mathrm{E}[(Y_t - \lambda_t(\psi_0))^4 | \mathcal{F}_{t-1}]$ by **A13**. By Corollary 2, the limiting covariance of the unrestricted estimator, I_ψ^{-1}, is the reciprocal expected value of the first summand of (A.6). In the case **A14.a** we have that $h_t = 0$ and $k_t \leq 3v_t(\gamma_0)$, with equality if and only if $q(\cdot)$ is Gaussian. Hence, from (A.7) $I_\psi(\theta_0) \leq H_\psi(\theta_0)$ and $\Sigma_R \leq H_\psi^{-1}(\theta_0) \leq I_\psi^{-1}$ where the last inequality holds since the second summand in (A.6) is non-negative. In the case **A14.b** we have that $h_t > 0$ and $\partial\lambda_t(\psi_0)/\partial\psi \; \partial v_t(\gamma_0)/\partial\psi < 0$ or $h_t < 0$ and $\partial\lambda_t(\psi_0)/\partial\psi \; \partial v_t(\gamma_0)/\partial\psi > 0$. In both scenarios the second summand in (A.7) is negative so $I_\psi(\theta_0) < H_\psi(\theta_0)$. The result follows as above. □

A.2. Technical lemmas

Lemma 1. *Consider the PVQMLE in (5) with log-quasi-likelihood (4). Under conditions* **A4**–***A5****, almost surely as* $T \to \infty$, $\sup_{\theta\in\Theta} |\tilde{L}_T(\theta) - L_T(\theta)| \to 0$.

Proof of Lemma 1. From assumption **A4**, we have that

$$\begin{aligned}
\sup_{\theta\in\Theta} |l_t(\theta) - \tilde{l}_t(\theta)| &\leq \sup_{\theta\in\Theta}\left|\frac{[\tilde{\lambda}_t(\psi) - \lambda_t(\psi)][\tilde{\lambda}_t(\psi) + \lambda_t(\psi) - 2Y_t]}{2\tilde{v}_t^*(\gamma)} + \frac{[v_t^*(\gamma) - \tilde{v}_t^*(\gamma)][Y_t - \lambda_t(\psi)]^2}{2v_t^*(\gamma)\tilde{v}_t^*(\gamma)}\right| + \frac{1}{2}\sup_{\gamma\in\Gamma}\left|\log\frac{\tilde{v}_t^*(\gamma)}{v_t^*(\gamma)}\right| \\
&\leq \frac{1}{\underline{v}^*} a_t\left(a_t + |Y_t| + \sup_{\psi\in\Psi}|\lambda_t(\psi)|\right) + \frac{1}{\underline{v}^{*2}} b_t\left(Y_t^2 + \sup_{\psi\in\Psi}\lambda_t^2(\psi)\right) + \frac{1}{2}\sup_{\gamma\in\Gamma}\left|\log\left(1 + \frac{\tilde{v}_t^*(\gamma) - v_t^*(\gamma)}{v_t^*(\gamma)}\right)\right| \\
&\leq \frac{1}{\underline{v}^*} a_t\left(1 + |Y_t| + \sup_{\psi\in\Psi}|\lambda_t(\psi)|\right) + \frac{1}{\underline{v}^{*2}} b_t\left(Y_t^2 + \sup_{\psi\in\Psi}\lambda_t^2(\psi)\right) + \frac{1}{2\underline{v}^*} b_t,
\end{aligned}$$

for t large enough since, by assumption **A5**, a.s. $a_t \to 0$ as $t \to \infty$. Note that in the last inequality we have used the fact that $x/(x+1) \leq \log(1+x) \leq x$ for $x > -1$ and that $|\log(1+x)| \leq \max\{|x/(x+1)|, |x|\}$. Indeed, by setting the simplified notation $x = (\tilde{v} - v)/v$, it is clear that $x = \tilde{v}/v - 1 > -1$ since $\tilde{v}/v > 0$. By standard algebra we find that $|x/(x+1)| = |\tilde{v} - v|/\tilde{v}$. Therefore $|\log(1+x)| \leq \max\{|\tilde{v} - v|/\tilde{v}, |\tilde{v} - v|/v\} \leq b_t/\underline{v}^*$ where the last inequality follows by **A4** and the definition of b_t. Assumption **A5** and an application of Cesaro's lemma lead to

$$\sup_{\theta\in\Theta} |\tilde{L}_T(\theta) - L_T(\theta)| \leq T^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\tilde{l}_t(\theta) - l_t(\theta)| \to 0, \quad a.s.$$

as $T \to \infty$. □

Lemma 2. *Consider the PVQMLE in (5) with score (13). Under conditions* **A4**–***A5*** *and* ***A7****, almost surely as* $T \to \infty$, $\sqrt{T}\sup_{\theta\in\Theta}\|\tilde{S}_T(\theta) - S_T(\theta)\| \to 0$.

Proof of Lemma 2. We obtain that

$$\begin{aligned}
\sup_{\theta\in\Theta}\|s_t(\theta) - \tilde{s}_t(\theta)\| &\leq \sup_{\theta\in\Theta}\left\|\frac{1}{2\tilde{v}_t^*(\gamma)}\frac{\partial\tilde{v}_t^*(\gamma)}{\partial\theta} - \frac{1}{2v_t^*(\gamma)}\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\| + \sup_{\theta\in\Theta}\left\|\frac{Y_t - \tilde{\lambda}_t(\psi)}{\tilde{v}_t^*(\gamma)}\frac{\partial\tilde{\lambda}_t(\psi)}{\partial\theta} - \frac{Y_t - \lambda_t(\psi)}{v_t^*(\gamma)}\frac{\partial\lambda_t(\psi)}{\partial\theta}\right\| \\
&\quad + \sup_{\theta\in\Theta}\left\|\frac{[Y_t - \tilde{\lambda}_t(\psi)]^2}{2\tilde{v}_t^{*2}(\gamma)}\frac{\partial\tilde{v}_t^*(\gamma)}{\partial\theta} - \frac{[Y_t - \lambda_t(\psi)]^2}{2v_t^{*2}(\gamma)}\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\| \\
&= \delta_t^1 + \delta_t^2 + \delta_t^3,
\end{aligned}$$

with obvious notation. We now bound the single terms individually. In what follows the notation $o(1)$ almost surely, as $t \to \infty$, will be abbreviated to $o(1)$.

$$\delta_t^1 \leq \sup_{\theta\in\Theta}\left\| \frac{1}{2\tilde{v}_t^*(\gamma)}\left(\frac{\partial \tilde{v}_t^*(\gamma)}{\partial\theta} - \frac{\partial v_t^*(\gamma)}{\partial\theta}\right) + \frac{\left[v_t^*(\gamma) - \tilde{v}_t^*(\gamma)\right]}{2\tilde{v}_t^*(\gamma)v_t^*(\gamma)}\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\| \leq \frac{d_t}{2\underline{v}^*} + \frac{b_t}{2\underline{v}^{*2}}\sup_{\theta\in\Theta}\left\|\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\| .$$

Similarly,

$$\begin{aligned}
\delta_t^2 &\leq \sup_{\theta\in\Theta}\left\| \frac{Y_t - \tilde{\lambda}_t(\psi)}{\tilde{v}_t^*(\gamma)}\left(\frac{\partial\tilde{\lambda}_t(\psi)}{\partial\theta} - \frac{\partial\lambda_t(\psi)}{\partial\theta}\right)\right\| + \sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\psi)}{\partial\theta}\left(\frac{\lambda_t(\psi) - \tilde{\lambda}_t(\psi)}{\tilde{v}_t^*(\gamma)} + \frac{Y_t - \lambda_t(\psi)}{\tilde{v}_t^*(\gamma)} - \frac{Y_t - \lambda_t(\psi)}{v_t^*(\gamma)}\right)\right\| \\
&\leq \frac{c_t}{\underline{v}^*}\left(|Y_t| + \sup_{\psi\in\Psi}|\lambda_t(\psi)| + a_t\right) + \sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\psi)}{\partial\theta}\right\|\left(\frac{a_t}{\underline{v}^*} + \sup_{\theta\in\Theta}\left|\frac{\left[v_t^*(\gamma) - \tilde{v}_t^*(\gamma)\right]\left[Y_t - \lambda_t(\psi)\right]}{\tilde{v}_t^*(\gamma)v_t^*(\gamma)}\right|\right) \\
&\leq \frac{c_t}{\underline{v}^*}\left(|Y_t| + \sup_{\psi\in\Psi}|\lambda_t(\psi)| + o(1)\right) + \sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\psi)}{\partial\theta}\right\|\left(\frac{a_t}{\underline{v}^*} + \frac{b_t}{\underline{v}^{*2}}\left(|Y_t| + \sup_{\psi\in\Psi}|\lambda_t(\psi)|\right)\right).
\end{aligned}$$

Using similar arguments for δ_t^3 and assumption A5 leads to

$$\begin{aligned}
\delta_t^3 &\leq \frac{d_t}{\underline{v}^{*2}}\sup_{\theta\in\Theta}\left(Y_t^2 + \lambda_t^2(\psi) + a_t^2 + 2a_t\lambda_t(\psi)\right) + \sup_{\theta\in\Theta}\left\|\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\|\sup_{\theta\in\Theta}\left|\frac{[\tilde{\lambda}_t(\psi) - \lambda_t(\psi)][\tilde{\lambda}_t(\psi) + \lambda_t(\psi) - 2Y_t]}{2\tilde{v}_t^{*2}(\gamma)}\right| \\
&\quad + \sup_{\theta\in\Theta}\left\|\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\|\sup_{\theta\in\Theta}\left|\frac{\left[v_t^*(\gamma) - \tilde{v}_t^*(\gamma)\right]\left[v_t^*(\gamma) + \tilde{v}_t^*(\gamma)\right]\left[Y_t - \lambda_t(\psi)\right]^2}{2v_t^{*2}(\gamma)\tilde{v}_t^{*2}(\gamma)}\right| \\
&\leq \frac{d_t}{\underline{v}^{*2}}\left(Y_t^2 + \sup_{\psi\in\Psi}\lambda_t^2(\psi) + o(1)\right) + \sup_{\theta\in\Theta}\left\|\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\|\frac{a_t}{\underline{v}^{*2}}\left(|Y_t| + \sup_{\psi\in\Psi}|\lambda_t(\psi)| + o(1)\right) + \sup_{\theta\in\Theta}\left\|\frac{\partial v_t^*(\gamma)}{\partial\theta}\right\|\frac{2b_t}{\underline{v}^{*3}}\left(Y_t^2 + \sup_{\psi\in\Psi}\lambda_t^2(\psi)\right).
\end{aligned}$$

By assumption A7, $\delta_t^j = \mathcal{O}(t^{-\delta})$, for $\delta > 1/2$ and $j = 1,2,3$. Therefore $\sqrt{T}\sup_{\theta\in\Theta}\|S_T(\theta) - \tilde{S}_T(\theta)\| \leq T^{-1/2}\sum_{t=1}^T \mathcal{O}(t^{-\delta})$ converges to 0 almost surely as $T \to \infty$. □

Lemma 3. *Assumption* ***A9*** *is satisfied for the unrestricted PVQMLE (5) under the following sufficient condition.*

A9* *The random variables of the vectors $\partial\lambda_t(\psi_0)/\partial\psi$ and $\partial v_t^*(\gamma^*)/\partial\gamma$ are linearly independent. Moreover, one of the following conditions holds a.s. for some $t \geq 1$.*

1. $v_t^*(\gamma^*) = v_t$.
2. $v_t^*(\gamma^*) < v_t$, $\partial^2 v_t^*(\gamma^*)/\partial\theta\partial\theta'$ *is negative semi-definite.*
3. $v_t < v_t^*(\gamma^*) < 2v_t$, $\partial^2 v_t^*(\gamma^*)/\partial\theta\partial\theta'$ *is positive semi-definite.*
4. $v_t^*(\gamma^*) < 2v_t$, $\partial^2 v_t^*(\gamma^*)/\partial\theta_i\partial\theta_j = 0$ *for all* $i,j = 1,\ldots,m$.

The same result holds for the restricted PVQMLE (6) with $\partial\gamma_2$ instead of $\partial\gamma$.

Proof of Lemma 3. Condition A9 requires that for all $\eta \in \mathbb{R}^m$, $\eta'\,\mathrm{E}[-\partial^2 l_t(\theta_0)/\partial\theta\partial\theta']\eta > 0$, with $\eta \neq 0$, but $\mathrm{E}[-\partial^2 l_t(\theta_0)/\partial\theta\partial\theta'] = \mathrm{E}[\mathrm{E}[-\partial^2 l_t(\theta_0)/\partial\theta\partial\theta'|\mathcal{F}_{t-1}]]$ and following (A.2), we only need to show

$$\mathrm{E}\left(d_t\eta' f_t^\theta f_t^{\theta\prime}\eta + l_t\eta' h_t^\theta h_t^{\theta\prime}\eta + \eta' C_t^\theta\eta\right) > 0\,, \tag{A.8}$$

where

$$d_t = \frac{1}{v_t^*(\gamma^*)}\,,\quad f_t^\theta = \frac{\partial\lambda_t(\psi_0)}{\partial\theta}\,,\quad l_t = \frac{2v_t - v_t^*(\gamma^*)}{2v_t^{*3}(\gamma^*)}\,,\quad h_t^\theta = \frac{\partial v_t^*(\gamma^*)}{\partial\theta}\,,\quad C_t^\theta = \left(\frac{v_t^*(\gamma^*) - v_t}{2v_t^{*2}(\gamma^*)}\right)\frac{\partial^2 v_t^*(\gamma^*)}{\partial\theta\partial\theta'}\,.$$

Note that under the conditions in A9* we have that a.s. $l_t > 0$ and $\eta' C_t^\theta\eta \geq 0$. Moreover, a.s. $d_t > 0$, $\eta' f_t^\theta f_t^{\theta\prime}\eta = (\eta' f_t^\theta)^2 \geq 0$ and $\eta' h_t^\theta h_t^{\theta\prime}\eta = (\eta' h_t^\theta)^2 \geq 0$. Therefore a sufficient condition for (A.8) requires a.s. $\eta' f_t^\theta \neq 0$ or $\eta' h_t^\theta \neq 0$. Let 0_m be a m-dimensional vector of zeros. To prove the result recall that

$$f_t^\theta = \begin{pmatrix} f_t^\psi \\ 0_k \end{pmatrix}\,,\qquad h_t^\theta = \begin{pmatrix} h_t^\psi \\ h_t^\gamma \end{pmatrix}\,.$$

Hence, a.s. $\eta' f_t^\theta = \eta_1' f_t^\psi$. We can split $\eta = (\eta_1', \eta_2')'$ where η_1 has dimension p and η_2 has dimension k. Consider two cases: (i) $\eta_1 \neq 0$ and (ii) $\eta_1 = 0$. Under (i) the result is verified by a.s. $\eta_1' f_t^\psi \neq 0$. In the case (ii), the result follows by $\eta_2' h_t^\gamma \neq 0$ a.s. since $\eta' h_t^\theta = \eta_2' h_t^\gamma$.

Recall that $\eta' I(\theta_0)\eta = \mathrm{E}[(\eta' s_t(\theta_0))^2] \geq 0$. Therefore, to prove the positive definiteness of $I(\theta_0)$ we need to show that for all $\eta \in \mathbb{R}^m$, with $\eta \neq 0$, $\eta' s_t(\theta_0) \neq 0$ where

$$\eta' s_t(\theta_0) = \frac{e_t}{v_t^*(\gamma^*)}\eta' f_t^\theta + \frac{e_t^2 - v_t^*(\gamma^*)}{2v_t^{*2}(\gamma^*)}\eta' h_t^\theta\,, \tag{A.9}$$

and $e_t = Y_t - \lambda_t(\psi_0)$ and therefore a.s. $e_t \neq 0$, $e_t^2 - v_t^*(\gamma^*) \neq 0$ and $v_t^*(\gamma^*) > 0$. If only one between $\eta' f_t^\theta$ and $\eta' h_t^\theta$ is different from 0 a.s. then the result follows with argument identical to the Hessian matrix case above. In the case where both $\eta' f_t^\theta$ and $\eta' h_t^\theta$ are not 0 a.s. we show that this cannot imply that $\eta' s_t(\theta_0) = 0$ since by Eq. (A.9) this would entail

$$\eta' f_t^{\theta} = -\frac{e_t^2 - v_t^*(\gamma^*)}{2v_t^*(\gamma^*)e_t}\eta' h_t^{\theta},$$

where the left-hand side is $\mathcal{F}_{t-1}$-measurable whereas the right-hand side is not as it depends on Y_t. Therefore $\eta' s_t(\theta_0) \neq 0$ for any non-trivial vector η. This concludes the proof. □

Appendix B. Further numerical results

B.1. Outliers

We evaluate the robustness of the described test statistic by repeating the same simulation study as in Section 5 with the inclusion of an outlier defined as 3 times the standard deviation of the observations plus their sample mean. The results are summarized in Tables B.6–B.7 and Fig. 6 below.

B.2. Near unit root

We evaluate the test in case of near unit root by considering the same simulation setting as in Section 5 but setting the autoregressive parameter equal to 0.99. The results are summarized in Tables B.8–B.9 and Fig. 7 below.

Table B.6
Empirical size for test in (27) with outlier. The model considered under H_0 is an INAR(1) model with Poisson thinning as well as Poisson error with parameter values $a = 0.75$ and $\omega = 1$.

Nominal size	T				
	100	250	500	1000	2000
0.1000	0.1510	0.1544	0.1312	0.1154	0.1112
0.0500	0.0692	0.0730	0.0646	0.0550	0.0540
0.0100	0.0142	0.0116	0.0106	0.0112	0.0128

Table B.7
Mean of parameters estimated for unrestricted PVQMLE over 5000 simulations with the presence of outlier. Data are generated from an INAR(1) model with Poisson thinning as well as Poisson error with parameter values $a = 0.75$ and $\omega = 1$.

T	a	ω_1	ω_2	b
100	0.6827	1.2679	1.7522	0.8230
250	0.7235	1.1059	1.3657	0.7740
500	0.7372	1.0497	1.1805	0.7616
1000	0.7432	1.0268	1.1066	0.7509
2000	0.7467	1.0125	1.0562	0.7497

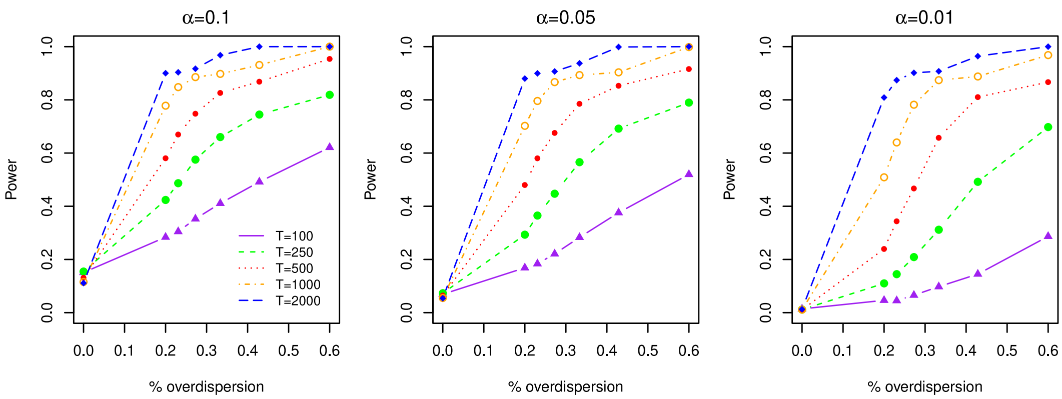

Fig. 6. Empirical power for test in (27) with outlier. The true parameter values of the INAR(1) model with negative binomial thinning and Poisson error are $a = 0.75$ and $\omega = 1$. The value of the dispersion parameter v changes as indicated in the horizontal axis through the % of overdispersion: $1 - a/(a + a^2/v)$.

Table B.8
Empirical size for test in (27). The model considered under H_0 is an INAR(1) model with Poisson thinning as well as Poisson error with parameter values $a = 0.99$ and $\omega = 1$.

Nominal size	T				
	100	250	500	1000	2000
0.1000	0.0524	0.0562	0.0568	0.0614	0.0660
0.0500	0.0250	0.0310	0.0300	0.0296	0.0254
0.0100	0.0074	0.0060	0.0076	0.0068	0.0050

Table B.9
Mean of parameters estimated for unrestricted PVQMLE over 5000 simulations with the presence of outlier. Data are generated from an INAR(1) model with Poisson thinning as well as Poisson error with parameter values $a = 0.99$ and $\omega = 1$.

T	a	ω_1	ω_2	b
100	0.9364	6.0910	16.8541	0.7962
250	0.9690	2.9188	6.8376	0.9150
500	0.9802	1.8379	3.3397	0.9578
1000	0.9857	1.3415	1.7776	0.9776
2000	0.9880	1.1619	1.1795	0.9866

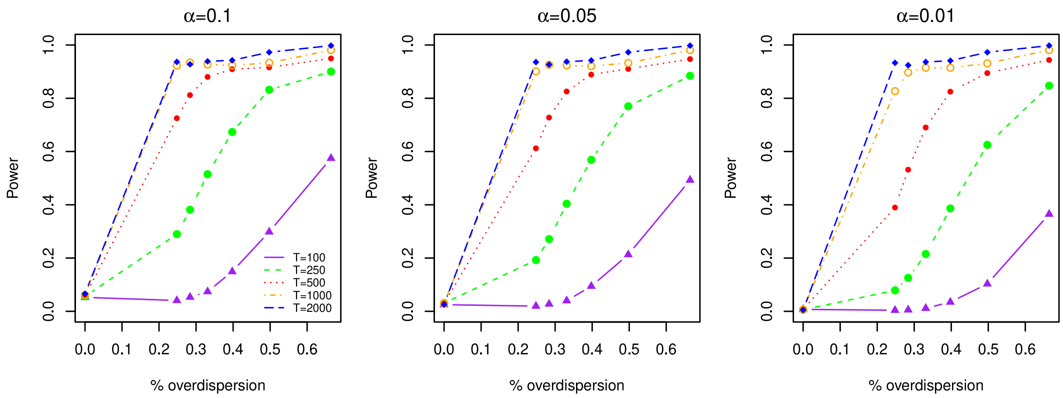

Fig. 7. Empirical power for test in (27). The true parameter values of the INAR(1) model with negative binomial thinning and Poisson error are $a = 0.99$ and $\omega = 1$. The value of the dispersion parameter υ changes as indicated in the horizontal axis through the % of overdispersion: $1 - a/(a + a^2/\upsilon)$.

B.3. BiNB thinning

We evaluate the power of the equidispersion test under BiNB thinning. In this specification $X_j \sim BerG(\mu, \pi)$ is the Bernoulli-Geometric distribution, defined as the sum of a Bernoulli distribution with probability π and a geometric distribution with mean μ where the distributions are independent and $\mu + \pi < 1$. Moreover, $a = \mu + \pi$, $b = \pi(1 - \pi) + \mu(1 - \mu)$ so $b/a = 1 + \mu - \pi$ and therefore $b > a$ if $\mu > \pi$. Fig. 8 below shows the power of the test in (27) to reject the null hypothesis.

B.4. INAR(2)

We consider the case of testing equidispersion of the thinning operator in an INAR(2) model. In this case the hypothesis test is the following

$$H_0 : b_i = a_i \quad \text{vs} \quad H_1 : b_i \neq a_i \quad \text{for } i = 1, 2\,. \tag{B.10}$$

The results of the test against negative binomial thinning are reported in Tables B.10–B.11 and Fig. 9.

Data availability

Codes and data to replicate the analyses in the paper are available at https://github.com/mirkoarmillotta/Pseudo_variance.

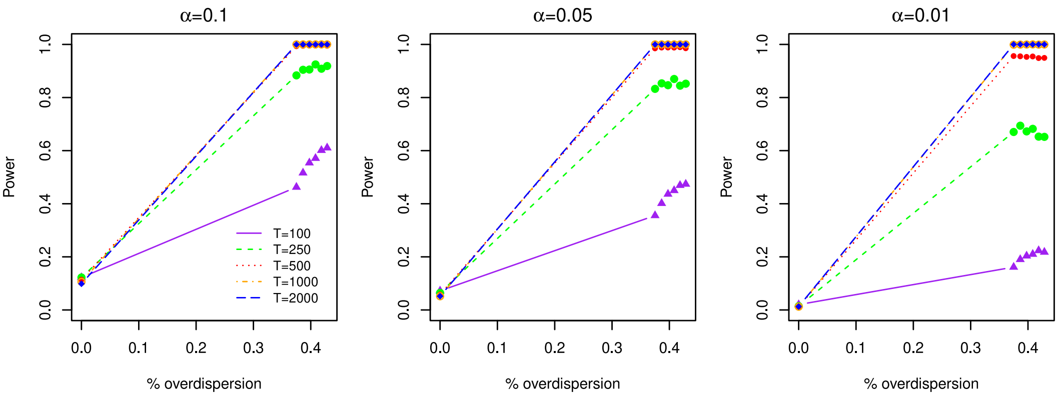

Fig. 8. Empirical power for test in (27). The true parameter values of the INAR(1) model with BiNB thinning and Poisson error are $\mu = 0.75$ and $\omega = 1$. The value of the Bernoulli parameter $\pi \in [0, 0.15]$ changes as indicated in the horizontal axis through the % of overdispersion: $1 - 1/(1 + \mu - \pi)$.

Table B.10
Empirical size for test in (B.10). The model considered under H_0 is an INAR(2) model with Poisson thinning as well as Poisson error with parameter values $a_1 = a_2 = 0.4$ and $\omega = 1$.

Nominal size	T				
	100	250	500	1000	2000
0.1000	0.1384	0.1460	0.1348	0.1182	0.1096
0.0500	0.0932	0.0822	0.0800	0.0632	0.0574
0.0100	0.0452	0.0242	0.0206	0.0166	0.0138

Table B.11
Mean of parameters estimated for unrestricted PVQMLE over 5000 simulations. Data are generated from an INAR(2) model with Poisson thinning as well as Poisson error with parameter values $a_1 = a_2 = 0.4$ and $\omega = 1$.

T	a_1	a_2	ω_1	ω_2	b_1	b_2
100	0.3835	0.3614	1.2498	1.0030	0.3939	0.3865
250	0.3937	0.3863	1.0858	0.9856	0.3978	0.3964
500	0.3966	0.3942	1.0386	0.9838	0.3997	0.3986
1000	0.3976	0.3977	1.0189	0.9977	0.3979	0.3991
2000	0.3991	0.3984	1.0099	0.9989	0.3999	0.3991

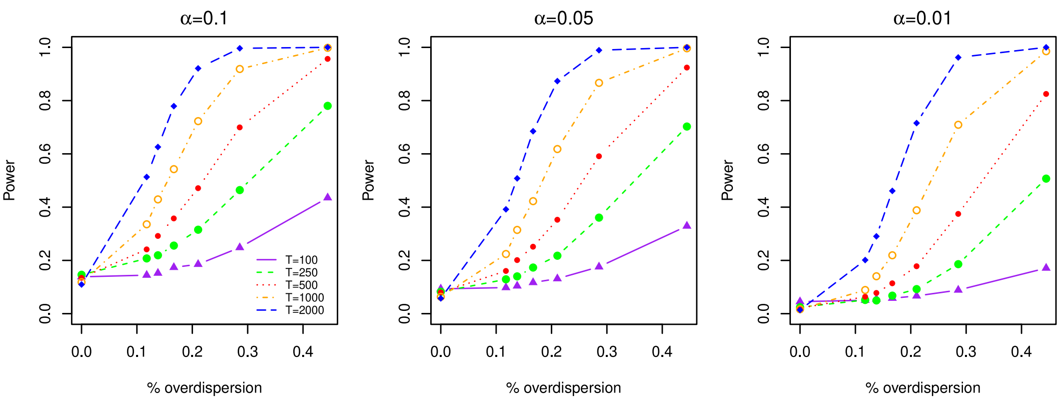

Fig. 9. Empirical power for test in (B.10). The true parameter values of the INAR(2) model with negative binomial thinning and Poisson error are $a_1 = a_2 = 0.4$ and $\omega = 1$. The value of the dispersion parameter v changes as indicated in the horizontal axis through the % of overdispersion: $1 - a_1/(a_1 + a_1^2/v)$. The horizontal axis is equal to $1 - a_2/(a_2 + a_2^2/v)$ since $a_1 = a_2$.

References

Ahmad, A., Francq, C., 2016. Poisson QMLE of count time series models. J. Time Series Anal. 37, 291–314.

Aknouche, A., Francq, C., 2023. Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models. J. Econometrics 237 (2), 105174.

Al-Osh, M., Alzaid, A.A., 1987. First-order integer-valued autoregressive (INAR (1)) process. J. Time Series Anal. 8, 261–275.

Aleksandrov, B., Weiß, C.H., 2020. Testing the dispersion structure of count time series using Pearson residuals. AStA Adv. Stat. Anal. 104 (3), 325–361.

Billingsley, P., 1961. The Lindeberg-Lévy theorem for martingales. Proc. Amer. Math. Soc. 12, 788–792.

Borges, P., Bourguignon, M., Molinares, F.F., 2017. A generalised NGINAR(1) process with inflated-parameter geometric counting series. Aust. N. Z. J. Stat. 59 (1), 137–150.

Borges, P., Molinares, F.F., Bourguignon, M., 2016. A geometric time series model with inflated-parameter Bernoulli counting series. Statist. Probab. Lett. 119, 264–272.

Bourguignon, M., Borges, P., Fajardo Molinares, F., 2018. A new geometric INAR(1) process based on counting series with deflation or inflation of zeros. J. Stat. Comput. Simul. 88 (17), 3338–3348.

Bourguignon, M., Weiß, C.H., 2017. An INAR(1) process for modeling count time series with equidispersion, underdispersion and overdispersion. Test 26 (4), 847–868.

Box, G.E., Jenkins, G.M., Reinsel, G., 1970. Time Series Analysis: Forecasting and Control. Holden-day San Francisco.

Bradley, R.C., 2005. Basic properties of strong mixing conditions. a survey and some open questions. Probab. Surv. 2, 107–144.

Christou, V., Fokianos, K., 2014. Quasi-likelihood inference for negative binomial time series models. J. Time Series Anal. 35, 55–78.

Christou, V., Fokianos, K., 2015. Estimation and testing linearity for non-linear mixed Poisson autoregressions. Electron. J. Stat. 9, 1357–1377. http://dx.doi.org/10.1214/15-EJS1044.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. J. Appl. Econometrics 28 (5), 777–795. http://dx.doi.org/10.1002/jae.1279.

Daniels, H.E., 1954. Saddlepoint approximations in statistics. Ann. Math. Stat. 25, 631–650.

Debaly, Z.M., Truquet, L., 2021. A note on the stability of multivariate non-linear time series with an application to time series of counts. Statist. Probab. Lett. 179, 1–7.

Doukhan, P., 1994. Mixing. In: Lecture Notes in Statistics, vol. 85, Springer-Verlag, New York, http://dx.doi.org/10.1007/978-1-4612-2642-0.

Doukhan, P., Fokianos, K., Li, X., 2012. On weak dependence conditions: The case of discrete valued processes. Statist. Probab. Lett. 82, 1941–1948.

Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Ferland, R., Latour, A., Oraichi, D., 2006. Integer-valued GARCH process. J. Time Series Anal. 27, 923–942.

Francq, C., Zakoian, J.-M., 2019. GARCH Models: Structure, Statistical Inference and Financial Applications. John Wiley & Sons.

Francq, C., Zakoian, J.-M., 2023. Optimal estimating function for weak location-scale dynamic models. J. Time Series Anal. 44, 533–555.

Goodman, J., 2022. Asymptotic accuracy of the saddlepoint approximation for maximum likelihood estimation. Ann. Statist. 50 (4), 2021–2046.

Gorgi, P., 2018. Integer-valued autoregressive models with survival probability driven by a stochastic recurrence equation. J. Time Series Anal. 39 (2), 150–171.

Gorgi, P., Koopman, S.J., 2023. Beta observation-driven models with exogenous regressors: A joint analysis of realized correlation and leverage effects. J. Econometrics 237 (2), 105177.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: Theory. Econometrica 52 (3), 681–700.

Guerrero, M.B., Barreto-Souza, W., Ombao, H., 2022. Integer-valued autoregressive processes with prespecified marginal and innovation distributions: a novel perspective. Stoch. Models 38 (1), 70–90.

Heinen, A., 2003. Modelling time series count data: an autoregressive conditional Poisson model. Technical Report MPRA Paper 8113, University Library of Munich, Germany, availabel at http://mpra.ub.uni-muenchen.de/8113/.

Latour, A., 1997. The multivariate GINAR (p) process. Adv. in Appl. Probab. 29, 228–248.

Latour, A., 1998. Existence and stochastic structure of a non-negative integer-valued autoregressive process. J. Time Series Anal. 19 (4), 439–455.

Leisen, F., Mena, R.H., Palma, F., Rossini, L., 2019. On a flexible construction of a negative binomial model. Statist. Probab. Lett. 152, 1–8.

Lu, Y., 2021. The predictive distributions of thinning-based count processes. Scand. J. Stat. 48, 42–67.

McKenzie, E., 1988. Some ARMA models for dependent sequences of Poisson counts. Adv. in Appl. Probab. 20, 822–835.

Miletić Ilić, A.V., 2016. A geometric time series model with a new dependent Bernoulli counting series. Comm. Statist. Theory Methods 45 (21), 6400–6415.

Nastić, A.S., Ristić, M.M., Ilić, A.V.M., 2017. A geometric time-series model with an alternative dependent Bernoulli counting series. Comm. Statist. Theory Methods 46 (2), 770–785.

Nocedal, J., Wright, S.J., 1999. Numerical Optimization. Springer.

Pedeli, X., Davison, A.C., Fokianos, K., 2015. Likelihood estimation for the INAR(p) model by saddlepoint approximation. J. Amer. Statist. Assoc. 110 (511), 1229–1238.

Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models. Springer-Verlag, Berlin, Asymptotic theory.

Ristić, M.M., Nastić, A.S., Miletić Ilić, A.V., 2013. A geometric time series model with dependent Bernoulli counting series. J. Time Series Anal. 34 (4), 466–476.

Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. Test 18 (3), 529.

Rudin, W., 1976. Principles of Mathematical Analysis. vol. 3, McGraw-hill New York.

Schweer, S., Weiß, C.H., 2014. Compound Poisson INAR (1) processes: Stochastic properties and testing for overdispersion. Comput. Statist. Data Anal. 77, 267–284.

Steutel, F., Van Harn, K., 1979. Discrete analogues of self-decomposability and stability. Ann. Probab. 7 (5), 893–899.

Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. Ann. Statist. 34 (5), 2449–2495.

Sun, J., McCabe, B.P., 2013. Score statistics for testing serial dependence in count data. J. Time Series Anal. 34 (3), 315–329.

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

Weiß, C.H., 2018a. An Introduction to Discrete-valued Time Series. John Wiley & Sons.

Weiß, C.H., 2018b. Goodness-of-fit testing of a count time series' marginal distribution. Metrika 81 (6), 619–651.

White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press.