JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **41**: 620–651 (2020)
Published online 06 May 2020 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12526

ORIGINAL ARTICLE

ROBUST ESTIMATION OF STATIONARY CONTINUOUS-TIME ARMA MODELS VIA INDIRECT INFERENCE

VICKY FASEN-HARTMANN[a*] AND SEBASTIAN KIMMIG[b]

[a] *Institute of Stochastics, Karlsruhe, Germany*
[b] *Württembergische Versicherung AG, Stuttgart, Germany*

In this article, we present a robust estimator for the parameters of a stationary, but not necessarily Gaussian, continuous-time ARMA(p, q) (CARMA(p, q)) process that is sampled equidistantly. Therefore, we propose an indirect estimation procedure that first estimates the parameters of the auxiliary AR(r) representation ($r \geq 2p - 1$) of the sampled CARMA process using a generalized M- (GM-)estimator. Since the map which maps the parameters of the auxiliary AR(r) representation to the parameters of the CARMA process is not given explicitly, a separate simulation part is necessary where the parameters of the AR(r) representation are estimated from simulated CARMA processes. Then, the parameters which take the minimum distance between the estimated AR parameters and the simulated AR parameters give an estimator for the CARMA parameters. First, we show that under some standard assumptions the GM-estimator for the AR(r) parameters is consistent and asymptotically normally distributed. Then, we prove that the indirect estimator is also consistent and asymptotically normally distributed when the asymptotically normally distributed LS-estimator is used in the simulation part. The indirect estimator satisfies several important robustness properties such as weak resistance, π_{d_n}-robustness and it has a bounded influence functional. The practical applicability of our method is illustrated in a small simulation study with replacement outliers.

Received 31 March 2018; Accepted 20 March 2020

Keywords: AR process; CARMA process; indirect estimator; influence functional; GM-estimator; LS-estimator; outlier; resistance; robustness.

MOS subject classification: 62F10; 62F12; 62M10; 60G10; 60G51.

1. INTRODUCTION

The article presents a robust estimator for the parameters of a discretely observed stationary continuous-time ARMA (CARMA) process. A *weak ARMA*(p, q) *process* in discrete-time is a weakly stationary solution of the stochastic difference equation

$$\phi(B)X_m = \theta(B)Z_m, \quad m \in \mathbb{Z}, \tag{1}$$

where B denotes the backward shift operator (i.e., $BX_m = X_{m-1}$),

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

are the autoregressive and the moving average polynomials respectively, with $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \in \mathbb{R}, \phi_p, \theta_q \neq 0$ and $(Z_m)_{m\in\mathbb{Z}}$ a *weak white noise*, that is, $(Z_m)_{m\in\mathbb{Z}}$ is an uncorrelated sequence with constant mean and constant variance. If $(Z_m)_{m\in\mathbb{Z}}$ is even an i.i.d. sequence then we call $(X_m)_{m\in\mathbb{Z}}$ a *strong ARMA process*. A natural continuous-time

---

* Correspondence to: Vicky Fasen-Hartmann, Institute of Stochastics, Englerstraße 2, D-76131 Karlsruhe, Germany.
E-mail: vicky.fasen@kit.edu

analog of this difference equation with i.i.d. noise $(Z_m)_{m\in\mathbb{Z}}$ is the formal pth order stochastic differential equation

$$a(D)Y_t = c(D)DL_t, \quad t \in \mathbb{R}, \tag{2}$$

where D denotes the differential operator with respect to t,

$$a(z) = z^p + a_1 z^{p-1} + \cdots + a_p \quad \text{and} \quad c(z) = c_0 z^q + c_1 z^{q-1} + \cdots + c_q$$

are the autoregressive and the moving average polynomials respectively, with $p > q$, and $a_1, \ldots, a_p, c_0, \ldots, c_q \in \mathbb{R}$, $a_p, c_0 \neq 0$. The process $(L_t)_{t\in\mathbb{R}}$ is a Lévy process, that is, a stochastic process with $L_0 = 0$ almost surely, independent and stationary increments and almost surely càdlàg sample paths. However, this is not the formal definition of a *CARMA*(p, q) *process* because a Lévy process is not differentiable. The idea is more that the differential operator on the autoregressive side acts like an integration operator on the moving average side. The precise definition of a CARMA process is given later. A rigorous foundation for CARMA$(p, 0)$ processes is provided in Bergstrom (1983, 1984) and for CARMA(p, q) processes in Brockwell (2001). A Lévy driven CARMA process can be defined via a controller canonical state space representation. Necessary and sufficient conditions for the existence of strictly stationary CARMA processes are given in Brockwell and Lindner (2009). From Brockwell and Lindner (2009) (see as well Thornton and Chambers, 2017) it is also well known that a discretely sampled stationary CARMA process $(Y_{mh})_{m\in\mathbb{Z}}$ ($h > 0$ fixed) admits a weak ARMA representation, but unfortunately this is in general for Lévy driven models not a strong ARMA representation. For an overview and a comprehensive list of references on CARMA processes we refer to Brockwell (2014) and Chambers *et al.* (2018).

In many situations it is more appropriate to specify a model in continuous time rather than in discrete time. In recent years the interest in these models has increased with the availability of high-frequency data in finance and turbulence but as well by irregularly spaced data, missing observations or situations when estimation and inference at various frequencies are to be carried out. It is not surprising that stationary CARMA processes are applied in many areas as, for example, signal processing and control (cf. Garnier and Wang, 2008; Larsson *et al.*, 2006), high-frequency financial econometrics (cf. Todorov, 2009) and financial mathematics (cf. Benth *et al.*, 2014a, 2014b). The first attempts for maximum-likelihood estimation of Gaussian stationary and non-stationary MCAR(p) models are going back (Harvey and Stock, 1985a, 1985b, 1989) and were further explored in the well-known paper of Zadrozny (1988). Zadrozny (1988) investigates continuous-time Brownian motion driven ARMAX models and allows stocks and flows at different frequencies, and higher-order integration. There exist a few papers dealing with the asymptotic properties of parameter estimators of discretely sampled stationary CARMA models as Schlemm and Stelzer (2012) and Brockwell *et al.* (2011) and for non-stationary CARMA models (Fasen-Hartmann and Scholz, 2019). The papers have in common that they use a quasi maximum likelihood estimator (QMLE). However, it is well known that a QMLE is sensitive to outliers and irregularities in the data. Hence, we are looking for an alternative robust approach.

In statistics the most fundamental question when considering *robustness* of an estimator is how the estimator behaves when the data does not satisfy the model assumptions (cf. Huber and Ronchetti, 2009; Maronna *et al.*, 2006; Olive, 2017). In the case of small deviations from the model assumptions a robust estimator should give estimations not too far away from the original model. The most common and best understood robustness property is *distributional robustness* where the shape of the true underlying distribution deviates slightly from the assumed model. The amount of measures for robustness is huge, for example, qualitative robustness, quantitative robustness, optimal robustness, efficiency robustness and the breakdown point, to mention only a few. In contrast to the case of i.i.d. random variables, in the case of time series, there exist several types of possible contamination of the data which makes it more difficult to characterize robustness. In particular, for AR processes it is well-known that the GM-estimator (cf. Boente *et al.*, 1987; Künsch, 1984; Martin, 1980) and the RA-estimator (cf. Ben *et al.*, 1999) satisfy different robustness properties in contrast to *M*- or *LS*-estimators which are sensitive to the presence of additive outliers (cf. Denby and Martin, 1979). However, for general ARMA models the GM-estimator and the RA-estimator are again sensitive to outliers and hence, non-robust (cf. Bustos and Yohai, 1986). Muler *et al.* (2009)

develop a robust estimation procedure for ARMA models by calculating the residuals of the ARMA models with the help of BIP-ARMA models. For their result it is essential to have a strong ARMA model. Unfortunately the results cannot easily be extended to weak ARMA models which we have in our context.

In this article, we use the indirect inference method originally proposed by Smith (1993) for nonlinear dynamic economic models. That paper was extended by Gallant and Tauchen (1996) and Gouriéroux *et al.* (1993) (see also the overview in Gouriéroux and Monfort, 1997) for models with intractable likelihood functions and moments. If the likelihood function and moments are intractable maximum likelihood estimation and generalized methods of moments are infeasible. The authors applied the indirect inference method to macroeconomics, microeconomics, finance, and auction models; see as well Monfort (1996) and Phillips and Yu (2009) for applications to continuous-time models, Gouriéroux *et al.* (2000) and Kyriacou *et al.* (2017) for applications to time series models, and Monfardini (1998) for applications to stochastic volatility models. In addition indirect inference is used for bias reduction in finite samples as, for example, in Gouriéroux *et al.* (2000), Gouriéroux *et al.* (2010), Yu (2011), Kyriacou *et al.* (2017), and do Rêgo Sousa *et al.* (2019). An alternative approach for bias correction is given in Wang *et al.* (2011) for univariate and multivariate diffusion models. There the discretization bias is set-up to have the opposite sign to the estimation bias. For estimators of the mean revision parameter based on the Euler approximation and the trapezoidal approximation for discretization, the authors calculate the bias and relate it to the estimation bias and discretization bias. Our motivation for the indirect inference method is robust estimation (cf. de Luna and Genton, 2000, 2001; Kyriacou *et al.*, 2017).

The core idea of the indirect estimation method is to avoid estimating the parameters of interest directly and instead fit an auxiliary model to the data, estimate the parameters of this auxiliary model, and then use this estimates with simulated data to construct an estimator for the original parameter of interest (see de Luna and Genton, 2001 for a schematic overview over the indirect estimation method). de Luna and Genton (2000, 2001) recognized that it is possible to construct robust estimators via this approach, even for model classes where direct robust estimation is difficult. The reason is that it is sufficient if the parameters of the auxiliary model are estimated by a robust estimation method. Therefore, de Luna and Genton (2001) present an indirect estimation procedure for strong ARMA processes (without detailed assumptions and rigorous proofs). They fit an AR(r) process to the ARMA model and estimate the parameters of the AR(r) process with a GM-estimator. We present a similar approach in our article for the estimation of the CARMA parameters. Since the discretely sampled stationary CARMA process admits a weak ARMA representation instead of a strong ARMA representation several proofs have to be added and identifiability issues have to be taken into account.

The article is structured as follows. In Section 2, we first present our parametric family of stationary CARMA processes and our model assumptions. Furthermore, we motivate that for any $r \geq 2p - 1$ any stationary CARMA process has an AR(r) representation. Then, in Section 3, we introduce the indirect estimation procedure and give sufficient criteria for indirect estimators to be consistent and asymptotically normally distributed independent of the model; we have to assume at least consistent and asymptotically normally distributed estimators in the estimation part and in the simulation part of the indirect estimation method. Since the auxiliary AR(r) parameters of the sampled CARMA process are estimated by a GM-estimator we give an introduction into GM-estimators in Section 4 and derive consistency and asymptotic normality of this estimator in our setup. Moreover, we see that the GM-estimator is still asymptotically normally distributed for CARMA processes with outliers as additive outliers and replacement outliers. Our conclusions extend the results of Bustos (1982). Finally, in Section 5, we are able to show that the indirect estimator for the parameters of the discretely observed stationary CARMA process is consistent and asymptotically normally distributed using in the estimation part a GM-estimator and in the simulation part a LS-estimator. Several robustness properties of this estimator are derived as well as qualitative robustness and a bounded influence functional. After all, an illustrative simulation study, in Section 6, shows the practical applicability of our indirect estimator and its robustness properties. Furthermore, we compare the indirect estimator with the non-robust QMLE. Conclusions are given in Section 7. The article ends with the proofs of the results in Section A.

Notation. We use as norms the Euclidean norm $\|\cdot\|$ in $\mathbb{R}^d$ and its operator norm $\|\cdot\|$ in $\mathbb{R}^{m\times d}$ which is sub-multiplicative. For a matrix $A \in \mathbb{R}^{m\times d}$ we denote by A^T its transpose. For a matrix function $f(\vartheta)$ in $\mathbb{R}^{m\times d}$

with $\vartheta \in \mathbb{R}^s$ the gradient with respect to the parameter vector ϑ is $\nabla_\vartheta f(\vartheta) = \frac{\partial \operatorname{vec}(f(\vartheta))}{\partial \vartheta^T} \in \mathbb{R}^{dm \times s}$ and similarly $\nabla^2_\vartheta f(\vartheta) = \frac{\partial \operatorname{vec}(\nabla_\vartheta f(\vartheta))}{\partial \vartheta^T} \in \mathbb{R}^{dms \times s}$. Finally, we write $\xrightarrow{\mathcal{D}}$ for weak convergence and $\xrightarrow{\mathbb{P}}$ for convergence in probability. In general C denotes a constant which may change from line to line.

2. PRELIMINARIES

2.1. The CARMA Model

In this article, we consider a parametric family of stationary CARMA processes. Let $\Theta \subseteq \mathbb{R}^{N(\Theta)}$ $(N(\Theta) \in \mathbb{N})$ be a parameter space, $p \in \mathbb{N}$ be fixed and for any $\vartheta \in \Theta$ let $a_1(\vartheta), \ldots, a_p(\vartheta), c_0(\vartheta), \ldots, c_{p-1}(\vartheta) \in \mathbb{R}$, $a_p(\vartheta) \neq 0$ and $c_j(\vartheta) \neq 0$ for some $j \in \{0, \ldots, p-1\}$. Furthermore, define

$$A_\vartheta := \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & 1 \\ -a_p(\vartheta) & -a_{p-1}(\vartheta) & \ldots & \ldots & -a_1(\vartheta) \end{pmatrix} \in \mathbb{R}^{p \times p},$$

$$q(\vartheta) = \inf\{j \in \{0, \ldots, p-1\} : c_l(\vartheta) = 0 \ \forall\, l > j\} \text{ with } \sup \emptyset := p-1,$$
$$c_\vartheta := (c_{q(\vartheta)}(\vartheta), c_{q(\vartheta)-1}(\vartheta), \ldots, c_0(\vartheta), 0, \ldots, 0)^T \in \mathbb{R}^p.$$

The CARMA process $(Y_t(\vartheta))_{t \in \mathbb{R}}$ is then defined via the controller canonical state space representation: let $(X_t(\vartheta))_{t \in \mathbb{R}}$ be a strictly stationary solution to the stochastic differential equation

$$\mathrm{d}X_t(\vartheta) = A_\vartheta X_t(\vartheta)\,\mathrm{d}t + e_p\,\mathrm{d}L_t, \quad t \in \mathbb{R}, \tag{3a}$$

where e_p denotes the pth unit vector in $\mathbb{R}^p$. Then the process

$$Y_t(\vartheta) := c_\vartheta^T X_t(\vartheta), \quad t \in \mathbb{R}, \tag{3b}$$

is said to be a (*stationary*) *CARMA process* of order $(p, q(\vartheta))$. Rewriting (3) line by line $(Y_t(\vartheta))_{t \in \mathbb{R}}$ can be interpreted as solution of the differential equation (2); see Brockwell (2001) and Marquardt and Stelzer (2007). This means that in our parametric family of CARMA processes the order of the autoregressive polynomial is fixed to p but the order of the moving average polynomial $q(\vartheta)$ may change. In addition, we investigate only stationary CARMA processes.

Furthermore, we have the discrete-time observations $Y_h, \ldots, Y_{nh}$ of the CARMA process $(Y_t)_{t \in \mathbb{R}} = (Y_t(\vartheta_0))_{t \in \mathbb{R}}$ with fixed grid distance $h > 0$. Hence, the true model parameter is ϑ_0. The aim of this article is to receive from the observations $Y_h, \ldots, Y_{nh}$ an estimator for ϑ_0. Throughout the article we will assume that the following Assumption A holds.

Assumption A.

(A.1) The parameter space Θ is a compact subset of $\mathbb{R}^{N(\Theta)}$.
(A.2) The true parameter ϑ_0 is an element of the interior of Θ.
(A.3) $\mathbb{E}[L_1] = 0$, $0 < \mathbb{E}L_1^2 = \sigma_L^2 < \infty$ and there exists a $\delta > 0$ such that $\mathbb{E}|L_1|^{4+\delta} < \infty$.
(A.4) The eigenvalues of A_ϑ have strictly negative real parts.
(A.5) For all $\vartheta \in \Theta$ the zeros of $c_\vartheta(z) = c_0(\vartheta)z^{q(\vartheta)} + c_1(\vartheta)z^{q(\vartheta)-1} + \cdots + c_{q(\vartheta)}$ are different from the eigenvalues of A_ϑ.

DOI: 10.1111/jtsa.12526 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

(A.6) For any $\vartheta, \vartheta' \in \Theta$ we have $(c_\vartheta, A_\vartheta) \neq (c_{\vartheta'}, A_{\vartheta'})$.
(A.7) For all $\vartheta \in \Theta$ the spectrum of A_ϑ is a subset of $\{z \in \mathbb{C} : -\frac{\pi}{h} < \mathrm{Im}(z) < \frac{\pi}{h}\}$ where $\mathrm{Im}(z)$ denotes the imaginary part of z.
(A.8) The maps $\vartheta \mapsto A_\vartheta$ and $\vartheta \mapsto c_\vartheta$ are three times continuous differentiable.

Remark 2.1.

(i) (A.1) and (A.2) are standard assumptions in point estimation theory.
(ii) (A.4) guarantees that there exists a stationary solution of the state process (3a) and hence, a stationary CARMA process $(Y_t(\vartheta))_{t\in\mathbb{R}}$ (see Marquardt and Stelzer, 2007). For this reason we can and will assume throughout the article that $(Y_t(\vartheta))_{t\in\mathbb{R}}$ is stationary. The assumption of a stationary CARMA process $(Y_t(\vartheta))_{t\in\mathbb{R}}$ is essential for the indirect estimation approach of this article.
(iii) A consequence of (A.4), (A.8), the compactness of Θ and the fact that the eigenvalues of a matrix are continuous functions of its entries (cf. Bernstein (2009, Fact 10.11.2)) is $\sup_{\vartheta\in\Theta} \max\{|\lambda| : \lambda \text{ is eigenvalue of } \mathrm{e}^{A_\vartheta}\} < 1$ and hence, $\sup_{\vartheta\in\Theta} \|\mathrm{e}^{A_\vartheta u}\| \leq C\mathrm{e}^{-\rho u}$ for some $C, \rho > 0$.
(iv) Due to (A.5) the state space representation (3) of the CARMA process is minimal (cf. Bernstein (2009, Proposition 12.9.3) and Hannan and Deistler (2012, Theorem 2.3.3)).
(v) A consequence of (A.5) and (A.6) is that the family of stationary CARMA processes $(Y_t(\vartheta))_{t\in\mathbb{R}}$ is identifiable from their spectral densities and in combination with (A.7) that the same is true for the discrete-time process $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ (cf. Schlemm and Stelzer (2012, Theorem 3.13)).
(vi) The CARMA process has to be sampled sufficiently finely to ensure that (A.7) holds so that the parameters can be identified from the discrete data.

In the following we denote the autocovariance function of the stationary CARMA process $(Y_t(\vartheta))_{t\in\mathbb{R}}$ as $(\gamma_\vartheta(t))_{t\in\mathbb{R}}$ which has by Schlemm and Stelzer (2012, Proposition 3.1) the form

$$\gamma_\vartheta(t) = \mathrm{Cov}(Y_{s+t}(\vartheta), Y_s(\vartheta)) = c_\vartheta^T \mathrm{e}^{A_\vartheta t} \Sigma_\vartheta c_\vartheta, \quad t \geq 0, \tag{4}$$

with $\Sigma_\vartheta = \sigma_L^2 \int_0^\infty \mathrm{e}^{A_\vartheta u} e_p e_p^T \mathrm{e}^{A_\vartheta u} \mathrm{d}u$. Due to Assumption A the autocovariance function is three times continuous differentiable as well.

2.2. The AR(r) Representation of a Stationary CARMA Process

First, we define the auxiliary AR(r) representation of the sampled CARMA process $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$.

Proposition 2.2. For every $\vartheta \in \Theta$ and every $r \geq 2p - 1$, there exists a unique

$$\pi(\vartheta) := (\pi_1(\vartheta), \dots, \pi_r(\vartheta), \sigma(\vartheta)) \in \mathbb{R}^r \times [0, \infty)$$

such that

$$U_m(\vartheta) := Y_{mh}(\vartheta) - \sum_{k=1}^{r} \pi_k(\vartheta) Y_{(m-k)h}(\vartheta) \tag{5}$$

is stationary with $\mathbb{E}[U_1(\vartheta)] = 0$, $\mathrm{Var}(U_1(\vartheta)) = \sigma^2(\vartheta)$ and

$$\mathbb{E}\left[U_m(\vartheta) Y_{(m-k)h}(\vartheta)\right] = 0 \quad \text{for } k = 1, \dots, r. \tag{6}$$

We call $\pi(\vartheta)$ the *auxiliary parameter* of the AR(r) representation of $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$.

  J. Time Ser. Anal. **41**: 620–651 (2020)
Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12526

Remark 2.3. $U_m(\vartheta)$ can be interpreted as the error of the best linear predictor of $Y_\vartheta(mh)$ in terms of $Y_{(m-1)h}(\vartheta), \dots, Y_{(m-r)h}(\vartheta)$. Per construction, however, the sequence $(U_m(\vartheta))_{m\in\mathbb{Z}}$ is not an uncorrelated sequence, $U_m(\vartheta)$ is only uncorrelated with $Y_{(m-1)h}(\vartheta), \dots, Y_{(m-r)h}(\vartheta)$.

Definition 2.4. Let $\Pi \subseteq \mathbb{R}^r \times [0,\infty)$ be the parameter space containing all possible parameter vectors of stationary AR(r) processes. The map $\pi : \Theta \to \Pi$ with $\vartheta \mapsto \pi(\vartheta)$ and $\pi(\vartheta)$ as given in Proposition 2.2 is called the *link function* or binding function.

Lemma 2.5. Let $r \geq 2p-1$. Then, $\pi(\vartheta)$ is injective and three times continuously differentiable.

Finally, due to Lemma 2.5 we suppose throughout the article:

Assumption B. Let $r \geq 2p-1$.

3. INDIRECT ESTIMATION

For fixed r, denote by $\hat{\pi}_n$ an estimator of $\pi(\vartheta_0)$ that is calculated from the observations $\mathcal{Y}^n = (Y_h, \dots, Y_{nh})$. If we were able to analytically invert the link function π and calculate $\pi^{-1}(\hat{\pi}_n)$, then $\pi^{-1}(\hat{\pi}_n)$ would be an estimator for $\vartheta_0 = \pi^{-1}(\pi(\vartheta_0))$. However, this is not possible in general since no analytic representation of π^{-1} exists. To overcome this problem, we perform a second estimation, which is based on simulations, and constitutes the other building block of indirect estimation. We fix a number $s \in \mathbb{N}$ and simulate a sample path of length sn of a Lévy process $(L_t^S)_{t\in\mathbb{R}}$ with $\mathbb{E}L_1^S = 0$ and $\mathbb{E}(L_1^S)^2 = \sigma_L^2$. Then, for a fixed parameter $\vartheta \in \Theta$ we generate a sample path of the associated CARMA process $(Y_t^S(\vartheta))_{t\in\mathbb{R}}$ using the simulated path $(L_t^S)_{t\in\mathbb{R}}$. This gives us a vector of 'pseudo-observations' $\mathcal{Y}_S^{sn}(\vartheta) = (Y_h^S(\vartheta), \dots, Y_{snh}^S(\vartheta))$ of length sn. From this observation $\mathcal{Y}_S^{sn}(\vartheta)$ we estimate again $\pi(\vartheta)$ by an estimator $\hat{\pi}_{sn}^S(\vartheta)$. The idea is now to choose that value of ϑ as estimator for ϑ_0 which minimizes a suitable distance between $\hat{\pi}_n$ and $\hat{\pi}_n^S(\vartheta)$. The formal definition is as follows.

Definition 3.1. Let $\hat{\pi}_n$ be an estimator for $\pi(\vartheta_0)$ calculated from the data $\mathcal{Y}^n$, let $\hat{\pi}_{sn}^S(\vartheta)$ be an estimator for $\pi(\vartheta)$ calculated from the pseudo-observations $\mathcal{Y}_S^{sn}(\vartheta) = (Y_h^S(\vartheta), \dots, Y_\vartheta^S(snh))$ and let $\Omega \in \mathbb{R}^{N(\Theta)\times N(\Theta)}$ be a symmetric positive definite weighting matrix. The function $\mathcal{L}_{\text{Ind}} : \Theta \to [0,\infty)$ is defined as

$$\mathcal{L}_{\text{Ind}}(\vartheta, \mathcal{Y}^n) := [\hat{\pi}_n - \hat{\pi}_{sn}^S(\vartheta)]^T \Omega [\hat{\pi}_n - \hat{\pi}_{sn}^S(\vartheta)].$$

Then, the indirect estimator for ϑ_0 is

$$\hat{\vartheta}_n^{\text{Ind}} = \arg\min_{\vartheta\in\Theta} \mathcal{L}_{\text{Ind}}(\vartheta, \mathcal{Y}^n).$$

We are able to present general conditions under which this indirect estimator is consistent and asymptotically normally distributed.

Theorem 3.2.

(a) Suppose that the following assumptions are satisfied:

(C.1) $\hat{\pi}_n \xrightarrow{\mathbb{P}} \pi(\vartheta_0)$ as $n \to \infty$.

(C.2) $\sup_{\vartheta\in\Theta} \|\hat{\pi}_n^S(\vartheta) - \pi(\vartheta)\| \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$.

Define the map

$$\mathcal{Q}_{\text{Ind}} : \Theta \to [0,\infty) \quad \text{as} \quad \vartheta \mapsto [\pi(\vartheta) - \pi(\vartheta_0)]^T \Omega [\pi(\vartheta) - \pi(\vartheta_0)]. \tag{7}$$

Then

$$\sup_{\vartheta\in\Theta}|\mathcal{L}_{\text{Ind}}(\vartheta,\mathcal{Y}^n)-\mathcal{Q}_{\text{Ind}}(\vartheta)|\xrightarrow{\mathbb{P}}0\quad\text{and}\quad\hat{\vartheta}_n^{\text{Ind}}\xrightarrow{\mathbb{P}}\vartheta_0.$$

If we replace in (C.1) and (C.2) convergence in probability by almost sure convergence then we can replace in the statement convergence in probability by almost sure convergence as well.

(b) Assume additionally to (C.1) and (C.2):

(C.3) $\sqrt{n}(\hat{\pi}_n^{\text{S}}(\vartheta)-\pi(\vartheta))\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Xi_{\text{S}}(\vartheta))$ as $n\to\infty$ for any $\vartheta\in\Theta$.

(C.4) $\sqrt{n}(\hat{\pi}_n-\pi(\vartheta_0))\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Xi_{\text{D}}(\vartheta_0))$ as $n\to\infty$.

(C.5) For any sequence $(\overline{\vartheta}_n)_{n\in\mathbb{N}}$ with $\overline{\vartheta}_n\xrightarrow{\mathbb{P}}\vartheta_0$ as $n\to\infty$ the asymptotic behaviors

$$\begin{aligned}\nabla_\vartheta\hat{\pi}_n^{\text{S}}(\overline{\vartheta}_n)&\xrightarrow{\mathbb{P}}\nabla_\vartheta\pi(\vartheta_0),\\ \nabla_\vartheta^2\hat{\pi}_n^{\text{S}}(\overline{\vartheta}_n)&=O_P(1),\end{aligned}$$

hold as $n\to\infty$ and $\nabla_\vartheta\pi(\vartheta_0)$ has full column rank $N(\Theta)$.

Then, as $n\to\infty$,

$$\sqrt{n}(\hat{\vartheta}_n^{\text{Ind}}-\vartheta_0)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Xi_{\text{Ind}}(\vartheta_0)),$$

where

$$\Xi_{\text{Ind}}(\vartheta_0)=\mathcal{J}_{\text{Ind}}(\vartheta_0)^{-1}\mathcal{I}_{\text{Ind}}(\vartheta_0)\mathcal{J}_{\text{Ind}}(\vartheta_0)^{-1}$$

with

$$\begin{aligned}\mathcal{J}_{\text{Ind}}(\vartheta_0)&=[\nabla_\vartheta\pi(\vartheta_0)]^T\Omega[\nabla_\vartheta\pi(\vartheta_0)]\quad\text{and}\\ \mathcal{I}_{\text{Ind}}(\vartheta_0)&=[\nabla_\vartheta\pi(\vartheta_0)]^T\Omega\left[\Xi_{\text{D}}(\vartheta_0)+\frac{1}{s}\Xi_S(\vartheta_0)\right]\Omega[\nabla_\vartheta\pi(\vartheta_0)].\end{aligned}$$

Gouriéroux *et al.* (1993) develop for a dynamic model as well the consistency and the asymptotic normality of the indirect estimator but under different assumptions mainly based on $\mathcal{L}_{\text{Ind}}(\vartheta,\mathcal{Y}^n)$ (see as well Smith, 1993). These results are again summarized in Gouriéroux and Monfort (1997). In the context of indirect estimation of ARMA models, de Luna and Genton (2001, p. 22) mention the asymptotic normality of their indirect estimator but without stating any regularity conditions and only referring to Gouriéroux and Monfort (1997, Proposition 4.2).

Remark 3.3.

(a) The asymptotic covariance matrix can be written as

$$\Xi_{\text{Ind}}(\vartheta_0)=\mathcal{H}(\vartheta_0)\left(\Xi_{\text{D}}(\vartheta_0)+\frac{1}{s}\Xi_S(\vartheta_0)\right)\mathcal{H}(\vartheta_0)^T,$$

where $\mathcal{H}(\vartheta_0)=[\nabla_\vartheta\pi(\vartheta_0)^T\Omega\nabla_\vartheta\pi(\vartheta_0)]^{-1}[\nabla_\vartheta\pi(\vartheta_0)]^T\Omega$. This is the analog form of de Luna and Genton (2001, eq. (4.4)).

(b) Note that the asymptotic results hold for any $r\geq 2p-1$. But increasing the auxiliary AR order does not necessarily yield better results. On the other hand, increasing s increases the efficiency. For $s\to\infty$ we receive $\Xi_{\text{Ind}}(\vartheta_0)\to\mathcal{H}(\vartheta_0)\Xi_{\text{D}}(\vartheta_0)\mathcal{H}(\vartheta_0)^T$. The best efficiency is received for $\Omega=[\Xi_{\text{D}}(\vartheta_0)]^{-1}$ in which case $\Xi_{\text{Ind}}(\vartheta_0)\xrightarrow{s\to\infty}\left[\nabla_\vartheta\pi(\vartheta_0)^T\Xi_{\text{D}}(\vartheta_0)^{-1}\nabla_\vartheta\pi(\vartheta_0)\right]^{-1}$.

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12526

Remark 3.4. A fundamental assumption for Proposition 2.2 is (A.4) resulting in the existence of stationary CARMA processes. In particular, in the case of integrated CARMA processes $(Y_t(\vartheta))_{t\in\mathbb{R}}$, where A_ϑ has eigenvalue 0, the result of Proposition 2.2 does not hold in general. For this reason the indirect estimation approach of this article cannot be extended to integrated CARMA processes which are non-stationary. Even for integrated CARMA processes it is well known that estimators for the parameter determining the integration have a n convergence instead of a $\sqrt{n}$ convergence (cf. Chambers and McCrorie, 2007; Fasen-Hartmann and Scholz, 2019; Chambers *et al.*, 2018).

Remark 3.5. The discretely observed stationary CARMA$(p, q(\vartheta))$ process $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ admits a representation as a stationary ARMA$(p, p-1)$ process with weak white noise of the form

$$\phi(B)Y_{mh}(\vartheta) = \theta(B)\epsilon_m(\vartheta), \tag{8}$$

where $\phi(z) = \prod_{i=1}^{p}(1 - e^{h\lambda_i}z)$ (the λ_i being the eigenvalues of A_ϑ), $\theta(z)$ is a monic, Schur-stable polynomial and $(\varepsilon_m(\vartheta))_{m\in\mathbb{Z}}$ is a weak white noise (see Brockwell and Lindner (2009, Lemma 2.1)), that is, $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ is a weak ARMA$(p, p-1)$ process. Such an exact discrete-time ARMA representation for multivariate CARMA processes was generalized in Thornton and Chambers (2017, Theorem 1) to possible non-stationary multivariate CARMA processes. Thus, it is as well possible to do an indirect estimation procedure by estimating the parameters of the discrete-time ARMA$(p, p-1)$ representation, for example, using maximum-likelihood estimation, instead of estimating the parameters of the auxiliary AR(r) model. Then the map π is replaced by the map π_1 which maps the parameters of the CARMA process to the coefficients of the weak ARMA$(p, p-1)$ representation of its sampled version (8). Using $\pi_1(\vartheta)$ instead of $\pi(\vartheta)$ in Theorem 3.2, Theorem 3.2 can be adapted under the same assumptions giving asymptotic normality of the indirect estimator based on the discrete-time ARMA representation of the CARMA process. In particular, it is as well possible to derive an estimation procedure for non-stationary CARMA processes. However, until now there does not exist robust estimators for the parameters of weak ARMA processes such that this approach does not give robust estimators for the parameters of the stationary CARMA process, which is the topic of this article.

4. ESTIMATING THE AUXILIARY AR(r) PARAMETERS OF A CARMA PROCESS WITH OUTLIERS

To apply the indirect estimator to a discretely sampled stationary CARMA process we need strongly consistent and asymptotically normally distributed estimators for the parameters of the auxiliary AR(r) representation. We will study generalized M- (GM-) estimators. The GM-estimator will be applied to a stationary CARMA process afflicted by outliers because we want to study some robustness properties of our estimator as well. Outliers can be thought as typical observations that do not arise because of the model structure but due to some external influence, for example, measurement errors. Therefore, a whole sample of observations which contains outliers does not come from the true model anymore but it is still close to it as long as the total number of outliers is not overwhelmingly large.

Definition 4.1. Let $g : [0,1] \to [0,1]$ be a function that satisfies $g(\gamma) - \gamma = o(\gamma)$ for $\gamma \to 0$. Let $(V_m)_{m\in\mathbb{Z}}$ be a stochastic process taking only the values 0 and 1 with

$$\mathbb{P}(V_m = 1) = g(\gamma)$$

and let $(Z_m)_{m\in\mathbb{Z}}$ be a real-valued stochastic process. The *disturbed process* $(Y^\gamma_{mh}(\vartheta))_{m\in\mathbb{Z}}$ is defined as

$$Y^\gamma_{mh}(\vartheta) = (1 - V_m)Y_{mh}(\vartheta) + V_m Z_m. \tag{9}$$

The disturbed process $(Y^\gamma_{mh}(\vartheta))_{m\in\mathbb{Z}}$ is in general not a sampled CARMA process anymore.

DOI: 10.1111/jtsa.12526 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

Remark 4.2.

(a) The interpretation of this model is that at each point $m \in \mathbb{Z}$ an outlier is observed with probability $g(\gamma)$ while the true value $Y_{mh}(\vartheta)$ is observed with probability $1 - g(\gamma)$. The model has the advantage that one can obtain both additive and replacement outliers by choosing the processes $(Z_m)_{m\in\mathbb{Z}}$ and $(V_m)_{m\in\mathbb{Z}}$ adequately. Specifically, to model replacement outliers, one assumes that $(Z_m)_{m\in\mathbb{Z}}$, $(V_m)_{m\in\mathbb{Z}}$, and $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ are jointly independent. Then, if the realization of V_m is equal to 1, the value $Y_{mh}(\vartheta)$ will be replaced by the realization of Z_m justifying the use of the name replacement outliers. On the other hand, modeling additive outliers can be achieved by taking $Z_m = Y_{mh}(\vartheta) + W_m$ for some process $(W_m)_{m\in\mathbb{Z}}$ and assuming that $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ is independent from $(V_m)_{m\in\mathbb{Z}}$. Then we have $Y^{\gamma}_{mh}(\vartheta) = Y_{mh}(\vartheta) + V_m W_m$ such that the realization of W_m is added to the realization of $Y_{mh}(\vartheta)$ if V_m is 1.

(b) Another advantage of this general outlier model is that one can easily model the temporal structure of outliers. On the one hand, if $(V_m)_{m\in\mathbb{Z}}$ is chosen as an i.i.d. sequence with $\mathbb{P}(V_m = 1) = \gamma$, then outliers typically appear isolated, that is, between two outliers there is usually a period of time where no outliers are present. On the other hand, one can also model patchy outliers by letting $(B_m)_{m\in\mathbb{Z}}$ be an i.i.d. process of Bernoulli random variables with success probability ϵ and setting $V_m = \max(B_{m-l}, \dots, B_m)$ for a fixed $l \in \mathbb{N}$. Then as $\epsilon \to 0$,

$$\mathbb{P}(V_m = 1) = 1 - (1-\epsilon)^l = l\epsilon + o(\epsilon),$$

which results in $\gamma = l\epsilon$. For ϵ sufficiently small, outliers then appear in a block of size l.

Recall the following notion:

Definition 4.3. A stationary stochastic process $Y = (Y_t)_{t\in I}$ with $I = \mathbb{R}$ or $I = \mathbb{Z}$ is called strongly (or α-) mixing if

$$\alpha_l := \sup\left\{|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \,:\, A \in \mathcal{F}^0_{-\infty},\, B \in \mathcal{F}^{\infty}_l\right\} \overset{l\to\infty}{\to} 0$$

where $\mathcal{F}^0_{-\infty} = \sigma(Y_t \,:\, t \le 0)$ and $\mathcal{F}^{\infty}_l = \sigma(Y_t \,:\, t \ge l)$. If $\alpha_l \le C\alpha^l$ for some constants $C > 0$ and $0 < \alpha < 1$ we call $Y = (Y_t)_{t\in I}$ exponentially strongly mixing.

Assumption D.

(D.1) The processes $(V_m)_{m\in\mathbb{Z}}$ and $(Z_m)_{m\in\mathbb{Z}}$ are strictly stationary with $\mathbb{E}|V_1| < \infty$ and $\mathbb{E}|Z_1| < \infty$.

(D.2) Either we have the replacement model where the processes $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$, $(V_m)_{m\in\mathbb{Z}}$, and $(Z_m)_{m\in\mathbb{Z}}$ are jointly independent, and $(V_m)_{m\in\mathbb{Z}}$ and $(Z_m)_{m\in\mathbb{Z}}$ are exponentially strongly mixing, that is, $\alpha_V(m) \le C\rho^m$ and $\alpha_Z(m) \le C\rho^m$ for some $C > 0$, $\rho \in (0,1)$ and any $m \in \mathbb{N}$. Or we have the additive model with $Z_m = Y_{mh}(\vartheta) + W_m$ where the processes $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$, $(V_m)_{m\in\mathbb{Z}}$ and $(W_m)_{m\in\mathbb{Z}}$ are jointly independent, and $(V_m)_{m\in\mathbb{Z}}$ and $(W_m)_{m\in\mathbb{Z}}$ are exponentially strongly mixing.

(D.3) For all $a \in \mathbb{R}, \pi \in \mathbb{R}^r$ with $|a| + \|\pi\| > 0$:

$$\mathbb{P}(aY^{\gamma}_{(r+1)h}(\vartheta) + \pi_1 Y^{\gamma}_{rh}(\vartheta) + \cdots + \pi_r Y^{\gamma}_{h}(\vartheta) = 0) = 0.$$

We largely follow the ideas of Bustos (1982) for the GM-estimation of AR(r) parameters, however our model and our assumptions are slightly different. Assumption D corresponds to Bustos (1982, Assumption (M2), (M4), (M5)). The main difference is that the sampled stationary CARMA process $(Y_{mh})_{m\in\mathbb{Z}}$ is in Bustos (1982) an infinite-order moving average process whose noise is Φ-mixing which is in general not satisfied for a sampled stationary CARMA process. However, we already know from Marquardt and Stelzer (2007, Proposition 3.34) that a stationary CARMA process is exponentially strongly mixing which is weaker than Φ-mixing. Therefore, we assume that $(V_m)_{m\in\mathbb{Z}}$, $(Z_m)_{m\in\mathbb{Z}}$, and $(W_m)_{m\in\mathbb{Z}}$ are exponentially strongly mixing instead of Φ-mixing as in Bustos (1982).

In the following we define GM-estimators. Let two functions $\phi:\mathbb{R}^r\times\mathbb{R}\to\mathbb{R}$ and $\chi:\mathbb{R}\to\mathbb{R}$ be given. Moreover, assume that we have observations $\mathcal{Y}^{n,\gamma}(\vartheta)=(Y_h^\gamma(\vartheta),Y_{2h}^\gamma(\vartheta),\dots,Y_{nh}^\gamma(\vartheta))$ from the disturbed process in (9). The parameter

$$\pi^{\mathrm{GM}}(\vartheta^\gamma)=(\pi_1^{\mathrm{GM}}(\vartheta^\gamma),\dots,\pi_r^{\mathrm{GM}}(\vartheta^\gamma),\sigma^{\mathrm{GM}}(\vartheta^\gamma))$$

is defined as the solution of the equations

$$\mathbb{E}\left[\phi\left(\begin{pmatrix}Y_h^\gamma(\vartheta)\\ \vdots\\ Y_{rh}^\gamma(\vartheta)\end{pmatrix},\frac{Y_{(r+1)h}^\gamma(\vartheta)-\pi_1Y_{rh}^\gamma(\vartheta)-\cdots-\pi_rY_h^\gamma(\vartheta)}{\sigma}\right)\begin{pmatrix}Y_h^\gamma(\vartheta)\\ \vdots\\ Y_{rh}^\gamma(\vartheta)\end{pmatrix}\right]=0, \tag{10a}$$

$$\mathbb{E}\left[\chi\left(\left(\frac{Y_{(r+1)h}^\gamma(\vartheta)-\pi_1Y_{rh}^\gamma(\vartheta)-\cdots-\pi_rY_h^\gamma(\vartheta)}{\sigma}\right)^2\right)\right]=0 \tag{10b}$$

for $(\pi_1,\dots,\pi_r,\sigma)\in\mathbb{R}^r\times(0,\infty)$. The idea is again that these are the parameters of the auxiliary AR representation of $(Y_{mh}^\gamma(\vartheta))_{m\in\mathbb{Z}}$. Note that $\pi^{\mathrm{GM}}(\vartheta^\gamma)$ depends on the processes $(V_m)_{m\in\mathbb{Z}}$ and $(Z_m)_{m\in\mathbb{Z}}$ as well. We choose not to indicate this in the notation to make the exposition more readable. For the uncontaminated process $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ we also write $\pi^{\mathrm{GM}}(\vartheta)$ instead of $\pi^{\mathrm{GM}}(\vartheta^0)$. Now, the GM-estimator $\hat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma)=(\hat{\pi}_{n,1}^{\mathrm{GM}}(\vartheta^\gamma),\dots,\hat{\pi}_{n,r}^{\mathrm{GM}}(\vartheta^\gamma),\hat{\sigma}_n^{\mathrm{GM}}(\vartheta^\gamma))$ based on ϕ and χ is defined to satisfy

$$\frac{1}{n-r}\sum_{k=1}^{n-r}\phi\left(\begin{pmatrix}Y_{kh}^\gamma(\vartheta)\\ \vdots\\ Y_{(k+r-1)h}^\gamma(\vartheta)\end{pmatrix},\frac{Y_{(k+r)h}^\gamma(\vartheta)-\hat{\pi}_{n,1}^{\mathrm{GM}}(\vartheta^\gamma)Y_{(k+r-1)h}^\gamma(\vartheta)-\cdots-\hat{\pi}_{n,r}^{\mathrm{GM}}(\vartheta^\gamma)Y_{kh}^\gamma(\vartheta)}{\hat{\sigma}_n^{\mathrm{GM}}(\vartheta^\gamma)}\right)\begin{pmatrix}Y_{kh}^\gamma(\vartheta)\\ \vdots\\ Y_{(k+r-1)h}^\gamma(\vartheta)\end{pmatrix}=0, \tag{11a}$$

$$\frac{1}{n-r}\sum_{k=1}^{n-r}\chi\left(\left(\frac{Y_{(k+r)h}^\gamma(\vartheta^\gamma)-\hat{\pi}_{n,1}^{\mathrm{GM}}(\vartheta^\gamma)Y_{(k+r-1)h}^\gamma(\vartheta)-\cdots-\hat{\pi}_{n,r}^{\mathrm{GM}}(\vartheta^\gamma)Y_{kh}^\gamma(\vartheta)}{\hat{\sigma}_n^{\mathrm{GM}}(\vartheta^\gamma)}\right)^2\right)=0. \tag{11b}$$

Throughout the article we assume that there exists a solution of (11) although this is not always the case in practice.

Example 4.4.

(a) There are two main classes of GM-estimators, the so-called Mallows estimators and the Hampel–Krasker–Welsch estimators. More information on them can be found in Bustos (1982), Denby and Martin (1979), Martin (1980), and Martin and Yohai (1986). In the literature, this kind of estimators sometimes appear under the name BIF (for bounded influence) estimators. The class of Mallows estimators are defined as $\phi(y,u)=w(y)\psi(u)$, where w is a strictly positive weight function and ψ is a suitably chosen robustifying function. The Hampel–Krasker–Welsch estimators are of the form

$$\phi(y,u)=\frac{\psi(w(y)u)}{w(y)},$$

where w is a weight function and ψ is again a suitably chosen bounded function.

(b) Typical choices for ψ are the Huber ψ_k-functions (cf. Maronna et al. (2006, eq. (2.28))). Those functions are defined as $\psi_k(u)=\mathrm{sign}(u)\min\{|u|,k\}$ for a constant $k>0$. A possibility for w is, for example, $w(y)=\psi_k(|y|)/|y|$ for a Huber function ψ_k. Another choice for ψ is the so-called Tukey bisquare (or biweight) function which is given by

$$\psi(u) = u\left(1 - \frac{u^2}{k^2}\right)^2 \mathbb{1}_{\{|u| \le k\}},$$

where k is a tuning constant.

(c) For the function χ, a possibility is $\chi(x^2) = \psi^2(x) - \mathbb{E}_Z[\psi^2(Z)]$ with the same ψ function as in the definition of ϕ. The random variable Z is suitably distributed.

To develop an asymptotic theory and to obtain a robust estimator it is necessary to impose assumptions on ϕ and χ analogous to Bustos (1982, (E1)–(E6)) which we will do next:

Assumption E. Suppose $\phi : \mathbb{R}^r \times \mathbb{R} \to \mathbb{R}$ and $\chi : \mathbb{R} \to \mathbb{R}$ satisfy the following assumptions:

(E.1) For each $y \in \mathbb{R}^r$, the map $u \mapsto \phi(y, u)$ is odd, uniformly continuous and $\phi(y, u) \ge 0$ for $u \ge 0$.

(E.2) $(y, u) \mapsto \phi(y, u)y$ is bounded and there exists a $c > 0$ such that

$$|\phi(y, u)y - \phi(z, u)z| \le c\|y - z\| \quad \text{for all } u \in \mathbb{R}.$$

(E.3) The map $u \mapsto \frac{\phi(y,u)}{u}$ is non-increasing for $y \in \mathbb{R}^r$ and there exists a $u_0 \in \mathbb{R}$ such that $\frac{\phi(y,u_0)}{u_0} > 0$.

(E.4) $\phi(y, u)$ is differentiable with respect to u and the map $u \mapsto \frac{\partial\phi(y,u)}{\partial u}$ is continuous, while $(y, u) \mapsto \frac{\partial\phi(y,u)}{\partial u}y$ is bounded.

(E.5) $\mathbb{E}\left[\sup_{u\in\mathbb{R}}\left\{u\left(\frac{\partial}{\partial u}\phi\left(\begin{pmatrix} Y_h^\gamma(\vartheta) \\ \vdots \\ Y_{rh}^\gamma(\vartheta)\end{pmatrix}, u\right)\right)\left\|\begin{pmatrix} Y_h^\gamma(\vartheta) \\ \vdots \\ Y_{rh}^\gamma(\vartheta)\end{pmatrix}\right\|\right\}\right] < \infty.$

(E.6) χ is bounded and increasing on $\{x : -a \le \chi(x) < b\}$ where $b = \sup_{x\in\mathbb{R}} \chi(x)$ and $a = -\chi(0)$. Furthermore, χ is differentiable and $x \mapsto x\chi'(x^2)$ is continuous and bounded. Lastly, $\chi(u_0^2) > 0$.

In the remaining of this section we always assume that Assumptions D and E are satisfied.

Remark 4.5. As pointed out in Bustos (1982, p. 497) one can deduce from Maronna and Yohai (1981, Theorem 2.1) that there exists a solution $\pi^{\mathrm{GM}}(\vartheta^\gamma) \in \mathbb{R}^r \times (0, \infty)$ of Equation (10) if Assumption E holds. Moreover, there exists a compact set $K \subset \mathbb{R}^r \times (0, \infty)$ with $\pi^{\mathrm{GM}}(\vartheta^\gamma) \in K$ and for any $\pi \in K^c$ Equation (10) does not hold (see Bustos (1982, p. 500)).

In general it is not easy to verify that $\pi^{\mathrm{GM}}(\vartheta^\gamma)$ is unique. Additionally, one would like to have that $\pi^{\mathrm{GM}}(\vartheta^0) = \pi^{\mathrm{GM}}(\vartheta) = \pi(\vartheta)$ are the parameters of the auxiliary AR(r) model in the case that the GM-estimator is applied to realizations of an uncontaminated sampled stationary CARMA process $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$. The following proposition gives a sufficient condition.

Proposition 4.6. Suppose that $U_{r+1}(\vartheta)$ as defined in Equation (5) satisfies

$$(U_{r+1}(\vartheta), Y_{rh}(\vartheta), \dots, Y_h(\vartheta)) \overset{\mathcal{D}}{=} (-U_{r+1}(\vartheta), Y_{rh}(\vartheta), \dots, Y_h(\vartheta)). \tag{12}$$

Assume further that the function $u \mapsto \phi(y, u)$ is non-decreasing and strictly increasing for $|u| \le u_0$, where u_0 satisfies Assumptions (E.3) and (E.6), and the function χ is chosen in such a way that

$$\mathbb{E}\left[\chi\left(\left(\frac{U_1(\vartheta)}{\sigma(\vartheta)}\right)^2\right)\right] = 0. \tag{13}$$

Finally, assume that $\gamma = 0$ so that $(Y^{\gamma}_{mh}(\vartheta))_{m\in\mathbb{Z}} = (Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$. Then the auxiliary parameter $\pi(\vartheta)$ as defined in Proposition 2.2 is the unique solution of (10), that is, $\pi^{\mathrm{GM}}(\vartheta^0) = \pi(\vartheta)$.

Remark 4.7.

(a) Assumption (12) holds if the distribution of $U_{r+1}(\vartheta)$ is symmetric and $U_{r+1}(\vartheta)$ is independent of $(Y_{rh}(\vartheta), \ldots, Y_h(\vartheta))$. This again is satisfied if $(L_t)_{t\in\mathbb{R}}$ is a Brownian motion.
(b) The monotonicity assumption on ϕ is valid, for example, for both the Mallows and Hampel–Krasker–Welsch estimators when the function ψ is chosen as a Huber ψ_k-function with $u_0 = k$.
(c) The assumption on χ is fulfilled, for example, if χ is chosen as in Example 4.4(c) with $Z \overset{\mathcal{D}}{=} U_1(\vartheta)/\sqrt{\mathrm{Var}(U_1(\vartheta))}$. In the case that the driving Lévy process is a Brownian motion this means that $Z \sim \mathcal{N}(0, 1)$.

Theorem 4.8. Suppose that there exists a unique solution $\pi^{\mathrm{GM}}(\vartheta^{\gamma})$ of (10). Then $\widehat{\pi}^{\mathrm{GM}}_n(\vartheta^{\gamma}) \overset{n\to\infty}{\to} \pi^{\mathrm{GM}}(\vartheta^{\gamma})$ $\mathbb{P}$-a.s.

The proof goes in the same vein as the proof of Bustos (1982, Theorem 2.1) and is therefore omitted.

Next, we would like to deduce the asymptotic normality of the GM-estimator. Let the set K be given as in Remark 4.5 and for $\pi = (\pi_1, \ldots, \pi_r, \sigma) \in K$ define

$$\mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^{\gamma}) = \begin{pmatrix} \mathbb{E}\left[\phi\left(\begin{pmatrix} Y^{\gamma}_h(\vartheta) \\ \vdots \\ Y^{\gamma}_{rh}(\vartheta)\end{pmatrix}, \frac{Y^{\gamma}_{(r+1)h}(\vartheta) - \pi_1 Y^{\gamma}_{rh}(\vartheta) - \cdots - \pi_r Y^{\gamma}_h(\vartheta)}{\sigma}\right)\begin{pmatrix} Y^{\gamma}_h(\vartheta) \\ \vdots \\ Y^{\gamma}_{rh}(\vartheta)\end{pmatrix}\right] \\ \mathbb{E}\left[\chi\left(\left(\frac{Y^{\gamma}_{(r+1)h}(\vartheta) - \pi_1 Y^{\gamma}_{rh}(\vartheta) - \cdots - \pi_r Y^{\gamma}_h(\vartheta)}{\sigma}\right)^2\right)\right] \end{pmatrix}. \tag{14}$$

For the proof of the asymptotic normality of the GM estimator we use a Taylor expansion of $\mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^{\gamma})$ at $\pi^{\mathrm{GM}}(\vartheta^{\gamma})$. With the knowledge of the asymptotic behavior $\mathcal{Q}_{\mathrm{GM}}(\widehat{\pi}^{\mathrm{GM}}_n(\vartheta^{\gamma}), \vartheta^{\gamma})$ and $\nabla_{\pi}\mathcal{Q}_{\mathrm{GM}}(\widehat{\pi}^{\mathrm{GM}}_n(\vartheta^{\gamma}), \vartheta^{\gamma})$ it is then straightforward to derive the asymptotic behavior of the GM-estimator $\widehat{\pi}^{\mathrm{GM}}_n(\vartheta^{\gamma})$.

We need the following auxiliary result which is the analog of Bustos (1982, Lemma 3.1) under our different model assumptions.

Lemma 4.9. Define the map $\Psi : \mathbb{R}^{r+1} \times \mathbb{R}^r \times (0, \infty) \to \mathbb{R}^{r+1}$ as

$$\Psi(y, \pi) = \begin{pmatrix} \phi\left(\begin{pmatrix} y_1 \\ \vdots \\ y_r\end{pmatrix}, \frac{y_{r+1} - \pi_1 y_r - \cdots - \pi_r y_1}{\sigma}\right)\begin{pmatrix} y_1 \\ \vdots \\ y_r\end{pmatrix} \\ \chi\left(\left(\frac{y_{r+1} - \pi_1 y_r - \cdots - \pi_r y_1}{\sigma}\right)^2\right) \end{pmatrix}.$$

Furthermore, define the stochastic process $\Psi(\vartheta^{\gamma}) = (\Psi_k(\vartheta^{\gamma}))_{k\in\mathbb{N}}$ as $\Psi_k(\vartheta^{\gamma}) = \Psi(Y^{\gamma}_{kh}(\vartheta), \ldots, Y^{\gamma}_{(k+r+1)h}(\vartheta), \pi^{\mathrm{GM}}(\vartheta^{\gamma}))$. Then

$$\frac{1}{\sqrt{n-r}}\sum_{k=1}^{n-r}\Psi_k(\vartheta^{\gamma}) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \mathcal{I}_{\mathrm{GM}}(\vartheta^{\gamma})),$$

where the (i, j)th component of $\mathcal{I}_{\mathrm{GM}}(\vartheta^{\gamma})$ is

$$[\mathcal{I}_{\mathrm{GM}}(\vartheta^{\gamma})]_{ij} = \mathbb{E}\left[\Psi_{1,i}(\vartheta^{\gamma})\Psi_{1,j}(\vartheta^{\gamma})\right] + 2\sum_{k=1}^{\infty}\mathbb{E}\left[\Psi_{1,i}(\vartheta^{\gamma})\Psi_{1+k,j}(\vartheta^{\gamma})\right] \tag{15}$$

DOI: 10.1111/jtsa.12526 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

and $\Psi_{k,i}(\vartheta^\gamma)$ denotes the ith component of $\Psi_k(\vartheta^\gamma)$, $i = 1, \dots, r+1$. Especially, each $[\mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma)]_{ij}$ is finite for $i, j \in \{1, \dots, r+1\}$.

First, we derive the asymptotic behavior of the gradient $\nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi_n, \vartheta^\gamma)$.

Lemma 4.10. Let $\mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ be defined as in (14). Then the gradient $\nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ exists. Moreover, for any sequence $(\pi_n)_{n\in\mathbb{N}}$ with $\pi_n \xrightarrow{\mathbb{P}} \pi^{\mathrm{GM}}(\vartheta^\gamma)$ as $n \to \infty$ we have as $n \to \infty$,

$$\nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi_n, \vartheta^\gamma) \xrightarrow{\mathbb{P}} \nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi^{\mathrm{GM}}(\vartheta^\gamma), \vartheta^\gamma).$$

Next, we deduce the asymptotic normality of $\mathcal{Q}_{\mathrm{GM}}(\widehat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma), \vartheta^\gamma)$.

Lemma 4.11. Let $\mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ be defined as in (14) and suppose that $\nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ is non-singular. Furthermore, let $\mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma)$ be given as in (15) and suppose that $\widehat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma) \xrightarrow{\mathbb{P}} \pi^{GM}(\vartheta^\gamma)$ as $n \to \infty$. Then, as $n \to \infty$,

$$\sqrt{n-r}\,\mathcal{Q}_{\mathrm{GM}}(\widehat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma), \vartheta^\gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma)).$$

The following analog version of Bustos (1982, Theorem 2.2) holds in our setting which gives the asymptotic normality of the GM-estimator.

Theorem 4.12. Let $\mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ be defined as in (14) and suppose that $\mathcal{J}_{\mathrm{GM}}(\vartheta^\gamma) := \nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ is non-singular. Furthermore, let $\mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma)$ be given as in (15) and suppose that $\widehat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma) \xrightarrow{\mathbb{P}} \pi^{GM}(\vartheta^\gamma)$ as $n \to \infty$. Then, as $n \to \infty$,

$$\sqrt{n-r}(\widehat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma) - \pi^{\mathrm{GM}}(\vartheta^\gamma)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Xi_{\mathrm{GM}}(\vartheta^\gamma)),$$

where

$$\Xi_{\mathrm{GM}}(\vartheta^\gamma) := [\mathcal{J}_{\mathrm{GM}}(\vartheta^\gamma)]^{-1} \mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma) [\mathcal{J}_{\mathrm{GM}}(\vartheta^\gamma)]^{-1}. \tag{16}$$

5. THE INDIRECT ESTIMATOR FOR THE CARMA PARAMETERS

5.1. Asymptotic Normality

In Section 3 we already introduced the indirect estimator and presented in Theorem 3.2 sufficient criteria for the indirect estimator to be consistent and asymptotically normally distributed. In the following we want to show that these assumptions are satisfied in the setting of discretely sampled CARMA processes when we use as estimator $\widehat{\pi}_n^S(\vartheta)$ in the simulation part the least-squares- (LS-) estimator $\widehat{\pi}_n^{\mathrm{LS}}(\vartheta)$ and for $\widehat{\pi}_n$ the GM-estimator $\widehat{\pi}_n^{\mathrm{GM}}(\vartheta_0)$.

Definition 5.1. Based on the sample $\mathcal{Y}_S^{sn}(\vartheta) = (Y_h^S(\vartheta), \dots, Y_{snh}^S(\vartheta))$ the LS-estimator $\widehat{\pi}_{sn}^{\mathrm{LS}}(\vartheta) = (\widehat{\pi}_{sn,1}^{\mathrm{LS}}(\vartheta), \dots, \widehat{\pi}_{sn,r}^{\mathrm{LS}}(\vartheta), \widehat{\sigma}_{sn}^{\mathrm{LS}}(\vartheta))$ of $\pi(\vartheta)$ minimizes

$$\mathcal{L}_{\mathrm{LS}}(\pi, \mathcal{Y}_S^{sn}(\vartheta)) := \frac{1}{sn-r} \sum_{k=1}^{sn-r} \left(Y_{(k+r)h}^S(\vartheta) - \pi_1 Y_{(k+r-1)h}^S(\vartheta) - \cdots - \pi_r Y_{kh}^S(\vartheta) \right)^2 \tag{17}$$

in $\Pi' := \pi(\Theta)$ and $\widehat{\sigma}_{sn}^{\mathrm{LS}}(\vartheta)$ is defined as

$$\widehat{\sigma}^2_{\mathrm{LS},sn}(\vartheta) = \frac{1}{sn-r} \sum_{k=1}^{sn-r} \left(Y_{(k+r)h}^S(\vartheta) - \widehat{\pi}_{sn,1}^{\mathrm{LS}}(\vartheta) Y_{(k+r-1)h}^S(\vartheta) - \cdots - \widehat{\pi}_{sn,r}^{\mathrm{LS}}(\vartheta) Y_{kh}^S(\vartheta) \right)^2.$$

Remark 5.2. The quasi ML-function for the auxiliary AR(r) parameters of the discretely sampled CARMA process is defined as

$$\mathcal{L}_{\mathrm{QMLE}}(\pi, \mathcal{Y}_S^{sn}(\vartheta)) = \frac{1}{sn-r}\sum_{k=1}^{sn-r}\left(\log(\sigma^2) + \frac{(Y^S_{(k+r)h}(\vartheta) - \pi_1 Y^S_{(k+r-1)h}(\vartheta) - \cdots - \pi_r Y^S_{kh}(\vartheta))^2}{\sigma^2}\right)$$

and the quasi ML-estimator as $\hat{\pi}_{sn}^{\mathrm{QMLE}}(\vartheta) = \arg\min_{\pi\in\Pi'} \mathcal{L}_{\mathrm{QMLE}}(\pi, \mathcal{Y}_S^{sn}(\vartheta))$. It is well known that for the estimation of AR(r) parameters the ML-estimator and the LS-estimator are equivalent (this can be seen by straightforward calculations taking the derivatives of the ML-function $\mathcal{L}_{\mathrm{QMLE}}$ which are proportional to the derivatives of $\mathcal{L}_{\mathrm{LS}}$).

Theorem 5.3. Let Assumption A, B, D, and E hold. Suppose that the unique solution $\pi^{\mathrm{GM}}(\vartheta_0)$ of (10) for $(Y_{mh})_{m\in\mathbb{Z}}$ is $\pi(\vartheta_0)$, that $\nabla_\vartheta \pi(\vartheta_0)$ has full column rank $N(\Theta)$ and that $\mathcal{J}_{\mathrm{GM}}(\vartheta_0)$ is non-singular. Further, assume that $\mathbb{E}|L_1^S|^{2N^*}$ for some $N^* \in \mathbb{N}$ with $2N^* > \max(N(\Theta), 4+\delta)$. If $\hat{\pi}_n^S(\vartheta) = \hat{\pi}_n^{\mathrm{LS}}(\vartheta)$ and $\hat{\pi}_n = \hat{\pi}_n^{\mathrm{GM}}(\vartheta_0)$ then the indirect estimator $\hat{\vartheta}_n^{\mathrm{Ind}}$ is weakly consistent and

$$\sqrt{n}(\hat{\vartheta}_n^{\mathrm{Ind}} - \vartheta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Xi_{\mathrm{Ind}}(\vartheta_0)),$$

where

$$\Xi_{\mathrm{Ind}}(\vartheta_0) = \mathcal{J}_{\mathrm{Ind}}(\vartheta_0)^{-1}\mathcal{I}_{\mathrm{Ind}}(\vartheta_0)\mathcal{J}_{\mathrm{Ind}}(\vartheta_0)^{-1}$$

with

$$\begin{aligned}\mathcal{J}_{\mathrm{Ind}}(\vartheta_0) &= [\nabla_\vartheta \pi(\vartheta_0)]^T \Omega [\nabla_\vartheta \pi(\vartheta_0)] \quad \text{and}\\ \mathcal{I}_{\mathrm{Ind}}(\vartheta_0) &= [\nabla_\vartheta \pi(\vartheta_0)]^T \Omega \left[\Xi_{\mathrm{GM}}(\vartheta_0) + \frac{1}{s}\Xi_{LS}(\vartheta_0)\right]\Omega[\nabla_\vartheta \pi(\vartheta_0)],\end{aligned}$$

where the matrix $\Xi_{\mathrm{LS}}(\vartheta)$ is defined as in (16) with $\phi(y,u) = u$ and $\chi(x) = x-1$.

We have already proven that (C.1) and (C.4) of Theorem 3.2 are satisfied. To show the remaining conditions on the LS-estimator $\hat{\pi}_n^{\mathrm{LS}}(\vartheta)$ we require several auxiliary results. The remaining of this section is devoted to that.

Sufficient conditions for (C.2) and (C.5) are the weak uniform convergence of the LS-estimator and its derivatives. Since the LS-estimator is defined via the sample autocovariance function we first derive the uniform weak convergence of the sample autocovariance function and its derivatives.

Proposition 5.4. For $j, l \in \{0, \ldots, r\}$ define

$$\hat{\gamma}_{\vartheta,n}(l,j) = \frac{1}{n-r}\sum_{k=1}^{n-r} Y_{(k+l)h}(\vartheta) Y_{(k+j)h}(\vartheta).$$

Then for $i, u \in \{1, \ldots, N(\Theta)\}$ the following statements hold.

(a) $\sup_{\vartheta\in\Theta} |\hat{\gamma}_{\vartheta,n}(l,j) - \gamma_\vartheta(l-j)| \xrightarrow{\mathbb{P}} 0$.
(b) $\sup_{\vartheta\in\Theta} \left|\frac{\partial}{\partial\vartheta_i}\hat{\gamma}_{\vartheta,n}(l,j) - \frac{\partial}{\partial\vartheta_i}\gamma_\vartheta(l-j)\right| \xrightarrow{\mathbb{P}} 0$.
(c) $\sup_{\vartheta\in\Theta} \left|\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_u}\hat{\gamma}_{\vartheta,n}(l,j) - \frac{\partial}{\partial\vartheta_i\partial\vartheta_u}\gamma_\vartheta(l-j)\right| \xrightarrow{\mathbb{P}} 0$.

Then the proof of (C.2) follows from Proposition 5.5.

Proposition 5.5.

(a) $\sup_{\vartheta \in \Theta} |\widehat{\pi}_n^{\mathrm{LS}}(\vartheta) - \pi(\vartheta)| \stackrel{\mathbb{P}}{\rightarrow} 0$.
(b) $\sup_{\vartheta \in \Theta} |\nabla_\vartheta \widehat{\pi}_n^{\mathrm{LS}}(\vartheta) - \nabla_\vartheta \pi(\vartheta)| \stackrel{\mathbb{P}}{\rightarrow} 0$.
(c) $\sup_{\vartheta \in \Theta} |\nabla_\vartheta^2 \widehat{\pi}_n^{\mathrm{LS}}(\vartheta) - \nabla_\vartheta^2 \pi(\vartheta)| \stackrel{\mathbb{P}}{\rightarrow} 0$.

A direct consequence from this is the next corollary.

Corollary 5.6. Let $\overline{\vartheta}_n$ be a sequence in Θ with $\overline{\vartheta}_n \stackrel{\mathbb{P}}{\rightarrow} \vartheta_0$. Then the following statements hold:

(a) $\widehat{\pi}_n^{\mathrm{LS}}(\overline{\vartheta}_n) \stackrel{\mathbb{P}}{\rightarrow} \pi(\vartheta_0)$.
(b) $\nabla_\vartheta \widehat{\pi}_n^{\mathrm{LS}}(\overline{\vartheta}_n) \stackrel{\mathbb{P}}{\rightarrow} \nabla_\vartheta \pi(\vartheta_0)$.
(c) $\nabla_\vartheta^2 \widehat{\pi}_n^{\mathrm{LS}}(\overline{\vartheta}_n) \stackrel{\mathbb{P}}{\rightarrow} \nabla_\vartheta^2 \pi(\vartheta_0)$.

This corollary already gives (C.5).

Finally, (C.3) is a consequence of Proposition 5.7 which gives the asymptotic normality of the LS-estimator. In principle this follows from Theorem 4.12 by interpreting the least squares estimator as a particular GM-estimator with $\phi(y, u) = u$ and $\chi(x) = x - 1$.

Proposition 5.7. For any $\vartheta \in \Theta$ the LS-estimator $\widehat{\pi}_n^{\mathrm{LS}}(\vartheta)$ is strongly consistent and as $n \to \infty$,

$$\sqrt{n}(\widehat{\pi}_n^{\mathrm{LS}}(\vartheta) - \pi(\vartheta)) \stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}\left(0, \Xi_{\mathrm{LS}}(\vartheta)\right).$$

5.2. Robustness Properties

Roughly speaking an estimator is robust when small deviations from the nominal model have not much effect on the estimator. This property is known as qualitative robustness or resistance of the estimator and was originally introduced in Hampel (1971) for i.i.d. sequences. The same article also gives a slight extension to the case of data that are generated by permutation-invariant distributions, introducing the term π-robustness (Hampel (1971, p. 1893)). Of course, time series do not satisfy the assumption of permutation invariance in general. Therefore, there have been various attempts to generalize the concept of qualitative robustness to the time series setting. Boente *et al.* (1987, Theorem 3.1) prove that their π_{d_n}-robustness for time series is equivalent to Hampel's π-robustness for i.i.d. random variables and therefore, extends Hampel's π-robustness. They go ahead and define the term resistance as well. The concept of resistance has the intuitive appeal of making a statement about changes in the values of the estimator when comparing two deterministic samples. In contrast, π_{d_n}-robustness is only a statement concerning the distribution of the estimator, which is in general not easily tractable. The indirect estimator is weakly resistant and π_{d_n}-robust. The explicit definitions and the derivation of these properties for our indirect estimator are given in Section B.1 in the Supporting information.

Intuitively speaking, the influence functional measures the change in the asymptotic bias of an estimator caused by an infinitesimal amount of contamination in the data. This measure of robustness was originally introduced as influence curve by Hampel (1974) for i.i.d. processes. It was later generalized to the time series context by Künsch (1984) who explicitly studies the estimation of autoregressive processes. However, in the paper of Künsch only estimators which depend on a finite-dimensional marginal distribution of the data-generating process and a very specific form of contaminations are considered. To remedy this, a further generalization was then made by Martin and Yohai (1986) who consider the influence functional and explicitly allow for the estimators to depend on the measure of the process which makes more sense in the time series setup (cf. Martin and Yohai (1986, section 4)).

In the sense of Martin and Yohai (1986, section 4) the indirect estimator has a bounded influence functional; see Section B.2 in the Supporting information.

The breakdown point is (for a sample of data with fixed length n) the maximum percentage of outliers which can be contained in the data without 'ruining' the estimator. In this sense, it measures how much the observed data can deviate from the nominal model before catastrophic effects in the estimation procedure happen. However, the formal definition depends on the model and the estimator. Maronna and Yohai (1991) and Maronna *et al.* (1979) deal explicitly with the breakdown point of GM-estimators in regression models and Martin and Yohai (1985) and Martin (1980) study it in the time series context. A very general definition of the breakdown point is given in Genton and Lucas (2003, Definitions 1 and 2). Heuristically speaking, the fundamental idea of that definition is that the breakdown point is the smallest amount of outlier contamination with the property that the performance of the estimator does not get worse anymore if the contamination is increased further. As already mentioned in Martin (1980, p. 239) (the proof is given in the unpublished paper of Martin and Jong (1977)), and later in de Luna and Genton (2001, p. 377) and Genton and Lucas (2003, p. 89), the breakdown point of the GM-estimator applied to estimate the parameters of an AR(r) process is $1/(r+1)$. Hence, the breakdown point of our indirect estimator is as well $1/(r+1)$ since the other building block of the indirect estimator, the estimator $\hat{\pi}_n^{\mathrm{S}}(\vartheta)$ is applied to a simulated outlier-free sample.

6. SIMULATION STUDY

We simulate CARMA processes on the interval $[0, n]$ and choose a sampling distance of $h = 1$, resulting in n observations of the discrete-time process. The simulated processes are driven either by a standard Brownian motion or by a univariate NIG (normal inverse Gaussian) Lévy process. The increments of a NIG-Lévy process $L(t) - L(t-1)$ have the density

$$f_{NIG}(x; \mu, \alpha, \beta, \delta) = \frac{\alpha\delta}{\pi} \exp\left(\delta\sqrt{\alpha^2 - \beta^2} + \beta x\right) \frac{K_1(\alpha\sqrt{\delta^2 + x^2})}{\sqrt{\delta^2 + x^2}}, \quad x \in \mathbb{R},$$

$\mu \in \mathbb{R}$ is a location parameter, $\alpha \geq 0$ is a shape parameter, $\beta \in \mathbb{R}$ is a symmetry parameter, and K_1 is the modified Bessel function of the third kind with index 1. The variance of the process is then $\sigma_L^2 = \delta\alpha^2/(\alpha^2 - \beta^2)^{\frac{3}{2}}$. For the NIG Lévy process we use the parameters $\alpha = 3$, $\beta = 1$, $\delta = 2.5145$, and $\mu = -0.8890$. These parameters result in a zero-mean Lévy process with variance approximately 1 which allows for comparison of the results to the standard Brownian motion case. For the outlier model we choose additive outliers where the process $(V_m)_{m\in\mathbb{Z}}$ is a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(V_1 = 1) = \gamma$. The process $(Z_m)_{m\in\mathbb{Z}}$ is $Z_m = \xi$ for $m \in \mathbb{Z}$ where ξ and γ take different values in different simulations.

The indirect estimator is defined as in Section 5. We take $\hat{\pi}_n$ as GM-estimator $\hat{\pi}_n^{\mathrm{GM}}(\vartheta_0)$ using the `R` software. The `R` software provides the prebuilt function `arGM` in the package `robKalman` for applying GM-estimators to AR processes. This function uses a Mallows estimator as in Example 4.4(a). The weight function $w(y)$ is the Tukey bisquare function from Example 4.4(b) applied to $\|y\|$, for the function $\psi(u)$ the user can choose between the Huber ψ_k-function and the bisquare function. The function is implemented as an iterative least squares procedure as described by Martin (1980, p. 231ff.). We do 6 iterations using the Huber function and then 50 iterations with the bisquare function, which is the maximum number of iterations. The algorithm stops earlier if convergence is achieved. In our experiments we use $k = 4$ for the tuning constant of the ψ_k-function. In general, we set $s = 75$ to obtain the simulation-based observations $\mathcal{Y}_S^{sn}(\vartheta) = (Y_h^S(\vartheta), \dots, Y_{snh}^S(\vartheta))$ in the simulation part of the indirect procedure. The type of Lévy process used for the simulation part is of the same type as the Lévy process driving the CARMA process. For the estimator $\hat{\pi}_n^{\mathrm{S}}(\vartheta)$ we apply the least squares estimator and as weighting matrix Ω we take the identity matrix for convenience reasons. In some experiments we first estimated the asymptotic covariance matrix of the GM-estimator by the empirical covariance matrix of a suitable number of independent realizations of $\hat{\pi}_n$. Setting Ω as the inverse of that estimate did not significantly affect the procedure positively or negatively

DOI: 10.1111/jtsa.12526 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

Table I. Indirect estimation of a CARMA$(1,0)$ process with parameter $\vartheta_0 = -2$ driven by a Brownian motion with $n = 1000$

	$\xi = 0, \gamma = 0$ (uncontaminated)			$\xi = 5, \gamma = 0.01$			$\xi = 10, \gamma = 0.1$		
	Mean	Bias	Var	Mean	Bias	Var	Mean	Bias	Var
IIE $r = 1$	−2.0463	−0.0463	0.0791	−2.0481	−0.0481	0.0811	−1.7611	0.2389	0.0465
IIE $r = 2$	−2.0446	−0.0446	0.0740	−2.0675	−0.0675	0.0789	−1.7977	0.2023	0.0558
IIE $r = 3$	−2.0442	−0.0442	0.0726	−2.1006	−0.1006	0.0788	−1.8096	0.1904	0.0668

Table II. Estimation results for CARMA$(1,0)$ processes with parameter ϑ_0 driven by a Brownian motion with $n = 1000$ and $r = 1$

	$\xi = 0, \gamma = 0$ (uncontaminated)					
	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$\vartheta_0 = -2$	−2.0545	−0.0545	0.0658	−2.0463	−0.0463	0.0791
$\vartheta_0 = -0.2$	−0.2009	−0.0009	2e−04	−0.2027	−0.0027	6e−04
$\vartheta_0 = -0.02$	−0.0200	0.0000	0e+00	−0.0221	−0.0021	1e−04
	$\xi = 5, \gamma = 0.1$					
	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$\vartheta_0 = -2$	−2.3902	−0.3902	0.1989	−1.9730	0.0270	0.0727
$\vartheta_0 = -0.2$	−2.4616	−2.2616	0.0119	−0.1918	0.0082	0.0006
$\vartheta_0 = -0.02$	−1.8386	−1.8186	6e−04	−0.0210	−0.0010	1e−04
	$\xi = 5, \gamma = 0.15$					
	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$\vartheta_0 = -2$	−2.0794	−0.0794	0.2379	−2.0896	−0.0896	0.1099
$\vartheta_0 = -0.2$	−2.9602	−2.7602	0.0120	−0.1714	0.0286	0.0006
$\vartheta_0 = -0.02$	−1.9864	−1.9664	0.0240	−0.0187	0.0013	0.0001

so that the use of the convenient identity matrix seems justified. In each experiment, we calculate the indirect estimator and, for comparison purposes, the QMLE as defined in Schlemm and Stelzer (2012). For the indirect estimator as well for the QMLE we use 1000 independent samples and report on the average estimated value, the bias and the empirical variance of the parameter estimates.

First, CARMA(1,0) processes with $A_\vartheta = \vartheta$ and $c_\vartheta = 1$ for $\vartheta \in (-\infty, 0)$ are studied. These processes are of particular interest because their discretely sampled version admits an AR(1) representation. In Table I, we estimate contaminated and uncontaminated CARMA(1,0) processes with true parameter $\vartheta_0 = -2$ driven by a Brownian motion using in the indirect estimation method an auxiliary AR(r) process with $r = 1, 2, 3$. For all choices of r the indirect estimator performs very well: the estimator seems to converge, the variance is quite low. Furthermore, we get the impression that the choice of r has not a big influence on the estimation results (see as well Table D.1 in the Supporting information for a further study). Next, in Table II, we compare the indirect estimator with $r = 1$ and the QMLE for a Brownian motion driven CARMA$(1,0)$ process with either $\vartheta_0 = -2$, $\vartheta_0 = -0.2$ or $\vartheta_0 = -0.02$. For $\vartheta_0 = -0.02$ the CARMA$(1,0)$ process is not so far away from a non-stationary process. In all cases we see again that the QMLE and the indirect estimator work superb for uncontaminated CARMA(1,0) processes (top of Table II). If we allow additionally outliers in the CARMA(1,0) model, the indirect estimator performs vastly better than the QML estimator giving a much less biased estimate and lower variance. Indeed for contaminated CARMA(1,0) models with $\vartheta_0 = -0.2$ and $\vartheta_0 = -0.02$ the QML estimator is far away from the true parameter where the indirect estimator seems to converge to the true value.

   DOI: 10.1111/jtsa.12526

Table III. Estimation results for an uncontaminated CARMA(3, 1) process with parameter $\vartheta_0 = (-1, -2, -2, 0, 1)$ driven by a Brownian motion with $r = 5$

	$n = 200$					
	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$\vartheta_1 = -1$	−1.0210	−0.0210	0.0101	−1.1190	−0.1190	0.1582
$\vartheta_2 = -2$	−2.0095	−0.0095	0.0183	−2.1111	−0.1111	0.4637
$\vartheta_3 = -2$	−1.9734	0.0266	0.0218	−1.9914	0.0086	0.1012
$\vartheta_4 = 0$	0.0019	0.0019	0.0076	0.0028	0.0028	0.0208
$\vartheta_5 = 1$	0.9943	−0.0057	0.0136	0.9005	−0.0995	0.0820
	$n = 1000$					
	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$\vartheta_1 = -1$	−1.0064	−0.0064	0.0010	−1.0116	−0.0116	0.0035
$\vartheta_2 = -2$	−1.9988	0.0012	0.0015	−1.9959	0.0041	0.0052
$\vartheta_3 = -2$	−1.9948	0.0052	0.0019	−1.9981	0.0019	0.0074
$\vartheta_4 = 0$	0.0091	0.0091	0.0006	0.0044	0.0044	0.0031
$\vartheta_5 = 1$	1.0092	0.0092	0.0005	0.9976	−0.0024	0.0037
	$n = 5000$					
	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$\vartheta_1 = -1$	−1.0050	−0.0050	1e−04	−1.0013	−0.0013	5e−04
$\vartheta_2 = -2$	−2.0002	−0.0002	1e−04	−1.9993	0.0007	2e−04
$\vartheta_3 = -2$	−1.9948	0.0052	1e−04	−1.9996	0.0004	6e−04
$\vartheta_4 = 0$	0.0086	0.0086	1e−04	−0.0002	−0.0002	3e−04
$\vartheta_5 = 1$	1.0085	0.0085	0e+00	1.0004	0.0004	3e−04

In a further study we investigate a CARMA(3,1) process. This especially means that the sampled process is not a weak AR process anymore. The true parameter is $\vartheta_0 = \left(-1, -2, -2, 0, 1\right)$ and

$$A_\vartheta = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vartheta_1 & \vartheta_2 & \vartheta_3 \end{pmatrix} \in \mathbb{R}^{3\times 3} \quad \text{and} \quad c_\vartheta = (\vartheta_1, -\vartheta_3\vartheta_4 + \vartheta_5, \vartheta_4).$$

This is the echelon canonical form for CARMA(3, 2) models as presented in Schlemm and Stelzer (2012, Theorem 4.2) which satisfies Assumption A. The echelon canonical form is widely used in the VARMA context, see, for example, Lütkepohl and Poskitt (1996) and the textbooks of Lütkepohl (2005), or Hannan and Deistler (2012). For this model we choose $r = 5$, which is also the minimum order of the auxiliary AR representation to satisfy Assumption B. We also tried different values of r but they did not give better results (see Table D.1 in the Supporting information).

In the first instance, we compare the QMLE and the indirect estimator for uncontaminated CARMA(3, 1) processes in Tables III and IV respectively. In Table III the driving Lévy process is a Brownian motion where in Table IV it is a NIG-Lévy process. For the Brownian motion driven model the QML optimization failed for $n = 200$, 1000, and 5000 in 100, 74, and 20 cases respectively, where for the NIG driven model it failed in 92, 72, and 29 cases respectively. The indirect estimator never failed. The error occurs when the estimator is not an element of Θ anymore. The results in the table are averaged over experiments in which the algorithm did deliver a result, the failed attempts were discarded. The estimation results for the Brownian motion driven model and the NIG driven model are very similar. For $n = 200$ the QMLE has in most of the cases a lower absolute bias than the indirect estimator. But for $n = 1000$ and $n = 5000$ this changes and the absolute bias behaves very similar for both estimators. The QMLE has in general a lower variance than the indirect estimator but it

Table IV. Estimation results for an uncontaminated CARMA(3, 1) process with parameter $\vartheta_0 = (-1, -2, -2, 0, 1)$ driven by a NIG Lévy process with $r = 5$

	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$n = 200$						
$\vartheta_1 = -1$	−1.0218	−0.0218	0.0456	−1.1416	−0.1416	0.2758
$\vartheta_2 = -2$	−2.0056	−0.0056	0.0638	−2.12998	−0.12998	0.60993
$\vartheta_3 = -2$	−1.9919	0.0081	0.0319	−2.0162	−0.0162	0.1300
$\vartheta_4 = 0$	0.0033	0.0033	0.0092	−0.0008	−0.0008	0.0236
$\vartheta_5 = 1$	0.9971	−0.0029	0.0131	0.8884	−0.1116	0.0905
$n = 1000$						
$\vartheta_1 = -1$	−1.0054	−0.0054	0.0012	−1.0240	−0.0240	0.0202
$\vartheta_2 = -2$	−2.0030	−0.0030	0.0017	−2.0090	−0.0090	0.0568
$\vartheta_3 = -2$	−1.9916	0.0084	0.0021	−1.9878	0.0122	0.0124
$\vartheta_4 = 0$	0.0053	0.0053	0.0006	−0.0025	−0.0025	0.0030
$\vartheta_5 = 1$	1.0053	0.0053	0.0006	0.9835	−0.0165	0.0057
$n = 5000$						
$\vartheta_1 = -1$	−1.0044	−0.0044	1e−04	−1.0056	−0.0056	6e−04
$\vartheta_2 = -2$	−1.9994	0.0006	1e−04	−1.9977	0.0023	3e−04
$\vartheta_3 = -2$	−1.9962	0.0038	1e−04	−1.9971	0.0029	7e−04
$\vartheta_4 = 0$	0.0086	0.0086	1e−04	−0.0048	−0.0048	3e−04
$\vartheta_5 = 1$	1.0084	0.0084	1e−04	0.9947	−0.0053	5e−04

failed to give results in several cases. However, in general the performance of both estimators is excellent. We obtained similar results as in Tables III and IV for different parameter values (see Table D.4 in the Supporting information).

Furthermore, for the Brownian motion driven CARMA(3,1) process we estimate ϑ_0 for each of the following contamination configurations in Table V (see as well Table D.2 in the Supporting information for different values of n): $\xi = 5$ and $\gamma = 0.1$, $\xi = 10$ and $\gamma = 0.1$, $\xi = 5$ and $\gamma = 1/6$, and $\xi = 5$ and $\gamma = 0.25$. The indirect estimator performs quite well in the first three contamination cases with a low bias and a low variance. It is not surprising that for $\gamma = 1/6$ the estimation results are not as good as for $\gamma = 0.1$ because the breakdown point has for $r = 5$ the upper bound $1/6$. Hence, it is apparent that for $\gamma = 0.25$ the indirect estimator has a higher bias and variance because γ lies above the breakdown point. This is in accordance with the results of Section 5.2. However, increasing n decreases both the bias and the variance significantly (see Table D.2 in the Supporting information). The maximum likelihood estimator is severely biased and far from the true parameter value in all four scenarios. Especially the inclusion of a zero component in the true parameter seems to pose a major problem since this component is affected by the most bias. Even for $\gamma = 0.25$ the simulation results for the QMLE are worse than the results for the indirect estimator.

7. CONCLUSION

In this article we presented an indirect estimation procedure for the parameters of a discretely observed CARMA process by estimating the parameters of its auxiliary AR(r) representation using a GM-estimator. Since there does not exist an explicit form of the map between the AR parameters and the CARMA parameters, an additional simulation step to get back from the AR parameters to the CARMA parameters was necessary. Sufficient conditions were given such that the indirect estimator is consistent and asymptotically normally distributed, on the one hand, in a general context, but on the other hand, as well for the special case where $\hat{\pi}_n = \hat{\pi}_n^{\mathrm{GM}}(\vartheta_0)$ and $\hat{\pi}_n^{\mathrm{S}}(\vartheta) = \hat{\pi}_n^{\mathrm{LS}}(\vartheta)$.

Table V. Estimation results for a CARMA(3, 1) process with parameter $\vartheta_0 = (-1, -2, -2, 0, 1)$ driven by a Brownian motion with $n = 1000$ and $r = 5$

	QMLE			Indirect		
	Mean	bias	Var	Mean	bias	Var
$\xi = 5, \gamma = 0.1$						
$\vartheta_1 = -1$	−0.4392	0.5608	0.0028	−1.0109	−0.0109	0.0136
$\vartheta_2 = -2$	−2.9656	−0.9656	0.0096	−2.0348	−0.0348	0.0766
$\vartheta_3 = -2$	−2.3261	−0.3261	0.1019	−1.9860	0.0140	0.0167
$\vartheta_4 = 0$	1.9142	1.9142	0.0110	−0.0195	−0.0195	0.0059
$\vartheta_5 = 1$	1.8945	0.8945	0.0008	0.9407	−0.0593	0.0186
$\xi = 10, \gamma = 0.1$						
$\vartheta_1 = -1$	−0.0842	0.9158	0.0041	−1.0115	−0.0115	0.0161
$\vartheta_2 = -2$	−3.7754	−1.7754	2.1064	−2.0318	−0.0318	0.0418
$\vartheta_3 = -2$	−4.4277	−2.4277	63.5757	−1.9864	0.0136	0.0190
$\vartheta_4 = 0$	4.0479	4.0479	0.2430	−0.0195	−0.0195	0.0060
$\vartheta_5 = 1$	2.4057	1.4057	0.1731	0.9395	−0.0605	0.0203
$\xi = 5, \gamma = \frac{1}{6}$						
$\vartheta_1 = -1$	−0.1646	0.8354	0.0144	−0.9601	0.0399	0.0937
$\vartheta_2 = -2$	−4.3789	−2.3789	7.7940	−2.2333	−0.2333	0.5094
$\vartheta_3 = -2$	−10.7587	−8.7587	179.4913	−1.9791	0.0209	0.0745
$\vartheta_4 = 0$	2.8016	2.8016	0.2456	−0.0490	−0.0490	0.0078
$\vartheta_5 = 1$	2.2468	1.2468	0.2453	0.7249	−0.2751	0.0649
$\xi = 5, \gamma = 0.25$						
$\vartheta_1 = -1$	−0.1009	0.8991	0.0134	−0.3893	0.6107	7.0162
$\vartheta_2 = -2$	−4.6634	−2.6634	21.3373	−3.3718	−1.3718	32.4638
$\vartheta_3 = -2$	−8.2994	−6.2994	350.5529	−4.1272	−2.1272	147.8243
$\vartheta_4 = 0$	3.0630	3.0630	0.1360	1.9057	1.9057	13.9561
$\vartheta_5 = 1$	1.8534	0.8534	0.7225	1.2545	0.2545	16.2171

Moreover, the indirect estimator satisfies different robustness properties as weakly resistant, π_{d_n}-robustness and it has a bounded influence functional.

Summarizing the simulation studies, the indirect estimator performs convincingly for various orders p and q of the CARMA process, for different driving Lévy processes and for a variety of outlier configurations. Some simulations show as well that the estimator works well for other sampling frequencies (see Table D.3 in the Supporting information for $h = 2$ and $n = 2000$ where we have in total 1000 observations). For contaminated CARMA processes the indirect estimator is robust against several kind of outliers and estimates the parameters very well with a low variance. Whereas the QMLE is severely biased with a high variance. In contrast to the indirect estimator, the QMLE does not give reasonable results in the presence of outliers. Therefore, in the context of outliers the indirect estimator is preferred to the QMLE. For uncontaminated CARMA processes both the indirect estimator and the QMLE perform excellent. For small n the QMLE is less biased with a lower variance than the indirectly estimator. But obviously the bias in the indirect estimation procedure can be decreased by using in the estimation and in the simulation part the same type of estimator. Then the bias from the estimation part and the simulation part are canceling out (cf. Gouriéroux *et al.*, 2000; Gouriéroux *et al.*, 2010). But proving the asymptotic normality of the indirect estimator using as well in the simulation part the GM estimator is involved. Thus, the comparison of the indirect estimator and the QMLE needs some further exploration in the context of uncontaminated CARMA processes.

Of course, it is clear that the indirect estimator has its bounds as well for contaminated CARMA processes. Increasing γ too far eventually causes the indirect estimator to break down because we get above the breakdown point. The breakdown point $1/(r+1)$ is very low if r is large, which is, however, necessary if the order p of the CARMA(p, q) process is large. A possible alternative is instead of estimating the AR(r) parameters of the discretely observed CARMA process with the GM-estimator to estimate the weak ARMA$(p, p-1)$ parameters of the discretely observed CARMA process with the BMM-estimator of Muler *et al.* (2009) which has the largest possible breakdown point of $1/2$. However, a problem is that the results of Muler *et al.* (2009) are based on ARMA processes with i.i.d. noise where the discretely sampled CARMA process admits only an ARMA representation with an uncorrelated white noise. Therefore, this extension is not as obvious and topic of some future research.

ACKNOWLEDGEMENTS

The authors thank Thiago do Rêgo Sousa for helping with the simulation study. They also thank two anonymous referees for useful comments improving the article. Financial support by the Deutsche Forschungsgemeinschaft through the research grant FA 809/2-2 is gratefully acknowledged.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

SUPPORTING INFORMATION

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

REFERENCES

Ben MG, Martinez EJ, Yohai VJ. 1999. Robust estimation in vector autoregressive moving-average models. *Journal of Time Series Analysis* **20**(4): 381–399.
Benth FE, Klüppelberg C, Müller G, Vos L. 2014a. Futures pricing in electricity markets based on stable CARMA spot models. *Energy Economics* **44**: 392–406.
Benth FE, Koekebakker S, Zakamouline V. 2014b. The CARMA interest rate model. *International Journal of Theoretical and Applied Finance* **17**.
Bergstrom AR. 1983. Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica* **51**: 117–152.
Bergstrom AR. 1984. Continuous time stochastic models and issues of aggregation over time. In *Handbook of Econometrics*, Griliches Z., Intriligator MD. (eds.) North-Holland; 1145–1212.
Bernstein DS. 2009. *Matrix Mathematics: Theory, Facts, and Formulas*: Princeton University Press.
Billingsley P. 1999. *Convergence of Probability Measures*, 2nd ed. New York: John Wiley & Sons.
Boente G, Fraiman R, Yohai V. 1987. Qualitative robustness for stochastic processes. *The Annals of Statistics* **15**(3): 1293–1312.
Bradley RC. 2007. *Introduction to Strong Mixing Conditions*, Vol. 1. Heber City: Kendrick Press.
Brockwell PJ. 2001. Lévy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics* **53**: 113–124.
Brockwell PJ. 2014. Recent results in the theory and applications of CARMA processes. *Annals of the Institute of Statistical Mathematics* **66**: 647–685.
Brockwell PJ, Davis R. 1991. *Time Series: Theory and Methods*, 2nd ed. New York: Springer.
Brockwell PJ, Lindner A. 2009. Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Processes and their Applications* **119**: 2660–2681.
Brockwell PJ, Davis R, Yang Y. 2011. Estimation for non-negative Lévy-driven CARMA processes. *Journal of Business and Economic Statistics* **29**: 250–259.
Bustos OH. 1982. General M-estimates for contaminated pth-order autoregressive processes: consistency and asymptotic normality. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **59**(4): 491–504.
Bustos OH, Yohai VJ. 1986. Robust estimates for ARMA models. *Journal of the American Statistical Association* **81**: 155–168.
Chambers MJ, McCrorie JR. 2007. Frequency domain estimation of temporally aggregated Gaussian cointegrated systems. *Journal of Econometrics* **136**: 1–29.
Chambers MJ, McCrorie JR, Thornton MA. 2018. Continuous time modelling based on an exact discrete time representation. In *Continuous Time Modeling in the Behavioral and Related Sciences*, van Montfort K, Oud J, Voelkle M (eds.): Springer; 317–357.

de Luna X, Genton MG. 2000. Robust simulation-based estimation. *Statistics and Probability Letters* **48**: 253–259.

de Luna X, Genton MG. 2001. Robust simulation-based estimation of ARMA models. *Journal of Computational and Graphical Statistics* **10**(2): 370–387.

Denby L, Martin RD. 1979. Robust estimation of the first-order autoregressive parameter. *Journal of the American Statistical Association* **74**(365): 140–146.

do Rêgo Sousa T, Haug S, Klüppelberg C. 2019. Indirect Inference for Lévy-driven continuous-time GARCH models. *Scandinavian Journal of Statistics* **46**(3): 765–801.

Fasen-Hartmann V, Scholz M. 2019. Quasi-maximum likelihood estimation for cointegrated solutions of continuous-time state space models observed at discrete time points. *Electronic Journal of Statistics* **13**(2): 5151–5212.

Gallant RA, Tauchen G. 1996. Which moments to match?. *Econometric Theory* **12**: 657–668.

2008. Identification of Continuous-time Models from Sampled Data. In *Advances in Industrial Control* London: Springer.

Genton MG, Lucas A. 2003. Comprehensive definitions of breakdown points for independent and dependent observations. *Journal of the Royal Statistical Society, Series B* **65**(1): 81–94.

Gouriéroux C, Monfort A. 1997. *Simulation-based Econometric Methods*. Oxford: Oxford University Press.

Gouriéroux C, Monfort A, Renault E. 1993. Indirect inference. *Journal of Applied Econometrics* **8**: 85–118.

Gouriéroux C, Renault E, Touzi N. 2000. Calibration by simulation for small sample bias correction. In *Simulation-based Inference in Econometrics: Methods and Applications*, Mariano R, Schuermann T, Weeks MJ (eds.) Cambridge: Cambridge University Press; 328–358.

Gouriéroux C, Phillips PCB, Yu J. 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* **157**: 68–77.

Hampel FR. 1971. A general qualitative definition of robustness. *Annals of Mathematical Statistics*: 1887–1896.

Hampel FR. 1974. The influence curve and its role in robust estimation. *Journal of the American Statistical Association* **69**(346): 383–393.

Hannan EJ, Deistler M. 2012. *The Statistical Theory of Linear Systems*: Society for Industrial and Applied Mathematics. Reprint of the 1988 original.

Harvey AC, Stock JH. 1985a. The estimation of higher-order continuous time autoregressive models. *Econometric Theory* **1**: 97–112.

Harvey AC, Stock JH. 1985b. Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics and Control* **12**: 365–384.

Harvey AC, Stock JH. 1989. Estimating integrated higher-order continuous time autoregressions with an application to money-income causality. *Journal of Econometrics* **42**(3): 319–336.

Huber PJ, Ronchetti E. 2009. *Robust Statistics*, 2nd ed.: Wiley.

Ibragimov I. 1962. Some limit theorems for stationary processes. *Theory of Probability and Its Applications* **7**(4): 349–382.

Kimmig S. 2016. *Statistical Inference for MCARMA Processes*, PhD thesis, Karlsruher Institut of Technology, Karlsruhe.

Künsch H. 1984. Infinitesimal robustness for autoregressive processes. *The Annals of Statistics* **12**(3): 843–863.

Kyriacou M, Phillips PCB, Rossi F. 2017. Indirect inference in spatial autoregression. *Econometrics Journal* **20**: 168–189.

Larsson EK, Mossberg M, Söderström T. 2006. An overview of important practical aspects of continuous-time ARMA system identification. *Circuits, Systems, and Signal Processing* **25**: 17–46.

Lütkepohl HH. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

Lütkepohl HH, Poskitt DS. 1996. Specification of Echelon-form VARMA models. *Journal of Business and Economic Statistics* **14**: 69–79.

Maronna RA, Yohai VJ. 1981. Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **58**(1): 7–20.

Maronna RA, Yohai VJ. 1991. The breakdown point of simultaneous general M estimates of regression and scale. *Journal of the American Statistical Association* **86**(415): 699–703.

Maronna RA, Bustos OH, Yohai VJ. 1979. Bias- and efficiency-robustness of general M-estimators for regression with random carriers. In *Smoothing Techniques for Curve Estimation*: Springer; 91–116.

Maronna RA, Martin DR, Yohai VJ. 2006. *Robust Statistics*: Wiley.

Marquardt T, Stelzer R. 2007. Multivariate CARMA processes. *Stochastic Processes and their Applications* **117**(1): 96–120.

Martin DR. 1980. Robust estimation of autoregressive models. In *Directions in Time Series*, Brillinger DR, Tiao GC (eds.) Hayward, CA: Institute of Mathematical Statistics; 228–262.

Martin DR, Jong J. 1977. *Asymptotic Properties of Robust Generalized M-Estimates for the First Order Autoregressive Parameter*. Murray Hill, NJ: Unpublished Bell Labs Technical Memo.

Martin DR, Yohai VJ. 1985. Robustness in time series and estimating ARMA models. In *Handbook of Statistics*, Hannan EJ, Krishnaiah PR, Rao MM (eds.): Elsevier; 119–155.

Martin DR, Yohai VJ. 1986. Influence functionals for time series. *The Annals of Statistics* **14**(3): 781–818.

Monfardini C. 1998. Estimating stochastic volatility models through indirect inference. *The Econometrics Journal*: C113–C128.

Monfort A. 1996. A reappraisal of misspecified econometric models. *Econometric Theory* **12**(12): 597–619.

Muler N, Peña D, Yohai VJ. 2009. Robust estimation for ARMA models. *The Annals of Statistics* **37**(2): 816–840.
Olive DJ. 2017. *Robust Multivariate Analysis*. Cham: Springer.
Phillips PCB, Yu J. 2009. Maximum likelihood and Gaussian estimation of continuous time models in finance. In *Handbook of Financial Time Series*, Mikosch T, Kreiß J-P, Davis RA, Andersen TG (eds.) Berlin, Heidelberg: Springer; 497–530.
Pollard D. 1990. *Empirical Processes: Theory and Applications*, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.
Schlemm E, Stelzer R. 2012. Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electronic Journal of Statistics* **6**: 2185–2234.
Smith AA. 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**(S1): S63–S84.
Thornton MA, Chambers MJ. 2017. Continuous time ARMA processes: discrete time representation and likelihood evaluation. *Journal of Economic Dynamics and Control* **79**: 48–65.
Todorov V. 2009. Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics* **148**: 131–148.
Wang X, Phillips PCB, Yu J. 2011. Bias in estimating multivariate and univariate diffusions. *Journal of Econometrics* **161**: 228–245.
Yu J. 2011. Simulation-based estimation methods for financial time series models. In *Handbook of Computational Finance*, Duan J-C, Gentle JE, Härdle W (eds.) Berlin: Springer; 427–465.
Zadrozny P. 1988. Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* **4**(1): 108–124.

APPENDIX A. PROOFS

A.1. Proofs of Section 2

Proof of Proposition 2.2. First, we need to show that for any $r \in \mathbb{N}$ the covariance matrix of $(Y_h(\vartheta), \dots, Y_{(r+1)h}(\vartheta))$ is non-singular. To see this, note that the autocovariance function of $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ is $\gamma_\vartheta(mh) = c_\vartheta^T e^{A_\vartheta hm}\Sigma_\vartheta c_\vartheta$, $m \in \mathbb{N}_0$ (see (4)). Since Σ_ϑ is non-singular (cf. Schlemm and Stelzer (2012, Corollary 3.9)) and $c_\vartheta \neq 0_p$ we have that $\gamma_\vartheta(0) > 0$. Moreover, the eigenvalues of A_ϑ have strictly negative real parts by Assumption (A.4) and therefore, $\gamma_\vartheta(mh) \to 0$ as $m \to \infty$ holds. By Brockwell and Davis (1991, Proposition 5.1.1), it follows that the covariance matrix of $(Y_h(\vartheta), \dots, Y_{(r+1)h}(\vartheta))$ is non-singular for every $r \in \mathbb{N}$. Thus, a conclusion of Brockwell and Davis (1991, §8.1) is that there exist unique $\pi_1(\vartheta), \dots, \pi_r(\vartheta), \sigma^2(\vartheta)$ which solve the set of $r+1$ Yule–Walker equations, namely

$$
\begin{aligned}
\pi^*(\vartheta) := \begin{pmatrix} \pi_1(\vartheta) \\ \vdots \\ \pi_r(\vartheta) \end{pmatrix} &= \begin{pmatrix} \gamma_\vartheta(0) & \gamma_\vartheta(h) & \cdots & \gamma_\vartheta((r-1)h) \\ \gamma_\vartheta(h) & \gamma_\vartheta(0) & \cdots & \gamma_\vartheta((r-2)h) \\ \vdots & \vdots & & \vdots \\ \gamma_\vartheta((r-1)h) & \gamma_\vartheta((r-2)h) & \cdots & \gamma_\vartheta(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma_\vartheta(h) \\ \vdots \\ \gamma_\vartheta(rh) \end{pmatrix} \\
&=: \Gamma^{(r-1)}(\vartheta)^{-1}\gamma^{(r-1)}(\vartheta), \qquad \text{(A1a)} \\
\sigma^2(\vartheta) &= \gamma_\vartheta(0) - \pi^*(\vartheta)^T\gamma^{(r-1)}(\vartheta). \qquad \text{(A1b)}
\end{aligned}
$$

□

Proof of Lemma 2.5. We make use of the fact that the discretely observed stationary CARMA$(p, q(\vartheta))$ process $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ admits a representation as a stationary ARMA$(p, p-1)$ process with weak white noise as is given in (8). Then we can decompose the map $\pi : \Theta \to \Pi$ into three separate maps for which we define the following spaces:

$$
\begin{aligned}
\mathcal{M} &:= \{(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_{p-1}, \sigma) \in \mathbb{R}^{2p} : \text{ The coefficients define a weak ARMA}(p, p-1) \\
&\qquad \text{model as in (1) for which } \phi(z) \text{ and } \theta(z) \text{ have no common zeros}\} \subseteq \mathbb{R}^{2p}, \\
\mathcal{G} &:= \{\gamma = (\gamma_0, \dots, \gamma_r) \in \mathbb{R}^{r+1} : \text{ The coefficients define the autocovariances up to order}
\end{aligned}
$$

r of a stationary stochastic process where $\Gamma^{(r-1)}$ is non-singular$\} \subseteq \mathbb{R}^{r+1}$,

$\Pi := \{(\pi_1, \dots, \pi_r, \sigma) \in \mathbb{R}^r \times (0, \infty) : (\pi_1, \dots, \pi_r)$ are the coefficients of a stationary AR(r) process and σ^2 is the variance of the noise$\} \subseteq \mathbb{R}^{r+1}$,

where $\Gamma^{(r-1)}$ is defined as $\Gamma^{(r-1)}(\vartheta)$ in (A1a). Denote by $\pi_1 : \Theta \to \mathcal{M}$ the map which maps the parameters of a CARMA process to the coefficients of the weak ARMA$(p, p-1)$ representation of its sampled version as in (8). Denote by $\pi_2 : \mathcal{M} \to \mathcal{G}$ the map which maps the parameters of a weak ARMA$(p, p-1)$ process to its autocovariances of lags $0, \dots, r$. Lastly, denote by $\pi_3 : \mathcal{G} \to \Pi$ the map which maps a vector of autocovariances $(\gamma_0, \dots, \gamma_r)$ to the parameters of the auxiliary AR(r) model. Then we have that $\pi = \pi_3 \circ \pi_2 \circ \pi_1$. We will show that π_i is injective for $i = 1, 2, 3$ and receive from this the injectivity of π. The three-times continuous-differentiability of the map π follows from the representation (A1) and the three-times continuous-differentiability of the autocovariance function γ_ϑ.

Step 1: π_1 is injective.
Due to Assumption A and Schlemm and Stelzer (2012, Theorem 3.13) the family of sampled processes $\{(Y_{mh}(\vartheta)_{m\in\mathbb{Z}} : \vartheta \in \Theta)\}$ is identifiable from their spectral densities and hence, for any $\vartheta \neq \vartheta' \in \Theta$ the parameters of the weak ARMA process in (8) differ.

Step 2: π_2 is injective if $r \geq 2p - 1$.
The reason is that by the method of Brockwell and Davis (1991, p. 93), the autocovariances of ARMA$(p, p-1)$ processes are completely determined as solutions of difference equations with p boundary conditions which depend on the coefficient vector $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_{p-1}, \sigma)$. If $r \geq 2p-1$, the number of equations r is greater than or equal to the number of variables $2p-1$ which results in the injectivity of π_2 (see also de Luna and Genton (2001, section 4.1)). To be more precise, let $\boldsymbol{\theta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_{p-1}, \sigma) \in \mathcal{M}$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\phi}_1, \dots, \widetilde{\phi}_p, \widetilde{\theta}_1, \dots, \widetilde{\theta}_{p-1}, \widetilde{\sigma}) \in \mathcal{M}$.

Case 1. $(\phi_1, \dots, \phi_p) \neq (\widetilde{\phi}_1, \dots, \widetilde{\phi}_p)$. Define $\Gamma^{(p)}(\theta) \in \mathbb{R}^{(p+1)\times(p+1)}$ similarly to $\Gamma^{(p-1)}(\vartheta)$ in (A1a). Due to Brockwell and Davis (1991, (3.3.9))

$$(-\phi_p \quad \dots \quad -\phi_1 \quad 1)\Gamma^{(p)}(\theta) = (0 \quad 0 \quad \cdots \quad 0),$$
$$(-\widetilde{\phi}_p \quad \dots \quad -\widetilde{\phi}_1 \quad 1)\Gamma^{(p)}(\widetilde{\theta}) = (0 \quad 0 \quad \cdots \quad 0).$$

But since the vectors $(-\phi_p \quad \dots \quad -\phi_1 \quad 1)$ and $(-\widetilde{\phi}_p \quad \dots \quad -\widetilde{\phi}_1 \quad 1)$ are linear independent this is only possible if $\Gamma^{(p)}(\theta) \neq \Gamma^{(p)}(\widetilde{\theta})$ which implies $\pi_2(\theta) \neq \pi_2(\widetilde{\theta})$.

Case 2. $(\phi_1, \dots, \phi_p) = (\widetilde{\phi}_1, \dots, \widetilde{\phi}_p)$. Assume that $\pi_2(\theta) = \pi_2(\widetilde{\theta})$. But then due to Brockwell and Davis (1991, (3.3.9)), $(\gamma_\theta(k))_{k\in\mathbb{N}_0} = (\gamma_{\widetilde{\theta}}(k))_{k\in\mathbb{N}_0}$ and hence, $\theta = \widetilde{\theta}$.

Step 3: π_3 is injective.
We can also rewrite the linear Eqs. (A1) as a linear system with $(r+1)$ equations and the $(r+1)$ unknown variables $\gamma_0, \dots, \gamma_r$ which gives the injectivity of π_3. □

A.2. Proofs of Section 3

Proof of Theorem 3.2.
(a) We first start by proving the consistency. With the definition of $\mathcal{Q}_{\text{Ind}}$ we obtain

$$\begin{aligned}
&\sup_{\vartheta\in\Theta} |\mathcal{L}_{\text{Ind}}(\vartheta, \mathcal{Y}^n) - \mathcal{Q}_{\text{Ind}}(\vartheta)| \\
&= \sup_{\vartheta\in\Theta} |[\widehat{\pi}_n - \widehat{\pi}^{\text{S}}_{sn}(\vartheta)]^T \Omega [\widehat{\pi}_n - \widehat{\pi}^{\text{S}}_{sn}(\vartheta)] - [\pi(\vartheta) - \pi(\vartheta_0)]^T \Omega [\pi(\vartheta) - \pi(\vartheta_0)]|
\end{aligned}$$

DOI: 10.1111/jtsa.12526 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

$$\leq |\hat{\pi}_n^T\Omega\hat{\pi}_n - \pi(\vartheta_0)^T\Omega\pi(\vartheta_0)| + \sup_{\vartheta\in\Theta}|\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta)^T\Omega\hat{\pi}_n - \pi(\vartheta)^T\Omega\pi(\vartheta_0)|$$
$$+ \sup_{\vartheta\in\Theta}|\hat{\pi}_n^T\Omega\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta) - \pi(\vartheta_0)^T\Omega\pi(\vartheta)| + \sup_{\vartheta\in\Theta}|\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta)^T\Omega\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta) - \pi(\vartheta)^T\Omega\pi(\vartheta)|.$$

The four summands on the right-hand side converge in probability to zero as $n \to \infty$. For the first one, this is a consequence of Assumption (C.1). For the remaining three terms, the arguments are similar, so that we treat only the second one exemplary. We have

$$\begin{aligned}
&\sup_{\vartheta\in\Theta}|\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta)^T\Omega\hat{\pi}_n - \pi(\vartheta)^T\Omega\pi(\vartheta_0)| \\
&\quad\leq \sup_{\vartheta\in\Theta}|\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta)^T\Omega\hat{\pi}_n - \pi(\vartheta)^T\Omega\hat{\pi}_n| + \sup_{\vartheta\in\Theta}|\pi(\vartheta)^T\Omega\hat{\pi}_n - \pi(\vartheta)^T\Omega\pi(\vartheta_0)| \\
&\quad\leq \|\Omega\|\sup_{\vartheta\in\Theta}\|\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta) - \pi(\vartheta)\|\|\hat{\pi}_n\| + \|\Omega\|\sup_{\vartheta\in\Theta}\|\pi(\vartheta)\|\|\hat{\pi}_n - \pi(\vartheta_0)\| \xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

Here, we used the fact that $\sup_{\vartheta\in\Theta}\|\pi(\vartheta)\|$ is finite due to the continuity of the map π and the compactness of Θ as well as both Assumptions (C.1) and (C.2). Therefore, the function $\mathcal{L}_{\mathrm{Ind}}(\vartheta, \mathcal{Y}^n)$ converges uniformly in ϑ in probability to the limiting function $\mathcal{Q}_{\mathrm{Ind}}(\vartheta)$. Per construction, $\hat{\vartheta}_n^{\mathrm{Ind}}$ minimizes $\mathcal{L}_{\mathrm{Ind}}(\vartheta, \mathcal{Y}^n)$ and $\mathcal{Q}_{\mathrm{Ind}}(\vartheta)$ has a unique minimum at $\vartheta = \vartheta_0$. Therefore, weak consistency of $\hat{\vartheta}_n^{\mathrm{Ind}}$ follows by arguing as in the proof of Schlemm and Stelzer (2012, Theorem 2.4); although in their proof convergence in probability is replaced by almost sure convergence, this doesn't matter because we can use the subsequence criterion which says that a sequence converges in probability iff any subsequence has a further subsequence which converges almost surely.

The proof of strong consistency goes similarly by replacing convergence in probability by almost sure convergence.

(b) For the asymptotic normality, note that

$$\sqrt{n}(\hat{\pi}_n - \hat{\pi}_{sn}^{\mathrm{S}}(\vartheta_0)) = \sqrt{n}(\hat{\pi}_n - \pi(\vartheta_0)) + \sqrt{n}(\pi(\vartheta_0) - \hat{\pi}_{sn}^{\mathrm{S}}(\vartheta_0)).$$

Since both estimators are independent from each other, we obtain with Assumptions (C.3) and (C.4) that

$$\sqrt{n}(\hat{\pi}_n - \hat{\pi}_{sn}^{\mathrm{S}}(\vartheta_0)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Xi_{\mathrm{D}}(\vartheta_0) + \frac{1}{s}\Xi_S(\vartheta_0)\right). \tag{A2}$$

Moreover,

$$0_{N(\Theta)} = \nabla_\vartheta\mathcal{L}_{\mathrm{Ind}}(\vartheta, \mathcal{Y}^n)\big|_{\vartheta=\hat{\vartheta}_n^{\mathrm{Ind}}} = 2[\nabla_\vartheta\hat{\pi}_{sn}^{\mathrm{S}}(\hat{\vartheta}_n^{\mathrm{Ind}})]^T\Omega[\hat{\pi}_{sn}^{\mathrm{S}}(\hat{\vartheta}_n^{\mathrm{Ind}}) - \hat{\pi}_n].$$

We now use a Taylor expansion of order 1 around the true value ϑ_0 to obtain

$$\begin{aligned}
0_{N(\Theta)} &= \sqrt{n}\nabla_\vartheta\mathcal{L}_{\mathrm{Ind}}(\hat{\vartheta}_{\mathrm{Ind}}^n, \mathcal{Y}^n) \\
&= \sqrt{n}\nabla_\vartheta\mathcal{L}_{\mathrm{Ind}}(\vartheta_0, \mathcal{Y}^n) + \sqrt{n}\nabla_\vartheta^2\mathcal{L}_{\mathrm{Ind}}(\overline{\vartheta}_n, \mathcal{Y}^n)(\hat{\vartheta}_n^{\mathrm{Ind}} - \vartheta_0) \\
&= 2[\nabla_\vartheta\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta_0)]^T\Omega\sqrt{n}[\hat{\pi}_{sn}^{\mathrm{S}}(\vartheta_0) - \hat{\pi}_n] + 2\left[[(\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n) - \hat{\pi}_n)^T\Omega] \otimes I_{N(\Theta)}\right]\left[\nabla_\vartheta^2\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n)\right]\sqrt{n}(\hat{\vartheta}_n^{\mathrm{Ind}} - \vartheta_0) \\
&\quad + 2[\nabla_\vartheta\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n)]^T\Omega[\nabla_\vartheta\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n)]\sqrt{n}(\hat{\vartheta}_n^{\mathrm{Ind}} - \vartheta_0).
\end{aligned}$$

Here, $\overline{\vartheta}_n$ is such that $\|\overline{\vartheta}_n - \vartheta_0\| \leq \|\hat{\vartheta}_n^{\mathrm{Ind}} - \vartheta_0\|$ and hence, $\overline{\vartheta}_n \xrightarrow{\mathbb{P}} \vartheta_0$ as $n \to \infty$. Moreover,

$$[\nabla_\vartheta^2\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}^n)]^T\Omega[\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n) - \hat{\pi}_n] + [\nabla_\vartheta\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n)]^T\Omega[\nabla_\vartheta\hat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n)] \xrightarrow{\mathbb{P}} [\nabla_\vartheta\pi(\vartheta_0)]^T\Omega[\nabla_\vartheta\pi(\vartheta_0)] \tag{A3}$$

due to Assumptions (C.1), (C.2) and (C.5) and the continuity of $\pi(\vartheta)$. Furthermore, the right-hand side is non-singular since $\nabla_\vartheta \pi(\vartheta_0)$ has full column rank and Ω is non-singular. Finally, we write

$$\sqrt{n}(\widehat{\vartheta}_n^{\mathrm{Ind}} - \vartheta_0) = \left([\nabla_\vartheta^2 \widehat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}^n)]^T \Omega [\widehat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n) - \widehat{\pi}_n] + [\nabla_\vartheta \widehat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n)]^T \Omega [\nabla_\vartheta \widehat{\pi}_{sn}^{\mathrm{S}}(\overline{\vartheta}_n)] \right)^{-1} \cdot [\nabla_\vartheta \widehat{\pi}_{sn}^{\mathrm{S}}(\vartheta_0)]^T \Omega \sqrt{n} (\widehat{\pi}_{sn}^{\mathrm{S}}(\vartheta_0) - \widehat{\pi}_n)$$

and use (A2), (A3), and Assumption (C.5) to obtain as $n \to \infty$,

$$\sqrt{n}(\widehat{\vartheta}_n^{\mathrm{Ind}} - \vartheta_0) \xrightarrow{D} \left([\nabla_\vartheta \pi(\vartheta_0)]^T \Omega [\nabla_\vartheta \pi(\vartheta_0)] \right)^{-1} [\nabla_\vartheta \pi(\vartheta_0)]^T \Omega \cdot \mathcal{N}\left(0, \Xi_{\mathrm{D}}(\vartheta_0) + \frac{1}{s} \Xi_S(\vartheta_0)\right).$$

This completes the proof. □

A.3. Proofs of Section 4

Proof of Proposition 4.6. Using similar arguments as in Maronna and Yohai (1981, Lemma 2.1) ($\lim_{x\to 0} \chi(x) < 0$, $\lim_{x\to\infty} \chi(x) = \infty$, the continuity and boundedness of χ and the intermediate value theorem) we can show that for each fixed $(\pi_1, \ldots, \pi_r) \in \mathbb{R}^r$ there exists a unique solution σ of the equation

$$\mathbb{E}\left[\chi\left(\left(\frac{Y_{(r+1)h}(\vartheta) - \pi_1 Y_{rh}(\vartheta) - \cdots - \pi_r Y_h(\vartheta)}{\sigma} \right)^2 \right) \right] = 0.$$

By assumption (13), the function χ is chosen in such a way that for $(\pi_1(\vartheta), \ldots, \pi_r(\vartheta))$ this unique solution is $\sigma(\vartheta)$. Therefore, we have that $\pi(\vartheta)$ is a solution of (10b). Next, we show that $\pi(\vartheta)$ is a solution of (10a) as well. Since the function $\phi(y, u)$ is odd in u by Assumption (E.1), it holds that

$$\begin{aligned} \mathbb{E}\left[\phi\left(\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}, \frac{U_{r+1}(\vartheta)}{\sigma(\vartheta)} \right) \begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix} \right] &= \mathbb{E}\left[-\phi\left(\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}, -\frac{U_{r+1}(\vartheta)}{\sigma(\vartheta)} \right) \begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix} \right] \\ &= -\mathbb{E}\left[\phi\left(\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}, \frac{U_{r+1}(\vartheta)}{\sigma(\vartheta)} \right) \begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix} \right], \end{aligned} \tag{A4}$$

where the last equality follows from (12). From this equation we can conclude that

$$\mathbb{E}\left[\phi\left(\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}, \frac{U_{r+1}(\vartheta)}{\sigma(\vartheta)} \right) \begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix} \right] = 0,$$

and therefore, $\pi(\vartheta)$ is a solution of Equation (10a).

Next, we show similarly to Maronna and Yohai (1981, Theorem 2.2(a)) for regression models that $\pi(\vartheta)$ is the unique solution. Assume that another solution $\pi' = (\pi_1', \ldots, \pi_r', \sigma')$ of (10) exists. But then $(\pi_1', \ldots, \pi_r') \neq (\pi_1, \ldots, \pi_r)$. Note that the arguments in the derivation of (A4) still hold if we replace $\sigma(\vartheta)$ in the denominator of the second argument of ϕ by σ'. Thus, we obtain that

$$\mathbb{E}\left[\phi\left(\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}, \frac{U_{r+1}(\vartheta)}{\sigma'} \right) \begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix} \right] = 0,$$

DOI: 10.1111/jtsa.12526 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

and therefore

$$\mathbb{E}\left[\left[\phi\left(\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}, \frac{Y_{(r+1)h}(\vartheta) - \pi_1' Y_{rh}(\vartheta) - \cdots - \pi_r' Y_h(\vartheta)}{\sigma'}\right)\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}\right.\right.$$
$$\left.\left. -\phi\left(\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}, \frac{U_{r+1}(\vartheta)}{\sigma'}\right)\right]\begin{pmatrix} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{pmatrix}\right] = 0. \tag{A5}$$

Since $\mathbb{P}((Y_h(\vartheta), \ldots, Y_{rh}(\vartheta)) = (0, \ldots, 0)) = 0$ and $\mathbb{P}(Y_{(r+1)h}(\vartheta) - \pi_1' Y_{rh}(\vartheta) - \cdots - \pi_r' Y_h(\vartheta) = U_{r+1}(\vartheta)) = 0$ due to Assumption (D.3) for $\gamma = 0$ and $u \mapsto \phi(y, u)$ is strictly increasing on the interval $(-u_0, u_0)$, for every $y \in \mathbb{R}^r$ we have that

$$\left|\frac{Y_{(r+1)h}(\vartheta) - \pi_1' Y_{rh}(\vartheta) - \cdots - \pi_r' Y_h(\vartheta)}{\sigma'}\right| \geq u_0 \quad \mathbb{P}\text{-a.s.} \tag{A6}$$

because otherwise (A5) cannot hold. Now, π' is by assumption also a solution of (10b) and hence, we have due to (A6) and Assumption (E.6)

$$0 = \mathbb{E}\left[\chi\left(\left(\frac{Y_{(r+1)h}(\vartheta) - \pi_1' Y_{rh}(\vartheta) - \cdots - \pi_r' Y_h(\vartheta)}{\sigma'}\right)^2\right)\right] \geq \chi(u_0^2) > 0$$

which is a contradiction. □

Proof of Lemma 4.9. By the Cramer–Wold device, the statement of the lemma is equivalent to the assertion that $\frac{1}{\sqrt{n-r}} x^T \sum_{k=1}^{n-r} \Psi_k(\vartheta^\gamma)$ converges to a univariate normal distribution with mean 0 and variance $x^T \mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma) x$ for every $x \in \mathbb{R}^{r+1}$. According to Ibragimov (1962, Theorem 1.7), this holds if we can show that

$$\mathbb{E}|x^T \Psi_k(\vartheta^\gamma)|^{2+\delta} < \infty \tag{A7}$$

and that $(x^T \Psi_k(\vartheta^\gamma))_{k \in \mathbb{N}}$ is strongly mixing with mixing coefficients $\alpha_{x^T\Psi(\vartheta^\gamma)}(m)$ satisfying

$$\sum_{m=1}^{\infty} \alpha_{x^T\Psi(\vartheta^\gamma)}^{\delta/(2+\delta)}(m) < \infty \quad \text{for some } \delta > 0. \tag{A8}$$

The same theorem then also states that $x^T \mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma) x < \infty$ from which we then deduce that for $i, j \in \{1, \ldots, r+1\}$ the entry $[\mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma)]_{ij}$ is finite and therefore, $\mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma)$ is well-defined.
We start to show the existence of the $(2+\delta)$th moment of $x^T \Psi_k(\vartheta^\gamma)$ in (A7). Therefore, note that

$$\mathbb{E}|x^T \Psi_k(\vartheta^\gamma)|^{2+\delta} \leq C\|x\|^{2+\delta} \sum_{i=1}^{r+1} \mathbb{E}\|\Psi_{k,i}(\vartheta^\gamma)\|^{2+\delta} < \infty, \tag{A9}$$

where the last inequality holds since $\Psi_{k,i}(\vartheta^\gamma)$ is bounded by Assumptions (E.2) and (E.6).

Finally, the process $(Y_{mh}^\gamma(\vartheta))_{m \in \mathbb{Z}}$ is strongly mixing and the mixing coefficients satisfy the above condition (A8) for the following reason. Either we have in the case of replacement outliers that $Y_{mh}^\gamma(\vartheta) = G(V_m, Z_m, Y_{mh}(\vartheta))$ for some measurable function G and the three processes (V_m), (Z_m) and $(Y_{mh}(\vartheta))$ are independent, or in the case of additive outliers we have $Y_{mh}^\gamma(\vartheta) = G(V_m, W_m, Y_{mh}(\vartheta))$ for some measurable function G and the three processes (V_m), (W_m) and $(Y_{mh}(\vartheta))$ are independent. Hence, by Bradley (2007, Theorem 6.6(II)), Assumption (D.2) and Marquardt and Stelzer (2007, Proposition 3.34) we receive

$$\alpha_{Y^\gamma(\vartheta)}(m) \leq \alpha_V(m) + \alpha_Z(m) + \alpha_{Y(\vartheta)}(m) \leq C\rho^m$$

   DOI: 10.1111/jtsa.12526

respectively, $\alpha_{Y^\gamma(\vartheta)}(m) \leq \alpha_V(m) + \alpha_W(m) + \alpha_{Y(\vartheta)}(m) \leq C\rho^m$ for some $C > 0$ and $\rho \in (0, 1)$. Furthermore, $\Psi_k(\vartheta^\gamma)$ depends only on the finitely many values $Y^\gamma_{kh}(\vartheta), \dots, Y^\gamma_{(k+r)h}(\vartheta)$ and by Bradley (2007, Remark 1.8(b)) this ensures that $\alpha_{\Psi(\vartheta^\gamma)}(m) \leq \alpha_{Y^\gamma(\vartheta)}(m + r) \leq C\rho^m$. Thus, the strong mixing coefficients $\alpha_{x^T\Psi(\vartheta^\gamma)}(m)$ of $x^T\Psi(\vartheta^\gamma)$ satisfy the summability condition (A8) and the lemma is proven. □

Proof of Lemma 4.10. Note, first that for $i, j = 1, \dots, r$,

$$\sup_{\pi\in K}\left|\frac{\partial}{\partial\pi_i}\phi\left(\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}, \frac{Y^\gamma_{(r+1)h}(\vartheta) - \pi_1 Y^\gamma_{rh}(\vartheta) - \cdots - \pi_r Y^\gamma_h(\vartheta)}{\sigma}\right)Y^\gamma_{jh}(\vartheta)\right|$$

$$= \sup_{\pi\in K}\left\|\left(\frac{\partial}{\partial u}\phi\right)\left(\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}, \frac{Y^\gamma_{(r+1)h}(\vartheta) - \pi_1 Y^\gamma_{rh}(\vartheta) - \cdots - \pi_r Y^\gamma_h(\vartheta)}{\sigma}\right)Y^\gamma_{jh}(\vartheta)\right\|\left|\frac{Y^\gamma_{(r+1-i)h}(\vartheta)}{\sigma}\right|$$

$$\leq \sup_{u\in\mathbb{R}} C\left\|\left(\frac{\partial}{\partial u}\phi\right)\left(\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}, u\right)\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}\right\|\left\|\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}\right\| \leq C\left\|\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}\right\|$$

due to Assumption (E.4) and the boundedness of $1/\sigma$ on the compact set K. By Assumption (D.1) and (A.3) the expectation on the right-hand side is finite. Similarly,

$$\sup_{\pi\in K}\left\|\frac{\partial}{\partial\sigma}\phi\left(\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}, \frac{Y^\gamma_{(r+1)h}(\vartheta) - \pi_1 Y^\gamma_{rh}(\vartheta) - \cdots - \pi_r Y^\gamma_h(\vartheta)}{\sigma}\right)\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}\right\|$$

$$\leq C\sup_{u\in\mathbb{R}}\left\|u\left(\frac{\partial}{\partial u}\phi\right)\left(\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}, u\right)\begin{pmatrix}Y^\gamma_h(\vartheta)\\ \vdots\\ Y^\gamma_{rh}(\vartheta)\end{pmatrix}\right\|.$$

The expectation on the right-hand side is finite due to Assumption (E.5). Similar arguments, using Assumption (E.6), also show that $\left|\frac{\partial}{\partial\pi_i}\chi\left(\left(\frac{Y^\gamma_{(r+1)h}(\vartheta)-\pi_1 Y^\gamma_{rh}(\vartheta)-\cdots-\pi_r Y^\gamma_h(\vartheta)}{\sigma}\right)^2\right)\right|$ for $i = 1, \dots, r$ and $\left|\frac{\partial}{\partial\sigma}\chi\left(\left(\frac{Y^\gamma_{(r+1)h}(\vartheta)-\pi_1 Y^\gamma_{rh}(\vartheta)-\cdots-\pi_r Y^\gamma_h(\vartheta)}{\sigma}\right)^2\right)\right|$ are uniformly dominated by integrable random variables. Therefore, by Billingsley (1999, Theorem 16.8(ii)) (that is an application of dominated convergence) $\nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ exists on K and the order of differentiation and expectation can be changed.

Moreover, due Assumptions (E.4), (E.6) and Billingsley (1999, Theorem 16.8(i)) the map $\pi \mapsto \nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi, \vartheta^\gamma)$ is continuous.
Hence, if $\pi_n \xrightarrow{\mathbb{P}} \pi^{\mathrm{GM}}(\vartheta^\gamma) \in K$ then $\nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi_n, \vartheta^\gamma) \xrightarrow{\mathbb{P}} \nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\pi^{\mathrm{GM}}(\vartheta^\gamma), \vartheta^\gamma)$. □

Proof of Lemma 4.11. We use the decomposition

$$\sqrt{n-p}\,\mathcal{Q}_{\mathrm{GM}}(\hat{\pi}^{\mathrm{GM}}_n(\vartheta^\gamma), \vartheta^\gamma) = \frac{1}{\sqrt{n-r}}\sum_{k=1}^{n-r}\left[\mathcal{Q}_{\mathrm{GM}}(\hat{\pi}^{\mathrm{GM}}_n(\vartheta^\gamma), \vartheta^\gamma) + \Psi_k(\vartheta^\gamma)\right] - \frac{1}{\sqrt{n-r}}\sum_{k=1}^{n-r}\Psi_k(\vartheta^\gamma).$$

The first term is of order $o_P(1)$ due to Bustos (1982, Lemma 3.5) (cf. Kimmig (2016, Lemma A.5) in our setting). The second term converges to $\mathcal{N}(0, \mathcal{I}_{\mathrm{GM}}(\vartheta^\gamma))$ due to Lemma 4.9. Hence, we receive the statement. □

Proof of Theorem 4.12. Due to (10) we have $\mathcal{Q}_{\mathrm{GM}}(\pi(\vartheta^\gamma),\vartheta^\gamma) = 0$. Next, a first-order Taylor expansion around $\hat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma)$ gives

$$\begin{aligned} 0 &= \sqrt{n-r}\mathcal{Q}_{\mathrm{GM}}(\pi(\vartheta^\gamma),\vartheta^\gamma) \\ &= \sqrt{n-r}\mathcal{Q}_{\mathrm{GM}}(\hat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma),\vartheta^\gamma) + \sqrt{n-r}\nabla_\pi \mathcal{Q}_{\mathrm{GM}}(\overline{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma),\vartheta^\gamma)(\pi^{\mathrm{GM}}(\vartheta^\gamma) - \hat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma)), \end{aligned}$$

where $\|\pi^{\mathrm{GM}}(\vartheta^\gamma) - \overline{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma)\| \le \|\pi^{\mathrm{GM}}(\vartheta^\gamma) - \hat{\pi}_n^{\mathrm{GM}}(\vartheta^\gamma)\|$. The statement follows then from a combination of Lemma 4.10 and Lemma 4.11. □

A.4. Proofs of Section 5

For the ease of notation we write in the following for the Lévy process $(L_t^S)_{t\in\mathbb{R}}$ shortly $(L_t)_{t\in\mathbb{R}}$ and hence, assume that $\mathbb{E}|L_1|^{2N^*}$ for some $2N^* > \max(N(\Theta), 4+\delta)$; similarly $(Y_t^S)_{t\in\mathbb{R}}$ is $(Y_t)_{t\in\mathbb{R}}$.

Lemma A.1. Define for any $\vartheta \in \Theta$ the function $f_\vartheta(u) = c_\vartheta^T \mathrm{e}^{A_\vartheta u} e_p 1_{[0,\infty)}(u)$ and

$$G_\vartheta(u) = \left(f_\vartheta(u), \nabla_\vartheta f_\vartheta(u), \nabla_\vartheta^2 f_\vartheta(u)\right).$$

Then $\mathbb{P}$-a.s. we have

$$\left(Y_{mh}(\vartheta), \nabla_\vartheta Y_{mh}(\vartheta), \nabla_\vartheta^2 Y_{mh}(\vartheta)\right)_{m\in\mathbb{Z}} = \left(\int_{-\infty}^{mh} G_\vartheta(mh-u)\, dL_u\right)_{m\in\mathbb{Z}}$$

which is strongly mixing and ergodic.

The proof is moved to Section C in the Supporting information.

Proof of Proposition 5.4. (a) First, we prove the pointwise convergence of the sample autocovariance function and second, that $\hat{\gamma}_{\vartheta,n}(l,j)$ is locally Hölder-continuous which results in a stochastic equicontinuity condition. Then we are able to apply Pollard (1990, Theorem 10.2) which gives the uniform convergence.

Step 1. Pointwise convergence. From Lemma A.1 we already know that $(Y_{mh}(\vartheta))_{m\in\mathbb{Z}}$ is a stationary and ergodic sequence with $\mathbb{E}|Y_{mh}(\vartheta)|^2 < \infty$ due to $\mathbb{E}|L_1|^2 < \infty$. Then Birkoff's Ergodic Theorem gives as $n \to \infty$,

$$\hat{\gamma}_{\vartheta,n}(l,j) \xrightarrow{\mathbb{P}} \gamma_\vartheta(l-j).$$

Step 2. $\hat{\gamma}_{\vartheta,n}(l,j)$ is locally Hölder-continuous. Let $\gamma \in [0, 1 - N(\Theta)/(2N^*))$ and

$$U_k := \sup_{\substack{0<\|\vartheta_1-\vartheta_2\|<1 \\ \vartheta_1,\vartheta_2\in\Theta}} \frac{|Y_{kh}(\vartheta_1) - Y_{kh}(\vartheta_2)|}{\|\vartheta_1-\vartheta_2\|^\gamma}.$$

Since $((Y_{mh}(\vartheta))_{\vartheta\in\Theta})_{m\in\mathbb{Z}}$ is a stationary sequence, $U_k \stackrel{d}{=} U_1$ and due to Lemma C.3 in the Supporting information, $\mathbb{E}U_1^{2N^*} < \infty$. In particular, for any $\vartheta_1, \vartheta_2 \in \Theta$ with $\|\vartheta_1 - \vartheta_2\| < 1$ the upper bound

$$|Y_{kh}(\vartheta_1) - Y_{kh}(\vartheta_2)| \le U_k \|\vartheta_1 - \vartheta_2\|^\gamma$$

and hence,

$$\begin{aligned}&|Y_{(k+l)h}(\vartheta_1)Y_{(k+j)h}(\vartheta_1) - Y_{(k+l)h}(\vartheta_2)Y_{(k+j)h}(\vartheta_2)|\\ &\leq \underbrace{\left(\sup_{\vartheta\in\Theta}|Y_{(k+l)h}(\vartheta)| + \sup_{\vartheta\in\Theta}|Y_{(k+j)h}(\vartheta)|\right)(U_{k+l}+U_{k+j})}_{=:U^*_{k+l,k+j}}\|\vartheta_1-\vartheta_2\|^{\gamma}\end{aligned}$$

hold. Finally,

$$|\hat{\gamma}_{\vartheta_1,n}(l,j) - \hat{\gamma}_{\vartheta_2,n}(l,j)| \leq \frac{1}{n-r}\sum_{k=1}^{n-r} U^*_{k+l,k+j}\|\vartheta_1-\vartheta_2\|^{\gamma} \quad \text{for } \|\vartheta_1-\vartheta_2\| < 1 \tag{A10}$$

with

$$\mathbb{E}(U^*_{k+l,k+j}) = \mathbb{E}(U^*_{1+l,1+j}) \leq 4\left(\mathbb{E}\left(\sup_{\vartheta\in\Theta} Y_h(\vartheta)^2\right)\mathbb{E}U_1^2\right)^{1/2} < \infty$$

where we used Lemma C.3 in the Supporting information to get the finite expectation.

Step 3. Let $\epsilon, \eta > 0$. Take $0 < \delta < \min\{1, \eta\epsilon/\mathbb{E}(U^*_{1+l,1+j})\}^{1/\gamma}$. Then (A10) and Markov's inequality give

$$\mathbb{P}\left(\sup_{\substack{0<\|\vartheta_1-\vartheta_2\|<\delta\\ \vartheta_1,\vartheta_2\in\Theta}} |\hat{\gamma}_{\vartheta_1,n}(l,j) - \hat{\gamma}_{\vartheta_2,n}(l,j)| > \eta\right) \leq \mathbb{E}(U^*_{1+l,1+j})\frac{\delta^{\gamma}}{\eta} < \epsilon.$$

A conclusion of this stochastic equicontinuity condition, the pointwise convergence in Step 1 and Pollard (1990, Theorem 10.2) is the uniform convergence.

The proof of (b,c) goes in the same vein as the proof of (a). □

Proof of Proposition 5.5. Define

$$\hat{\gamma}_n^{(r-1)}(\vartheta) = \begin{pmatrix}\hat{\gamma}_{\vartheta,n}(r,r-1)\\ \vdots\\ \hat{\gamma}_{\vartheta,n}(r,0)\end{pmatrix} \quad \text{and} \quad \hat{\Gamma}_n^{(r-1)}(\vartheta) = \begin{pmatrix}\hat{\gamma}_{\vartheta,n}(r-1,r-1) & \cdots & \hat{\gamma}_{\vartheta,n}(0,r-1)\\ \vdots & & \vdots\\ \hat{\gamma}_{\vartheta,n}(r-1,0) & \cdots & \hat{\gamma}_{\vartheta,n}(0,0)\end{pmatrix}.$$

Then

$$\begin{aligned}\hat{\pi}_n^*(\vartheta) :&= \begin{pmatrix}\hat{\pi}^{\mathrm{LS}}_{n,1}(\vartheta)\\ \vdots\\ \hat{\pi}^{\mathrm{LS}}_{n,r}(\vartheta)\end{pmatrix} = [\hat{\Gamma}_n^{(r-1)}(\vartheta)]^{-1}\hat{\gamma}_n^{(r-1)}(\vartheta),\\ \sigma^2_{\mathrm{LS},n}(\vartheta) &= \hat{\gamma}_{\vartheta,n}(r,r) - [\hat{\pi}_n^*(\vartheta)]^T\hat{\gamma}_n^{(r-1)}(\vartheta).\end{aligned} \tag{A11}$$

A conclusion of Proposition 5.4(a) and the definition of $\Gamma^{(r-1)}(\vartheta)$ and $\gamma^{(r-1)}(\vartheta)$ in (A1) is that

$$\sup_{\vartheta\in\Theta}\|\hat{\Gamma}_n^{(r-1)}(\vartheta) - \Gamma^{(r-1)}(\vartheta)\| \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \sup_{\vartheta\in\Theta}\|\hat{\gamma}_n^{(r-1)}(\vartheta) - \gamma^{(r-1)}(\vartheta)\| \xrightarrow{\mathbb{P}} 0. \tag{A12}$$

Due to the continuity and the positive definiteness of $\Gamma^{(r-1)}(\vartheta)$ (cf. proof of Proposition 2.2), and the compactness of Θ we receive $\sup_{\vartheta\in\Theta}\|\Gamma^{(r-1)}(\vartheta)^{-1}\| < \infty$. Hence, statement (a) is a consequence of (A11)–(A12) and (A1).

(b) Note that

$$\frac{\partial}{\partial\vartheta_i}\hat{\pi}_n^*(\vartheta) = -[\hat{\Gamma}_n^{(r-1)}(\vartheta)]^{-1}\left[\frac{\partial}{\partial\vartheta_i}\hat{\Gamma}_n^{(r-1)}(\vartheta)\right][\hat{\Gamma}_n^{(r-1)}(\vartheta)]^{-1}\hat{\gamma}_n^{(r-1)}(\vartheta) + [\hat{\Gamma}_n^{(r-1)}(\vartheta)]^{-1}\left[\frac{\partial}{\partial\vartheta_i}\hat{\gamma}_n^{(r-1)}(\vartheta)\right],$$
$$\frac{\partial}{\partial\vartheta_i}\pi^*(\vartheta) = -[\Gamma^{(r-1)}(\vartheta)]^{-1}\left[\frac{\partial}{\partial\vartheta_i}\Gamma^{(r-1)}(\vartheta)\right][\Gamma^{(r-1)}(\vartheta)]^{-1}\gamma^{(r-1)}(\vartheta) + [\Gamma^{(r-1)}(\vartheta)]^{-1}\left[\frac{\partial}{\partial\vartheta_i}\gamma^{(r-1)}(\vartheta)\right]. \tag{A13}$$

Applying Proposition 5.4(b) we receive that

$$\sup_{\vartheta\in\Theta}\left\|\frac{\partial}{\partial\vartheta_i}\hat{\Gamma}_n^{(r-1)}(\vartheta) - \frac{\partial}{\partial\vartheta_i}\Gamma^{(r-1)}(\vartheta)\right\| \xrightarrow{\mathbb{P}} 0 \text{ and } \sup_{\vartheta\in\Theta}\left\|\frac{\partial}{\partial\vartheta_i}\hat{\gamma}_n^{(r-1)}(\vartheta) - \frac{\partial}{\partial\vartheta_i}\gamma^{(r-1)}(\vartheta)\right\| \xrightarrow{\mathbb{P}} 0. \tag{A14}$$

Then the same arguments as in (a) and (A12)-(A14) lead to statement (b).
(c) The proof goes in analog lines as in (a) and (b). □

Proof of Corollary 5.6. (a) We use the upper bound

$$\|\hat{\pi}_n^{\text{LS}}(\overline{\vartheta}_n) - \pi(\vartheta_0)\| \le \sup_{\vartheta\in\Theta}\|\hat{\pi}_n^{\text{LS}}(\vartheta) - \pi(\vartheta)\| + \|\pi(\overline{\vartheta}_n) - \pi(\vartheta_0)\|.$$

The first term converges in probability to 0 due Proposition 5.5(a) and the second term because $\pi(\vartheta)$ is continuous (see Lemma 2.5) and $\overline{\vartheta}_n \xrightarrow{\mathbb{P}} \vartheta_0$. The proof of (b,c) goes on the same way. □

Proof of Proposition 5.7. Due to Proposition 5.5(a) we already know that the LS-estimator $\hat{\pi}_n^{\text{LS}}(\vartheta)$ is consistent. The asymptotic normality of $\hat{\pi}_n^{\text{LS}}(\vartheta)$ follows in principle from Theorem 4.12 by interpreting the least squares estimator as a particular GM-estimator with $\phi(y,u) = u$ and $\chi(x) = x - 1$. An assumption of Theorem 4.12 is that the Jacobian $\mathcal{J}_{\text{GM}}(\vartheta) = \nabla_\pi \mathcal{Q}_{\text{GM}}(\pi(\vartheta),\vartheta)$ is non-singular. For the LS-estimator this can be verified by direct calculations because

$$\nabla_\pi \mathcal{Q}_{\text{LS}}(\pi,\vartheta) = \mathcal{J}_{\text{LS}}(\vartheta) = -\frac{1}{\sigma(\vartheta)}\begin{pmatrix} \gamma_\vartheta(0) & \gamma_\vartheta(h) & \dots & \gamma_\vartheta((r-1)h) & 0 \\ \gamma_\vartheta(h) & \gamma_\vartheta(0) & \dots & \gamma_\vartheta((r-2)h) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_\vartheta((r-1)h) & \dots & \dots & \gamma_\vartheta(0) & 0 \\ 0 & 0 & \dots & 0 & 2 \end{pmatrix}.$$

Hence, $\mathcal{J}_{\text{LS}}(\vartheta)$ is non-singular if and only if the upper left $r\times r$ block is non-singular. However, the upper left block is up to a positive factor the covariance matrix of the random vector $(Y_h(\vartheta),\dots,Y_{rh}(\vartheta))$ which is non-singular (cf. proof of Proposition 2.2).

Still, we need to be careful because the function ϕ and χ do not satisfy Assumptions (E.2), (E.4), and (E.6) with respect to boundedness. However, a close inspection of the proof of Theorem 4.12 reveals that the boundedness is only used at two points. First, in Lemma 4.9 where we deduce the finiteness of the expectation in (A9). However, for the LS-estimator

$$\Psi_{k,i}(\vartheta) = \left[Y_{(k+r)h}(\vartheta) - \pi_1(\vartheta_0)Y_{(k+r-1)h}(\vartheta) - \cdots - \pi_r(\vartheta_0)Y_{kh}(\vartheta)\right]Y_{(k+i-1)h}(\vartheta)$$

for $i = 1,\dots,r$ and

$$\Psi_{k,r+1}(\vartheta) = \left(\frac{Y_{(k+r)h}(\vartheta) - \pi_1(\vartheta)Y_{(k+r-1)h}(\vartheta) - \cdots - \pi_r(\vartheta)Y_{kh}(\vartheta)}{\sigma(\vartheta)}\right)^2 - 1.$$

Therefore, inequality (A9) follows since the Lévy process $(L_t)_{\in\mathbb{R}}$ has finite $(4+\delta)$th moment which then transfers to $(Y_t(\vartheta))_{t\in\mathbb{R}}$ by Marquardt and Stelzer (2007, Proposition 3.30) and subsequently the $(2+\delta/2)$-moment of $\Psi_{k,i}(\vartheta)$.

Second, the boundedness assumptions are used in the proof of Lemma 4.10 to deduce the existence of $\nabla_\pi \mathcal{Q}_{\mathrm{LS}}(\pi, \vartheta)$ and its continuity. But by the above calculations $\nabla_\pi \mathcal{Q}_{\mathrm{LS}}(\pi, \vartheta)$ exists obviously and is continuous. □

Proof of Theorem 5.3. Assumption (C.1) and (C.4) follow from Theorem 4.8 and Theorem 4.12. Assumption (C.2) is proven in Proposition 5.5. Assumption (C.3) is a consequence of Proposition 5.7. Finally, Assumption (C.5) is derived in Corollary 5.6. □

DOI: 10.1111/jtsa.12526 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd