Robust estimation of continuous-time ARMA models via indirect inference

Vicky Fasen-Hartmann †‡  Sebastian Kimmig §

In this paper we present a robust estimator for the parameters of a continuous-time ARMA\((p,q)\) (CARMA\((p,q)\)) process sampled equidistantly which is not necessarily Gaussian. Therefore, an indirect estimation procedure is used. It is an indirect estimation because we first estimate 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 parameter which takes the minimum distance between the estimated AR parameters and the simulated AR parameters gives 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. Next, we prove that the indirect estimator is consistent and asymptotically normally distributed as well using in the simulation part the asymptotically normally distributed LS-estimator. The indirect estimator satisfies several important robustness properties such as weak resistance, \(\pi_{d_{n}}\)-robustness and it has a bounded influence functional. The practical applicability of our method is demonstrated through a simulation study with replacement outliers and compared to the non-robust quasi-maximum-likelihood estimation method.

AMS Subject Classification 2010: Primary: 62F10, 62F12, 62M10
Secondary: 60G10, 60G51

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

1 Introduction

The paper presents a robust estimator for the parameters of a discretely observed continuous-time ARMA (CARMA) process. A weak ARMA\((p,q)\) process in discrete-time is a weakly stationary

---

1 Institute of Stochastics, Englerstraße 2, D-76131 Karlsruhe, Germany.
2 Financial support by the Deutsche Forschungsgemeinschaft through the research grant FA 809/2-2 is gratefully acknowledged.
3 Württembergische Versicherung AG, Gutenbergstraße 30 D-70176 Stuttgart, Germany.
solution of the stochastic difference equation
\[ \phi(B)X_m = \theta(B)Z_m, \quad m \in \mathbb{Z}, \]  
(1.1)
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 polynomial, respectively, with \( \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \in \mathbb{R}, \phi_q \neq 0 \) and \((Z_m)_{m \in \mathbb{Z}} \) is a weak white noise. If \((Z_m)_{m \in \mathbb{Z}} \) is even an independent and identically distributed (i.i.d.) sequence then we call \((X_m)_{m \in \mathbb{Z}} \) a strong ARMA process. A natural continuous-time analog of this difference equation with i.i.d. noise \((Z_m)_{m \in \mathbb{Z}} \) is the formal \( p \)-th order stochastic differential equation
\[ a(D)Y_t = c(D)DL_t, \quad t \in \mathbb{R}, \]  
(1.2)
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 polynomial, 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, i.e., 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. Going back to Brockwell (2001) a 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) it is well known as well that a discretely sampled CARMA process \((Y_{m h})_{m \in \mathbb{Z}} (h > 0 \text{ fixed}) \) 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).

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 estimations of 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, e.g., 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. Important for their results is that they have a strong ARMA model.
Unfortunately the results can not easily be extended to weak ARMA models which we have in our context.

There exist a few papers dealing with the estimation of the parameters of a discretely sampled CARMA process as Schlemm and Stelzer (2012); Brockwell et al. (2011); Fasen and Kimmig (2017); Fasen-Hartmann and Scholz (2017). 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 this paper we use the indirect inference method, originally proposed by Smith (1993) and extended by Gouriéroux et al. (1993) (see also the overview in Gouriéroux and Monfort (1997)) for models with intractable likelihood function. The core idea of the indirect estimation method is to avoid estimating the parameter 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). This method has been successfully used in different contexts, see, e.g., Gouriéroux and Monfort (1997); Jiang and Turnbull (2004); de Luna and Genton (2001, 2000). The latter two papers 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 \( \text{AR}(r) \) process to the ARMA model and estimate the parameters of the \( \text{AR}(r) \) process with a GM-estimator. We present a similar ansatz in our paper for the estimation of the CARMA parameters. Since the discretely sampled 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 paper is structured as follows. In Section 2 we first present our parametric family of CARMA processes and our model assumptions. Furthermore, we motivate that for any \( r \geq 2p - 1 \) any CARMA process has an \( \text{AR}(r) \) 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. Since the auxiliary \( \text{AR}(r) \) 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 asymptotically 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 CARMA process is consistent and asymptotically normally distributed using in the estimation part a GM-estimator and in the simulation part a LS-estimator. Robustness properties of this estimator are topic of Section 6. In particular, qualitative robustness properties and the influence functional are studied. After all, the simulation part, in Section 7, shows the practical applicability of our indirect estimator and its robustness properties. We compare our estimator with the non-robust QMLE. The paper ends with a conclusion in Section 8.

The present paper is mainly based on the PhD-thesis Kimmig (2016). The same arguments for the indirect estimation as in Kimmig (2016) were later applied in do Rêgo Sousa et al. (2017) for the estimation of the COGARCH process using only the LS- and the Yule-Walker-estimator and without investigating any robustness properties.
Notation
We use as norms the Euclidean norm \( \| \cdot \| \) in \( \mathbb{R}^d \) and its operator \( \| \cdot \| \) in \( \mathbb{R}^{m \times d} \) which is submultiplicative. 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 \( \vartheta \in \mathbb{R}^d \) the gradient with respect to the parameter vector \( \vartheta \) is \( \nabla f(\vartheta) = \frac{\partial \text{vec}(f(\vartheta))}{\partial \vartheta} \in \mathbb{R}^{md \times s} \) and similarly \( \nabla^2 f(\vartheta) = \frac{\partial^2 \text{vec}(f(\vartheta))}{\partial \vartheta \partial \vartheta^T} \in \mathbb{R}^{md \times s} \). Finally, we write \( \overset{\text{d}}{\rightarrow} \) for weak convergence and \( \overset{\text{p}}{\rightarrow} \) 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 paper we consider a parametric family of 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

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

\[
q(\vartheta) = \sup\{ j \in \{0, \ldots, p-1\} : c_j(\vartheta) = 0 \ \forall \ l > j \} \quad \text{with} \quad \sup \vartheta := p - 1,
\]

\[
c_\vartheta := (c_{q(\vartheta)}(\vartheta), c_{q(\vartheta)}-1(\vartheta), \ldots, c_0(\vartheta), 0, \ldots, 0)^T \in \mathbb{R}^p.
\]

Due to [Brockwell (2001)] 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

\[
dX_t(\vartheta) = A_\vartheta X_t(\vartheta) dt + e_p dL_t, \quad t \in \mathbb{R},
\]

(2.1a)

where \( e_p \) denotes the \( p \)-th unit vector in \( \mathbb{R}^p \). Then the process

\[
Y_t(\vartheta) := c_\vartheta^T X_t(\vartheta), \quad t \in \mathbb{R},
\]

(2.1b)

is said to be a CARMA process of order \((p, q(\vartheta))\). 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.

Furthermore, we have the discrete-time observations \( Y_h, \ldots, Y_{ah} \) 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 \( \vartheta_0 \). The aim is to receive from the observations \( Y_h, \ldots, Y_{ah} \) an estimator for \( \vartheta_0 \). Throughout the paper we will assume that the following Assumption A holds.

Assumption A.

(A.1) The parameter space \( \Theta \) is a compact subset of \( \mathbb{R}^{\text{N}(\Theta)} \).

(A.2) The true parameter \( \vartheta_0 \) is an element of the interior of \( \Theta \).

(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 \( \mathbb{E}|L_1|^{4+\delta} < \infty \).
The eigenvalues of $A_\vartheta$ have strictly negative real parts.

For all $\vartheta \in \Theta$ the zeros of $c_\vartheta(z) = c_0(\vartheta)z^{q(\vartheta)} + c_1(\vartheta)z^{q(\vartheta)-1} + \ldots + c_{q(\vartheta)}$ are different from the eigenvalues of $A_\vartheta$.

For any $\vartheta, \vartheta' \in \Theta$ we have $(c_\vartheta, A_\vartheta) \neq (c_{\vartheta'}, A_{\vartheta'})$.

For all $\vartheta \in \Theta$ the zeros of $c_\vartheta(z) = c_0(\vartheta)z^{q(\vartheta)} + c_1(\vartheta)z^{q(\vartheta)-1} + \ldots + c_{q(\vartheta)}$ are different from the eigenvalues of $A_\vartheta$.

For any $\vartheta, \vartheta' \in \Theta$ we have $(c_\vartheta, A_\vartheta) \neq (c_{\vartheta'}, A_{\vartheta'})$.

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 space model $(2.1a)$ and hence, of the CARMA process $(Y_t(\vartheta))_{t \in \mathbb{R}}$ (see Marquardt and Stelzer (2007)). For this reason we will assume throughout the paper that $(Y_t(\vartheta))_{t \in \mathbb{R}}$ is stationary.

(iii) A consequence of $(A.4)$ and $(A.8)$, the compactness of $\Theta$ and the fact that the eigenvalues of a matrix are continuous functions of its entries (cf. Bernstein (2009, Fact 10.11.2)) is that $\max_{\vartheta \in \Theta} \{ |\lambda| : \lambda \text{ is eigenvalue of } e^{A_\vartheta t} \} < 1$ and hence, $\sup_{\vartheta \in \Theta} \| e^{A_\vartheta t} \| \leq Ce^{-\rho t}$ for some $C, \rho > 0$.

(iv) Due to $(A.5)$ the state space representation $(2.1)$ 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 CARMA processes $(Y_t(\vartheta))_{t \in \mathbb{R}}$ is identifiable from the spectral density 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)).

In the following we denote the autocovariance function of $(Y_t(\vartheta))_{t \in \mathbb{R}}$ as $(\gamma_{\vartheta}(t))_{t \in \mathbb{R}}$ which has by Assumption $A$ the form

$$\gamma_{\vartheta}(t) = c_\vartheta^T e^{A_\vartheta t} \Sigma_\vartheta e_{\vartheta}^T, \quad t \in \mathbb{R},$$

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

2.2 The AR($r$) representation of a 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

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

such that

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

(2.3)
is stationary with $\mathbb{E}[U_1(\vartheta)] = 0$, Var$(U_1(\vartheta)) = \sigma^2(\vartheta)$ and

$$
\mathbb{E}[U_m(\vartheta) Y_{(m-k)}(\vartheta)] = 0 \quad \text{for } k = 1, \ldots, r.
$$

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

**Proof.** First, we need to show that for any $r \in \mathbb{N}$ the covariance matrix of $(Y_{h}(\vartheta), \ldots, Y_{(r+1)h}(\vartheta))$ is non–singular. To see this, note that the autocovariance function of $(Y_{mh}(\vartheta))_{m \in \mathbb{Z}}$ is $\gamma_0(mh) = e^{T_\vartheta} c_{\vartheta}^T e^{A_{\vartheta} hm} \Sigma_{\vartheta} c_{\vartheta}, m \in \mathbb{Z}$ (see (2.2)). Since $\Sigma_{\vartheta}$ is non–singular (cf. Schlemm and Stelzer (2012, Corollary 3.9)) and $c_{\vartheta} \neq 0$, we have that $\gamma_0(0) > 0$. Moreover, the eigenvalues of $A_{\vartheta}$ have strictly negative real parts by (A.4) and therefore, $\gamma_0(mh) \to 0$ as $m \to \infty$ holds. By Brockwell and Davis (1991, Proposition 5.1.1), it follows that the covariance matrix of $(Y_{h}(\vartheta), \ldots, Y_{(r+1)h}(\vartheta))$ is non–singular for every $r \in \mathbb{N}$. Thus, a conclusion of Brockwell and Davis (1991, §8.1) is that there exist unique $\pi_1(\vartheta), \ldots, \pi_r(\vartheta), \sigma^2(\vartheta)$ which solve the set of $r + 1$ Yule–Walker equations, namely

$$
\pi^*(\vartheta) := \begin{pmatrix}
\pi_1(\vartheta) \\
\vdots \\
\pi_r(\vartheta)
\end{pmatrix} = \begin{pmatrix}
\gamma_0(0) & \gamma_0(h) & \cdots & \gamma_0((r-1)h) \\
\gamma_0(h) & \gamma_0(0) & \cdots & \gamma_0((r-2)h) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_0((r-1)h) & \gamma_0((r-2)h) & \cdots & \gamma_0(0)
\end{pmatrix}^{-1} \begin{pmatrix}
\gamma_0(h) \\
\vdots \\
\gamma_0(rh)
\end{pmatrix},
$$

$$
\sigma^2(\vartheta) = \gamma_0(0) - \pi^*(\vartheta)^T \gamma^r(\vartheta)^{-1}(\vartheta).
$$

**Remark 2.3.** $U_m(\vartheta)$ can be interpreted as the error of the best linear predictor of $Y_0(mh)$ in terms of $Y_{(m-1)h}(\vartheta), \ldots, Y_{(m-r)h}(\vartheta)$. Per construction, however, the sequence $(U_m(\vartheta))_{m \in \mathbb{Z}}$ is not an uncorrelated sequence, $U_m(\vartheta)$ is only uncorrelated with $Y_{(m-1)h}(\vartheta), \ldots, Y_{(m-r)h}(\vartheta)$.

**Definition 2.4.** Let $\Pi \subseteq \mathbb{R}^{r+1}$ 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 Theorem 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.

**Proof.** We make use of the fact that the discretely observed 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

$$
\phi(B)Y_{mh}(\vartheta) = \theta(B)e_m(\vartheta),
$$

where $\phi(z) = \prod_{k=1}^p (1 - e^{h \lambda_k} z)$ (the $\lambda_k$ being the eigenvalues of $A_{\vartheta}$), $\theta(z)$ is a monic, Schur–stable polynomial and $(e_m(\vartheta))_{m \in \mathbb{Z}}$ is a weak white noise (see Brockwell and Lindner (2009, Lemma 2.1)).

We can now decompose the map $\pi : \Theta \to \Pi$ into three separate maps for which we define the following spaces:

$$
\mathcal{M} := \{ (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_{p-1}, \sigma) \in \mathbb{R}^{2p} : \text{The coefficients define a weak ARMA($p,p-1$) model as in (1.1) for which $\phi(z)$ and $\theta(z)$ have no common zeros} \} \subseteq \mathbb{R}^{2p},
$$

$$
\mathcal{G} := \{ \gamma = (\gamma_0, \ldots, \gamma_r) \in \mathbb{R}^{r+1} : \text{The coefficients define the autocovariances up to order $r$ of a stationary stochastic process where $\Gamma^r = \text{non-singular} \} \subseteq \mathbb{R}^{r+1},
$$

where

$$
\gamma(\vartheta) := \begin{pmatrix}
\gamma_0(h) \\
\gamma_0(2h) \\
\vdots \\
\gamma_0((r-1)h)
\end{pmatrix} = (\pi^*(\vartheta)^T \gamma^r(\vartheta)^{-1}(\vartheta))^T
$$

and $\gamma^r(\vartheta)^{-1}(\vartheta) := \begin{pmatrix}
\gamma_0(h) & \gamma_0(2h) & \cdots & \gamma_0((r-1)h)
\end{pmatrix}$.
\[
\Pi := \{(\pi_1, \ldots, \pi_r, \sigma) \in \mathbb{R}^r \times (0, \infty) : (\pi_1, \ldots, \pi_r) \text{ are the coefficients of a stationary}
\]

\[
AR(r) \text{ process and } \sigma^2 \text{ is the variance of the noise} \subseteq \mathbb{R}^{r+1},
\]

where \( \Gamma^{(r-1)} \) is defined as \( \Gamma^{(r-1)}(\theta) \) in (2.5a). Denote by \( \pi_1 : \Theta \rightarrow \mathcal{M} \) 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 (2.6). Denote by \( \pi_2 : \mathcal{M} \rightarrow \mathcal{G} \) the map which maps the parameters of a weak ARMA\((p, p-1)\) process to its autocovariances of lags 0, \ldots, \( r \). Lastly, denote by \( \pi_3 : \mathcal{G} \rightarrow \Pi \) the map which maps a vector of autocovariances \((\gamma_0, \ldots, \gamma_r)\) to the parameters of the auxiliary AR\((r)\) model. Then we have that \( \pi = \pi_1 \circ \pi_2 \circ \pi_3 \). We will show that \( \pi_i \) is injective for \( i = 1, 2, 3 \) and receive from this the injectivity of \( \pi \). The three-times continuous-differentiability of the map \( \pi \) follows from the representation (2.5) and the three-times continuous-differentiability of the autocovariance function \( \gamma_0 \).

**Step 1:** \( \pi_1 \) is injective.

Due to Assumption A and Schlemm and Stelzer (2012, Theorem 3.13) the family of sampled processes \( \{(Y_{nm}(\theta))_{m \in \mathbb{Z}} : \theta \in \Theta\} \) is identifiable from the spectral densities and hence, for any \( \hat{\theta} \neq \hat{\theta}' \in \Theta \) the parameters of the weak ARMA process in (2.5d) differ.

**Step 2:** \( \pi_2 \) is injective if \( r \geq 2p-1 \).

The reason is that by the method of Brockwell and Davis (1991, p. 93), the autocovariances of an ARMA\((p, p-1)\) process are completely determined as the solution of a difference equation with \( p \) boundary conditions which depend on the coefficient vector \((\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_{p-1}, \sigma)\). 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 \( \pi_2 \).

**Case 1:** \((\phi_1, \ldots, \phi_p) \neq (\hat{\phi}_1, \ldots, \hat{\phi}_p)\). Define \( \Gamma^{(p)}(\theta) \in \mathbb{R}^{(p+1)\times(p+1)} \) similarly to \( \Gamma^{(r-1)}(\theta) \) in (2.5a). Due to Brockwell and Davis (1991, (3.3.9))

\[
(-\phi_p \ldots -\phi_1) \Gamma^{(p)}(\theta) = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}, \\
(-\hat{\phi}_p \ldots -\hat{\phi}_1) \Gamma^{(p)}(\hat{\theta}) = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}.
\]

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

**Case 2:** \((\phi_1, \ldots, \phi_p) = (\hat{\phi}_1, \ldots, \hat{\phi}_p)\). Assume that \( \pi_2(\theta) = \pi_2(\hat{\theta}) \). But then due to (Brockwell and Davis 1991, (3.3.9)), \((\gamma_0(k))_{k \in \mathbb{N}_0} = (\tilde{\gamma}_0(k))_{k \in \mathbb{N}_0}\) and hence, \( \theta = \hat{\theta} \).

**Step 3:** \( \pi_3 \) is injective.

We can also rewrite the linear equations (2.5) as a linear system with \( r+1 \) equations and the \( r+1 \) unknown variables \( \gamma_0, \ldots, \gamma_r \) which gives the injectivity of \( \pi_3 \).

Finally, due to Lemma 2.3 we suppose throughout the paper:

**Assumption B.** Let \( r \geq 2p-1 \).

### 3 Indirect estimation

For fixed \( r \), denote by \( \hat{\pi}_n \) an estimator of \( \pi(\theta_0) \) that is calculated from the observations \( Y^n = (Y_2, \ldots, Y_{n_0}) \). If we were able to analytically invert the link function \( \pi \) and calculate \( \pi^{-1}(\hat{\pi}_n) \), then \( \pi^{-1}(\hat{\pi}_n) \) we would be an estimator for \( \theta_0 = \pi^{-1}(\pi(\theta_0)) \). However, this is not possible in general since no analytic representation of \( \pi^{-1} \) 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 \((L_t^\nu)_{t \in \mathbb{R}}\) with \( \mathbb{E}L_1^\nu = 0 \) and \( \mathbb{E}(L_1^\nu)^2 = \sigma_L^2 \). Then, for a fixed parameter \( \vartheta \in \Theta \) we generate a sample path of the associated CARMA process \((Y_t^\vartheta(\theta))_{t \in \mathbb{R}}\) using the simulated path \((L_t^\nu)_{t \in \mathbb{R}}\). This gives us a vector of “pseudo–observations” \( Y_{sn}^\vartheta(\theta) = (Y_{h}^\vartheta(\theta), \ldots, Y_{snh}^\vartheta(\theta)) \) of length \( sn \). From this observation \( Y_{sn}^\vartheta(\theta) \) we estimate again \( \pi(\theta) \) by an estimator \( \hat{\pi}^{\vartheta}_{sn}(\theta) \). The idea is now to choose that value of \( \vartheta \) as estimator for \( \vartheta_0 \) which minimizes a suitable distance between \( \hat{\vartheta}_n \) and \( \hat{\pi}^{\vartheta}_{sn}(\theta) \). The formal definition is as follows.

**Definition 3.1.** Let \( \hat{\vartheta}_n \) be an estimator for \( \pi(\vartheta_0) \) calculated from the data \( \mathcal{Y}^n \), let \( \hat{\pi}^{\vartheta}_{sn}(\theta) \) be an estimator for \( \pi(\vartheta) \) calculated from the pseudo–observations \( \mathcal{Y}_{sn}^\vartheta(\theta) = (Y_{h}^\vartheta(\theta), \ldots, Y_{snh}^\vartheta(\theta)) \) 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

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

Then, the indirect estimator for \( \vartheta_0 \) is

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

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) \( \hat{\vartheta}_n \xrightarrow{P} \vartheta(\vartheta_0) \) as \( n \to \infty \).

(C.2) \( \sup_{\vartheta \in \Theta} \| \hat{\pi}^{\vartheta}_{sn}(\theta) - \pi(\theta) \| \xrightarrow{P} 0 \) as \( n \to \infty \).

Then

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

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) \( \sqrt{n}(\hat{\vartheta}_n - \vartheta(\vartheta_0)) \xrightarrow{D} \mathcal{N}(0, \Xi(\vartheta)) \) as \( n \to \infty \) for any \( \vartheta \in \Theta \).

(C.4) \( \sqrt{n}(\hat{\vartheta}_n - \vartheta(\vartheta_0)) \xrightarrow{P} \mathcal{N}(0, \Xi(\vartheta_0)) \) as \( n \to \infty \).

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

\[
\nabla_\vartheta \hat{\pi}^{\vartheta}_{sn}(\tilde{\vartheta}_n) \xrightarrow{P} \nabla_\vartheta \pi(\vartheta_0), \quad \nabla_\vartheta \hat{\pi}^{\vartheta}_{sn}(\tilde{\vartheta}_n) = O_P(1),
\]

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

Then, as \( n \to \infty \),

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

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

where

\[
\mathcal{J}_{\text{Ind}}(\vartheta_0) = \mathcal{J}_{\text{Ind}}(\vartheta_0) \mathcal{J}_{\text{Ind}}(\vartheta_0) \mathcal{J}_{\text{Ind}}(\vartheta_0)^{-1}
\]
with
\[
\mathcal{J}_{\text{Ind}}(\hat{\theta}_0) = [\nabla_\theta \pi(\theta_0)]^T \Omega [\nabla_\theta \pi(\theta_0)] \quad \text{and}
\]
\[
\mathcal{J}_{\text{Ind}}(\hat{\theta}_0) = [\nabla_\theta \pi(\theta_0)]^T \Omega \left[ \mathcal{Z}_D(\theta_0) + \frac{1}{s} \mathcal{Z}_S(\theta_0) \right] \Omega [\nabla_\theta \pi(\theta_0)].
\]

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 \(\mathcal{L}_{\text{Ind}}(\theta, \mathcal{Y}^n)\) (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 refereeing to Gouriéroux and Monfort (1997, Proposition 4.2).

**Proof.**
(a) We first start by proving the consistency. To this end, we define the map
\[
\mathcal{L}_{\text{Ind}} : \Theta \to [0, \infty) \quad \text{as} \quad \theta \mapsto [\pi(\theta) - \pi(\theta_0)]^T \Omega [\pi(\theta) - \pi(\theta_0)].
\] (3.1)

With this, we obtain
\[
\sup_{\theta \in \Theta} |\mathcal{L}_{\text{Ind}}(\theta, \mathcal{Y}^n) - \mathcal{L}_{\text{Ind}}(\theta)| = \sup_{\theta \in \Theta} |[\pi_\theta - \pi_{\theta_0}(\theta)]^T \Omega [\pi_\theta - \pi_{\theta_0}(\theta)] - [\pi(\theta) - \pi(\theta_0)]^T \Omega [\pi(\theta) - \pi(\theta_0)]| \\
\leq |\pi_\theta^T \Omega \pi_\theta - \pi(\theta_0)^T \Omega \pi(\theta_0)| + \sup_{\theta \in \Theta} |\pi_\theta^T \Omega \pi_\theta - \pi(\theta)^T \Omega \pi(\theta)| \quad + \sup_{\theta \in \Theta} |\pi_\theta^T \Omega \pi_\theta - \pi(\theta)^T \Omega \pi(\theta)|.
\]

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 \((C.1)\). For the remaining three terms, the arguments are similar, so that we treat only the second one exemplary. We have
\[
\sup_{\theta \in \Theta} |\pi_\theta^T \Omega \pi_\theta - \pi(\theta)^T \Omega \pi(\theta)| \\
\leq \sup_{\theta \in \Theta} |\pi_\theta^T \Omega \pi_\theta - \pi(\theta)^T \Omega \pi_\theta| + \sup_{\theta \in \Theta} |\pi(\theta)^T \Omega \pi_\theta - \pi(\theta)^T \Omega \pi(\theta)| \\
\leq \|\Omega\| \sup_{\theta \in \Theta} |\pi_\theta - \pi(\theta)||\pi_\theta|| + \|\Omega\| \sup_{\theta \in \Theta} |\pi(\theta)||\pi_\theta - \pi(\theta)|| \xrightarrow{n \to \infty} 0.
\]

Here, we used the fact that \(\sup_{\theta \in \Theta} \|\pi(\theta)\||\pi_\theta||\pi(\theta)||\pi_\theta - \pi(\theta)|| \xrightarrow{n \to \infty} 0\). This means that \(\sup_{\theta \in \Theta} \|\pi(\theta)\||\pi_\theta - \pi(\theta)|| \xrightarrow{n \to \infty} 0\).

Therefore, the function \(\mathcal{L}_{\text{Ind}}(\theta, \mathcal{Y}^n)\) converges uniformly in \(\theta\) to the limiting function \(\mathcal{L}_{\text{Ind}}(\theta)\). Per construction, \(\hat{\theta}_n\) minimizes \(\mathcal{L}_{\text{Ind}}(\theta, \mathcal{Y}^n)\) and \(\mathcal{L}_{\text{Ind}}(\theta)\) has a unique minimum at \(\theta = \theta_0\). Therefore, weak consistency of \(\hat{\theta}_n\) follows by arguing as in the proof of Schlemm and Steizler (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
\[
\sqrt{n}(\hat{\pi}_n - \pi^{S}(\vartheta_0)) = \sqrt{n}(\hat{\pi}_n - \pi(\vartheta_0)) + \sqrt{n}(\pi(\vartheta_0) - \pi^{S}(\vartheta_0)).
\]
Since both estimators are independent from each other, we obtain with \([C.3]\) and \([C.4]\) that
\[
\sqrt{n}(\hat{\pi}_n - \pi^{S}(\vartheta_0)) \xrightarrow{D} N \left( 0, \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) + \frac{1}{s} \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) \right) \right) \right).
\]
Moreover,
\[
0 = \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) \right) \big| \vartheta = \hat{\vartheta}_n = 2[\mathbb{E} \pi^{S}(\hat{\vartheta}_n) \mathcal{Y}_n(\hat{\vartheta}_n) - \pi(\vartheta_0) + 2[\mathbb{E} \pi^{S}(\hat{\vartheta}_n) \mathcal{Y}_n(\hat{\vartheta}_n) - \pi(\vartheta_0) - \pi^{S}(\vartheta_0)].
\]

We now use a Taylor expansion of order 1 around the true value \(\vartheta_0\) to obtain
\[
0 = \sqrt{n}[\mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) \right) \big| \vartheta = \hat{\vartheta}_n = 2[\mathbb{E} \pi^{S}(\hat{\vartheta}_n) \mathcal{Y}_n(\hat{\vartheta}_n) - \pi(\vartheta_0) + 2[\mathbb{E} \pi^{S}(\hat{\vartheta}_n) \mathcal{Y}_n(\hat{\vartheta}_n) - \pi(\vartheta_0) - \pi^{S}(\vartheta_0)].
\]

Here, \(\hat{\vartheta}_n\) is such that \[\parallel \hat{\vartheta}_n - \vartheta_0 \parallel \leq \parallel \hat{\vartheta}_n - \vartheta_0 \parallel\] and hence, \(\hat{\vartheta}_n \xrightarrow{P} \vartheta_0\) as \(n \to \infty\). Moreover,
\[
[\mathbb{E} \pi^{S}(\hat{\vartheta}_n) \mathcal{Y}_n(\hat{\vartheta}_n) - \pi(\vartheta_0) - \pi^{S}(\vartheta_0)]
\]
due to \([C.1]\), \([C.2]\), \([C.5]\) and the continuity of \(\pi(\vartheta)\). Furthermore, the right-hand side is non-singular since \(\mathbb{E} \pi(\vartheta_0)\) has full column rank and \(\Omega\) is non-singular. Finally, we write
\[
\sqrt{n}(\hat{\vartheta}_n^{\text{Ind}} - \vartheta_0)
\]
and use \([3.2], [3.3]\) and \([C.5]\) to obtain as \(n \to \infty\),
\[
\sqrt{n}(\hat{\vartheta}_n^{\text{Ind}} - \vartheta_0) \xrightarrow{D} \left( [\mathbb{E} \pi(\vartheta_0) + \pi(\vartheta_0)] - 1 [\mathbb{E} \pi(\vartheta_0) + \pi(\vartheta_0)] \right) \cdot \mathcal{N} \left( 0, \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) + \frac{1}{s} \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) \right) \right) \right).
\]

This completes the proof. \(\square\)

**Remark 3.3.**

(a) The asymptotic covariance matrix can be written as
\[
\mathbb{E}_n(\vartheta_0) = \mathcal{H}(\vartheta_0) \left( \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) + \frac{1}{s} \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) \right) \right) \right) \mathcal{H}(\vartheta_0)^T,
\]
where \(\mathcal{H}(\vartheta_0) = [\mathbb{E} \pi(\vartheta_0) + \pi(\vartheta_0)] - 1 [\mathbb{E} \pi(\vartheta_0) + \pi(\vartheta_0)] \cdot \mathcal{N} \left( 0, \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) + \frac{1}{s} \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) \right) \right) \right).\) 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 \(\mathbb{E}_n(\vartheta_0) \to \mathcal{H}(\vartheta_0) \mathbb{E} \left( \mathcal{Y}_n(\vartheta_0) \right) \mathcal{H}(\vartheta_0)^T\). The best efficiency is received for
\[
\Omega = [\mathbb{E} \pi(\vartheta_0)]^{-1} \text{ in which case } \mathbb{E}_n(\vartheta_0) \xrightarrow{s \to \infty} \left[ \mathbb{E} \pi(\vartheta_0) \mathbb{E} \pi(\vartheta_0) - 1 \mathbb{E} \pi(\vartheta_0) \right]^{-1}.\]
4 Estimating the auxiliary AR(r) parameters of a CARMA process with outliers

In order to apply the indirect estimator to a discretely sampled CARMA process we need strongly consistent and asymptotically normally distributed estimators for the parameters of the auxiliary AR(r) representation. In this section we will study generalized M- (GM-) estimators. The GM-estimator will be applied to a 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, e.g., 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 \( g : [0, 1] \rightarrow [0, 1] \) be a function that satisfies \( g(\gamma) - \gamma = o(\gamma) \) for \( \gamma \rightarrow 0 \). 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 \((Y^T_{mh}(\vartheta))_{m \in \mathbb{Z}}\) is defined as

\[
Y^T_{mh}(\vartheta) = (1 - V_m)Y_{mh}(\vartheta) + V_mZ_m.
\]  

(4.1)

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

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 \( Y^T_{mh}(\vartheta) = Y_{mh}(\vartheta) + V_mW_m \) such that the realization of \( W_m \) is added to the realization of \( Y_{mh}(\vartheta) \) if \( V_m = 1 \).

(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, i.e., 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 \( \varepsilon \) and setting \( V_m = \max(B_{m-l}, \ldots, B_m) \) for a fixed \( l \in \mathbb{N} \). Then as \( \varepsilon \rightarrow 0 \),

\[
\mathbb{P}(V_m = 1) = 1 - (1 - \varepsilon)^l = l\varepsilon + o(\varepsilon),
\]

which results in \( \gamma = l\varepsilon \). For \( \varepsilon \) sufficiently small, outliers then appear in a block of size \( l \).

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 \( \mathbb{E}|Z_1| < \infty \) for some \( \delta > 0 \).
Either we have the replacement model where the processes \((Y_{mb}(\theta))_{m \in \mathbb{Z}}, (V_m)_{m \in \mathbb{Z}}\) 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 strong mixing, i.e., \(\alpha_V(m) \leq Cp^m\) and \(\alpha_Z(m) \leq Cp^m\) for some \(C > 0, \rho \in (0, 1)\) and every \(m \in \mathbb{N}\). Or we have the additive model with \(Z_m = Y_{mb}(\theta) + W_m\) where the processes \((Y_{mb}(\theta))_{m \in \mathbb{Z}}, (V_m)_{m \in \mathbb{Z}}\) 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 strong mixing.

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

\[
P(a^r \gamma_{(r+1)h}(\theta) + \pi_1 Y_{rh}(\theta) + \ldots + \pi_r Y_{rh}(\theta) = 0) = 0.
\]

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 CARMA process \((Y_{mb})(m)\) is in Bustos (1982) an infinite-order moving average process whose noise is \(\Phi\)–mixing which is in general not satisfied for a sampled CARMA process. However, we already know from Marquardt and Stelzer (2007, Proposition 3.34) that a CARMA process is exponentially strong mixing which is weaker than \(\Phi\)-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 strong mixing instead of \(\Phi\)-mixing as in Bustos (1982).

In the following we define GM-estimators. Let two functions \(\phi : \mathbb{R}^r \times \mathbb{R} \rightarrow \mathbb{R}\) and \(\chi : \mathbb{R} \rightarrow \mathbb{R}\) be given. Moreover, assume that we have observations \(\gamma^r(\theta) = (Y_{h}^r(\theta), Y_{2h}^r(\theta), \ldots, Y_{nh}^r(\theta))\) from the disturbed process in (4.1). The parameter

\[
\pi^\Gamma(\theta) = (\pi_{1}^\Gamma(\theta), \ldots, \pi_{r}^\Gamma(\theta), \sigma_{1}^\Gamma(\theta), \ldots, \sigma_{r}^\Gamma(\theta))
\]

is defined as the solution of the equations

\[
E \left[ \phi \left( \frac{Y_{h}^r(\theta)}{\sigma}, \frac{Y_{(r+1)h}^r(\theta) - \pi_1 Y_{rh}^r(\theta) - \ldots - \pi_r Y_{rh}^r(\theta)}{\sigma} \right) \right] = 0, \quad (4.2a)
\]

\[
E \left[ \chi \left( \frac{Y_{(r+1)h}^r(\theta) - \pi_1 Y_{rh}^r(\theta) - \ldots - \pi_r Y_{rh}^r(\theta)}{\sigma} \right)^2 \right] = 0 \quad (4.2b)
\]

for \((\pi_1, \ldots, \pi_r, \sigma) \in \mathbb{R}^r \times (0, \infty)\). The idea is again that these are the parameters of the auxiliary AR representation of \((Y_{mb}(\theta))_{m \in \mathbb{Z}}\). Note that \(\pi^\Gamma(\theta)\) 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. Now, the GM-estimator \(\hat{\pi}_{n}^\Gamma(\theta) = (\hat{\pi}_{n,1}^\Gamma(\theta), \ldots, \hat{\pi}_{n,r}^\Gamma(\theta), \hat{\sigma}_{n}^\Gamma(\theta))\) based on \(\phi\) and \(\chi\) is defined to satisfy

\[
\frac{1}{n - r} \sum_{k=1}^{n-r} \phi \left( Y_{kh}^r(\theta), \ldots, Y_{(k+r-1)h}^r(\theta) \right) \frac{Y_{(k+r)h}^r(\theta) - \hat{\pi}_{n,1}^\Gamma(\theta) Y_{kh}^r(\theta) - \ldots - \hat{\pi}_{n,r}^\Gamma(\theta) Y_{rh}^r(\theta)}{\hat{\sigma}_{n}^\Gamma(\theta)} = 0, \quad (4.3a)
\]

\[
\frac{1}{n - r} \sum_{k=1}^{n-r} \chi \left( Y_{(k+r)h}^r(\theta) - \hat{\pi}_{n,1}^\Gamma(\theta) Y_{kh}^r(\theta) - \ldots - \hat{\pi}_{n,r}^\Gamma(\theta) Y_{rh}^r(\theta) \right)^2 = 0. \quad (4.3b)
\]

Throughout the paper we assume that there exists a solution of (4.3) although this is not always the
case in practice.

**Example 4.3.**

(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); 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 $\psi$ 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 $\psi$ is again a suitably chosen bounded function.

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

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

where $k$ is a tuning constant.

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

In order to develop an asymptotic theory and to obtain a robust estimator it is necessary to impose assumptions on $\phi$ and $\chi$ which we will do next analogous to Bustos (1982) (E1) - (E6):

**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 $\phi(y,u) \geq 0$ for $u \geq 0$.

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

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

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

(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 $(y,u) \mapsto \phi(y,u)/u$ is bounded.

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

(E.6) $\chi$ is bounded and increasing on $\{x : -a \leq \chi(x) < b\}$ where $b = \sup_{x \in \mathbb{R}} \chi(x)$ and $a = -\chi(0)$. Furthermore, $\chi$ is differentiable and $x \mapsto x\chi^2(x^2)$ is continuous and bounded. Lastly, $\chi(u_0^2) > 0.$
In the remaining of this section we always assume that Assumption D and E are satisfied.

\textbf{Remark 4.4.} As pointed out in Bustos (1982, p. 497) one can deduce from Maronna and Yohai (1981, Theorem 2.1) that there exists a solution \( \pi^\text{GM}(\theta^T) \in \mathbb{R}^r \times (0, \infty) \) of equation (4.2) if Assumption E holds. Moreover, there exists a compact set \( K \subseteq \mathbb{R}^r \times (0, \infty) \) with \( \pi^\text{GM}(\theta^T) \in K \) and for any \( \pi \in K^c \) equation (4.2) does not hold (see Bustos (1982, p. 500)).

In general it is not easy to verify that \( \pi^\text{GM}(\theta^T) \) is unique. Additionally, one would like to have that \( \pi^\text{GM}(\theta^0) = \pi(\theta) \) are the parameters of the auxiliary AR(1) model in the case that the GM-estimator is applied to realizations of an uncontaminated sampled CARMA process \( (Y_{mh}(\theta))^m \in Z \). The following proposition gives a sufficient condition.

\textbf{Proposition 4.5.} Suppose that \( U_{r+1}(\theta) \) as defined in equation (2.3) satisfies

\begin{equation}
(U_{r+1}(\theta), Y_{rh}(\theta), \ldots, Y_h(\theta)) = (-U_{r+1}(\theta), Y_{rh}(\theta), \ldots, Y_h(\theta)).
\end{equation}

Assume further that the function \( u \mapsto \phi(y,u) \) is nondecreasing and strictly increasing for \( |u| \leq u_0 \), where \( u_0 \) satisfies Assumptions (E.3) and (E.6) and the function \( \chi \) is chosen in such a way that

\begin{equation}
E \left[ \chi \left( \frac{U_i(\theta)}{\sigma(\theta)} \right)^2 \right] = 0.
\end{equation}

Finally, assume that \( \gamma = 0 \) so that \( (Y_{mh}(\theta))^m \in Z = (Y_{mh}(\theta))^m \in Z \). Then the auxiliary parameter \( \pi(\theta) \) as defined in (Theorem 2.2) is the unique solution of (4.2), i.e., \( \pi^\text{GM}(\theta^0) = \pi(\theta) \).

\textbf{Proof.} Using similar arguments as in Maronna and Yohai (1981, Lemma 2.1) \( \lim_{t \to 0} \chi(x) < 0 \), \( \lim_{t \to \infty} \chi(x) = \infty \), the continuity and boundedness of \( \chi \) 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 \( \sigma \) of the equation

\begin{equation}
E \left[ \chi \left( \frac{Y_{(r+1)h}(\theta) - \pi_1 Y_{rh}(\theta) - \ldots - \pi_ry_h(\theta)}{\sigma} \right)^2 \right] = 0.
\end{equation}

By assumption (4.5), the function \( \chi \) is chosen in such a way that for \( (\pi_1(\theta), \ldots, \pi_r(\theta)) \) this unique solution is \( \sigma(\theta) \). Therefore, we have that \( \pi(\theta) \) is a solution of (4.2b). Next, we show that \( \pi(\theta) \) is a solution of (4.2a) as well. Since the function \( \phi(y,u) \) is odd in \( u \) by Assumption (E.1) it holds that

\begin{equation}
E \left[ \phi \left( \begin{pmatrix} Y_{h}(\theta) \\ \vdots \\ Y_{rh}(\theta) \end{pmatrix}, \frac{U_{r+1}(\theta)}{\sigma(\theta)} \right) \begin{pmatrix} Y_{h}(\theta) \\ \vdots \\ Y_{rh}(\theta) \end{pmatrix} \right] = E \left[ -\phi \left( \begin{pmatrix} Y_{h}(\theta) \\ \vdots \\ Y_{rh}(\theta) \end{pmatrix}, \frac{U_{r+1}(\theta)}{\sigma(\theta)} \right) \begin{pmatrix} Y_{h}(\theta) \\ \vdots \\ Y_{rh}(\theta) \end{pmatrix} \right]
\end{equation}

where the last equality follows from (4.4). From this equation we can conclude that

\begin{equation}
E \left[ \phi \left( \begin{pmatrix} Y_{h}(\theta) \\ \vdots \\ Y_{rh}(\theta) \end{pmatrix}, \frac{U_{r+1}(\theta)}{\sigma(\theta)} \right) \begin{pmatrix} Y_{h}(\theta) \\ \vdots \\ Y_{rh}(\theta) \end{pmatrix} \right] = 0.
\end{equation}
and therefore, $\pi(\vartheta)$ is a solution of equation (4.23).

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 (4.2) exists. But then $(\pi'_1, \ldots, \pi'_r) \neq (\pi_1, \ldots, \pi_r)$. Note that the arguments in the derivation of (4.6) still hold if we replace $\sigma(\vartheta)$ in the denominator of the second argument of $\phi$ by $\sigma'$. Thus, we obtain that

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

and therefore

$$
E \left[ \phi \left( \left( \begin{array}{c} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{array} \right), \frac{Y_{(r+1)h}(\vartheta) - \pi'_1 Y_h(\vartheta) - \ldots - \pi'_r Y_h(\vartheta)}{\sigma'} \right) \left( \begin{array}{c} Y_h(\vartheta) \\ \vdots \\ Y_{rh}(\vartheta) \end{array} \right) \right] = 0. \tag{4.7}
$$

Since $P(Y_h(\vartheta), \ldots, Y_{rh}(\vartheta)) = (0, \ldots, 0) = 0$ and $P(Y_{(r+1)h}(\vartheta) - \pi'_1 Y_h(\vartheta) - \ldots - \pi'_r Y_h(\vartheta) = U_{r+1}(\vartheta)) = 0$ due to (D.3) for $\gamma = 0$ and $u \rightarrow \phi(y, u)$ is strictly increasing on the interval $(-u_0, u_0)$ for every $y \in \mathbb{R}^r$ we have that

$$
\left| \frac{Y_{(r+1)h}(\vartheta) - \pi_1 Y_h(\vartheta) - \ldots - \pi_r Y_h(\vartheta)}{\sigma'} \right| \geq u_0 \quad P\text{-a.s.} \tag{4.8}
$$

because otherwise (4.7) cannot hold. Now, $\pi'$ is by assumption also a solution of (4.2b) and hence, we have due to (4.8) and (E.6)

$$
0 = E \left[ \chi \left( \left( \frac{Y_{(r+1)h}(\vartheta) - \pi'_1 Y_h(\vartheta) - \ldots - \pi'_r Y_h(\vartheta)}{\sigma'} \right)^2 \right) \right] \geq \chi(u_0^2) > 0
$$

which is a contradiction. \hfill \Box

**Remark 4.6.**

(a) Assumption (4.4) 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 $\phi$ is valid, e.g., for both the Mallows and Hampel–Krasker–Welsch estimators when the function $\psi$ is chosen as a Huber $\psi_k$–function with $u_0 = k$.

(c) The assumption on $\chi$ is fulfilled, e.g., if $\chi$ is chosen as in Example 4.3(c) with $Z \overset{d}{=} U_1(\vartheta) / \sqrt{\text{Var}(U_1(\vartheta))}$. In the case that the driving Lévy process is a Brownian motion this means that $Z \sim \mathcal{N}(0, 1)$.

**Theorem 4.7.** Suppose that there exists a unique solution $\pi^{GM}(\vartheta^\gamma)$ of (4.2). Then $\pi_n^{GM}(\vartheta^\gamma) \overset{a.s.}{\to} \pi^{GM}(\vartheta^\gamma) \ P\text{-a.s.}$

**Proof.** The proof goes in the same vein as the proof of Bustos (1982, Theorem 2.1). \hfill \Box
Next, we would like to deduce the asymptotic normality of the GM-estimator. We need the following lemma which is the analog of Bustos [1982, Lemma 3.1] under our different model assumptions.

**Lemma 4.8.** Define the map \( \Psi : \mathbb{R}^r \times \mathbb{R}^r \times (0, \infty) \rightarrow \mathbb{R}^{r+1} \) as

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

Furthermore, define the stochastic process \( \Psi(\vartheta^γ) = (\Psi_k(\vartheta^γ))_{k \in \mathbb{N}} \) as \( \Psi_k(\vartheta^γ) = \Psi(Y_{kh}^γ(\vartheta), \ldots, Y_{(k+r-1)h}^γ(\vartheta), \chi^GM(\vartheta^γ)) \). Then

\[
\frac{1}{\sqrt{n-r}} \sum_{k=1}^{n-r} \Psi_k(\vartheta^γ) \xrightarrow{D} \mathcal{N}(0, \mathcal{I}_{GM}(\vartheta^γ)),
\]

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

\[
[\mathcal{I}_{GM}(\vartheta^γ)]_{ij} = \mathbb{E}[\Psi_{ij}(\vartheta^γ)\Psi_{ij}(\vartheta^γ)] + 2 \sum_{k=1}^{\infty} \mathbb{E}[\Psi_{ij}(\vartheta^γ)\Psi_{i+k,j}(\vartheta^γ)]
\]

and \( \Psi_{ij}(\vartheta^γ) \) denotes the \(i\)-th component of \( \Psi_k(\vartheta^γ) \), \(i = 1, \ldots, r+1\). Especially, each \([\mathcal{I}_{GM}(\vartheta^γ)]_{ij}\) is finite for \(i, j \in \{1, \ldots, r+1\}\).

**Proof.** By the Cramer–Wold device, the statement of the Lemma is equivalent to the assertion that

\[
\frac{1}{\sqrt{n-r}} x^T \sum_{k=1}^{n-r} \Psi_k(\vartheta^γ) \xrightarrow{D} \mathcal{N}(0, x^T \mathcal{I}_{GM}(\vartheta^γ)x)
\]

for every \(x \in \mathbb{R}^{r+1}\). According to [Ibragimov 1962, Theorem 1.7), this holds if we can show that

\[
\mathbb{E}|x^T \Psi_k(\vartheta^γ)|^{2+\delta} < \infty \quad (4.10)
\]

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

\[
\sum_{m=1}^{\infty} \alpha_{x^T \Psi(\vartheta^γ)}^{(2+\delta)}(m) < \infty \quad \text{for some } \delta > 0. \quad (4.11)
\]

The same theorem then also states that \(x^T \mathcal{I}_{GM}(\vartheta^γ)x < \infty\) from which we then deduce that for \(i, j \in \{1, \ldots, r+1\}\) the entry \([\mathcal{I}_{GM}(\vartheta^γ)]_{ij}\) is finite and therefore, \(\mathcal{I}_{GM}(\vartheta^γ)\) is well-defined.

We start to show the existence of the \((2+\delta)\)-th moment of \(x^T \Psi_k(\vartheta^γ)\) in (4.10). Therefore, note that

\[
\mathbb{E}|x^T \Psi_k(\vartheta^γ)|^{2+\delta} \leq C \|x\|^{2+\delta} \sum_{i=1}^{r+1} \mathbb{E}\|\Psi_{ij}(\vartheta^γ)\|^{2+\delta} < \infty,
\]

where the last inequality holds since \(\Psi_{ij}(\vartheta^γ)\) is bounded by \((E.2)\) and \((E.6)\).

Finally, the process \((Y_{mh}^γ(\vartheta))_{m \in \mathbb{Z}}\) is strongly mixing and the mixing coefficients satisfy the above condition \((4.11)\) for the following reason. Either we have in the case of replacement outliers that
\[ Y_{mb}(\theta) = G(V_{m}, Z_{m}, Y_{mb}(\theta)) \] for some measurable function \( G \) and the three processes \( (V_{m}), (Z_{m}) \) and \( (Y_{mb}(\theta)) \) are independent, or in the case of additive outliers we have \( Y_{mb}(\theta) = G(V_{m}, W_{m}, Y_{mb}(\theta)) \) for some measurable function \( G \) and the three processes \( (V_{m}), (W_{m}) \) and \( (Y_{mb}(\theta)) \) are independent. Hence, by Bradley (2007, Theorem 6.6(II)), Assumption [D.2] and Marquardt and Stelzer (2007, Proposition 3.34) we receive
\[ \alpha(Y(\theta))(m) \leq \alpha_{2}(m) + \alpha_{Y(\theta)}(m) \leq C \rho^{m} \]
respectively, \( \alpha_{Y(\theta)}(m) \leq \alpha_{V}(m) + \alpha_{W}(m) + \alpha_{Y(\theta)}(m) \leq C \rho^{m} \) for some \( C > 0 \) and \( \rho \in (0, 1) \). Furthermore, \( V_{k}(\theta) \) depends only on the finitely many values \( Y_{kh}(\theta), \ldots, Y_{(k+r)h}(\theta) \) and by Bradley (2007). Remark 1.8(b)) this ensures that \( \alpha_{Y(\theta)}(m) \leq \alpha_{Y(\theta)}(m + r) \leq C \rho^{m} \). Thus, the strong mixing coefficients \( \alpha_{X_{i}}, \Psi_{(\theta)}(m) \) of \( x^{T}\Psi(\theta) \) satisfy the summability condition \( 4.11 \) and the lemma is proven.

**Lemma 4.9.** Let the set \( K \) be given as in Remark 4.4 and for \( \pi = \{\pi_{1}, \ldots, \pi_{n}, \sigma\} \in K \) define
\[
\mathcal{D}_{GM}(\pi, \theta') = \left( \mathbb{E} \left[ \phi \left( Y_{h}^{\gamma} (\theta), \ldots, Y_{nh}^{\gamma} (\theta) \right) \right] \right) \left( \mathbb{E} \left[ \chi \left( \left( Y_{(r+1)h}^{\gamma} (\theta) - \pi_{1} Y_{rh}^{\gamma} (\theta) - \cdots - \pi_{r} Y_{h}^{\gamma} (\theta) \right) \right) \right] \right).
\]
Then \( \nabla_{\pi} \mathcal{D}_{GM}(\pi, \theta') \) exists. Moreover, for any sequence \( (\pi_{n})_{n \in \mathbb{N}} \) with \( \pi_{n} \overset{P}{\rightarrow} \pi^{GM}(\theta) \) as \( n \rightarrow \infty \) we have as \( n \rightarrow \infty \),
\[
\nabla_{\pi} \mathcal{D}_{GM}(\pi_{n}, \theta') \overset{P}{\rightarrow} \nabla_{\pi} \mathcal{D}_{GM}(\pi^{GM}(\theta'), \theta') \).
\]

**Proof.** Note, first that for \( i, j = 1, \ldots, r, \)
\[
\sup_{\pi \in K} \left| \frac{\partial}{\partial \pi_{i}} \phi \left( \left( Y_{h}^{\gamma} (\theta), \ldots, Y_{nh}^{\gamma} (\theta) \right) \right) \right| \leq C \]
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,
\[
\sup_{\pi \in K} \left| \frac{\partial}{\partial \sigma} \phi \left( \left( Y_{h}^{\gamma} (\theta), \ldots, Y_{nh}^{\gamma} (\theta) \right) \right) \right| \leq C \]

17
The expectation on the right-hand side is finite due to Assumption \ref{E5}. Similar arguments, using Assumption \ref{E6}, also show that
\[
\left| \frac{\partial}{\partial \pi} K \left( \frac{Y_i^T(\theta)}{\sigma} \right) \right| \text{ for } i = 1, \ldots, r
\]
are uniformly dominated by integrable random variables.

Therefore, by (Billingsley, 1999, Theorem 16.8(ii)) (that is an application of dominated convergence) \( \mathbb{V}_x \mathcal{D}_{\pi} \) exists on \( K \) and the order of differentiation and expectation can be changed.

Moreover, due \ref{E4}, \ref{E6} and (Billingsley, 1999, Theorem 16.8(i)) the map \( \pi \mapsto \mathbb{V}_x \mathcal{D}_{\pi} \) is continuous. Hence, if \( \pi_n \xrightarrow{P} \pi^\text{GM}(\theta^\gamma) \in K \) then \( \mathbb{V}_x \mathcal{D}_{\pi_n} \xrightarrow{P} \mathbb{V}_x \mathcal{D}_{\pi^\text{GM}(\theta^\gamma)} \).

\textbf{Lemma 4.10.} Let \( \mathcal{D}_{\pi}(\theta^\gamma) \) be defined as in \ref{4.13} and suppose that \( \mathbb{V}_x \mathcal{D}_{\pi} \) is non-singular. Furthermore, let \( \mathcal{J}_{\pi}(\theta^\gamma) \) be given as in \ref{4.9} and suppose that \( \hat{\pi}_n^\text{GM}(\theta^\gamma) \xrightarrow{P} \pi^\text{GM}(\theta^\gamma) \) as \( n \to \infty \). Then, as \( n \to \infty \),
\[
\sqrt{n_r} \mathcal{D}_{\pi_n} \left( \hat{\pi}_n^\text{GM}(\theta^\gamma), \theta^\gamma \right) \xrightarrow{D} \mathcal{N}(0, \mathcal{J}_{\pi}(\theta^\gamma)).
\]

\textbf{Proof.} We use the decomposition
\[
\sqrt{n_r} \mathcal{D}_{\pi_n} \left( \hat{\pi}_n^\text{GM}(\theta^\gamma), \theta^\gamma \right) = \frac{1}{\sqrt{n_r}} \sum_{k=1}^{n-r} \left[ \mathcal{D}_{\pi_n} \left( \hat{\pi}_n^\text{GM}(\theta^\gamma), \theta^\gamma \right) \right] - \frac{1}{\sqrt{n_r}} \sum_{k=1}^{n-r} \Psi_k(\theta^\gamma).
\]
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 \( \mathcal{N}(0, \mathcal{J}_{\pi}(\theta^\gamma)) \) due to Lemma \ref{4.8}. Hence, we receive the statement.

Finally, we obtain the following analog version of Bustos (1982, Theorem 2.2) in our setting.

\textbf{Theorem 4.11.} Let \( \mathcal{D}_{\pi}(\theta^\gamma), \theta^\gamma \) be defined as in \ref{4.13} and suppose that \( \mathcal{J}_{\pi}(\theta^\gamma) := \mathbb{V}_x \mathcal{D}_{\pi} \) is non-singular. Furthermore, let \( \mathcal{J}_{\pi}(\theta^\gamma) \) be given as in \ref{4.9} and suppose that \( \hat{\pi}_n^\text{GM}(\theta^\gamma) \xrightarrow{P} \pi^\text{GM}(\theta^\gamma) \) as \( n \to \infty \). Then, as \( n \to \infty \),
\[
\sqrt{n_r} \mathcal{D}_{\pi_n} \left( \hat{\pi}_n^\text{GM}(\theta^\gamma), \theta^\gamma \right) \xrightarrow{D} \mathcal{N}(0, \mathcal{J}_{\pi}(\theta^\gamma)),
\]
where
\[
\mathcal{E}_{\pi}(\theta^\gamma) := \left[ \mathcal{J}_{\pi}(\theta^\gamma) \right]^{-1} \mathcal{J}_{\pi}(\theta^\gamma) \mathcal{J}_{\pi}(\theta^\gamma) \left[ \mathcal{J}_{\pi}(\theta^\gamma) \right]^{-1}.
\]

\textbf{Proof.} Due to \ref{4.2} we have \( \mathcal{D}_{\pi}(\theta^\gamma), \theta^\gamma \) = 0. Next, a first-order Taylor expansion around \( \hat{\pi}_n^\text{GM}(\theta^\gamma) \) gives
\[
0 = \sqrt{n_r} \mathcal{D}_{\pi} \left( \hat{\pi}_n^\text{GM}(\theta^\gamma), \theta^\gamma \right)
= \sqrt{n_r} \mathcal{D}_{\pi} \left( \hat{\pi}_n^\text{GM}(\theta^\gamma), \theta^\gamma \right) + \sqrt{n_r} \mathcal{D}_{\pi} \left( \hat{\pi}_n^\text{GM}(\theta^\gamma), \theta^\gamma \right) \pi^\text{GM}(\theta^\gamma) - \hat{\pi}_n^\text{GM}(\theta^\gamma)),
\]
where \( \| \pi^\text{GM}(\theta^\gamma) - \hat{\pi}_n^\text{GM}(\theta^\gamma) \| \leq \| \pi^\text{GM}(\theta^\gamma) - \hat{\pi}_n^\text{GM}(\theta^\gamma) \| \). The statement follows then from a combination of Lemma \ref{4.9} and Lemma \ref{4.10}.
5 The indirect estimator for the CARMA parameters

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 the least-squares- (LS-) estimator \( \hat{\pi}_n \) and for \( \hat{\pi}_n \) the GM-estimator \( \hat{\pi}_n^{GM} \).

**Definition 5.1.** Based on the sample \( \mathcal{Y}^S_n(\vartheta) = (Y^S_n(\vartheta), \ldots, Y^S_{nh}(\vartheta)) \) the LS-estimator \( \hat{\pi}_n^{LS}(\vartheta) = (\hat{\pi}_{n,1}^{LS}(\vartheta), \ldots, \hat{\pi}_{n,r}^{LS}(\vartheta), \hat{\sigma}_n^{LS}(\vartheta)) \) of \( \vartheta \) minimizes

\[
L_{LS}(\pi, \mathcal{Y}^S_n(\vartheta)) := \frac{1}{sn-r} \sum_{k=1}^{sn-r} \left( Y^S_{(k+r)h}(\vartheta) - \pi_1 Y^S_{(k+r-1)h}(\vartheta) - \ldots - \pi_r Y^S_{kh}(\vartheta) \right)^2
\]

in \( \Pi' := \pi(\Theta) \) and \( \hat{\sigma}_n^{LS}(\vartheta) \) is defined as

\[
\hat{\sigma}_n^{LS}(\vartheta) = \frac{1}{sn-r} \sum_{k=1}^{sn-r} \left( Y^S_{(k+r)h}(\vartheta) - \hat{\pi}_{n,1}^{LS}(\vartheta) Y^S_{(k+r-1)h}(\vartheta) - \ldots - \hat{\pi}_{n,r}^{LS}(\vartheta) Y^S_{kh}(\vartheta) \right)^2.
\]

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

\[
L_{MLE}(\pi, \mathcal{Y}^S_n(\vartheta)) = \frac{1}{sn-r} \sum_{k=1}^{sn-r} \left( \log(\sigma^2) + \frac{(Y^S_{(k+r)h}(\vartheta) - \pi_1 Y^S_{(k+r-1)h}(\vartheta) - \ldots - \pi_r Y^S_{kh}(\vartheta))^2}{\sigma^2} \right)
\]

and the quasi ML-estimator as \( \hat{\pi}_n^{MLE}(\vartheta) = \arg\min_{\pi \in \Pi'} L_{MLE}(\pi, \mathcal{Y}^S_n(\vartheta)) \).

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 \( L_{MLE} \) which are proportional to the derivatives of \( L_{LS} \)).

**Theorem 5.3.** Let Assumptions A, B, D and E hold. Suppose that the unique solution \( \pi^{GM}(\vartheta)_0 \) of (4.2) for \( (Y_{mh})_{m \in \mathbb{Z}} = \pi(\vartheta)_0 \) that \( \nabla_{\vartheta} \pi(\vartheta)_0 \) has full column rank \( N(\Theta) \) and that \( J_{GM}(\vartheta)_0 \) is non-singular. Further, assume that \( E[|L|^2] \leq N^* \) for some \( N^* \in \mathbb{N} \) with \( 2N^* > \max\{N(\Theta), 4 + \delta\} \). If \( \hat{\pi}_n(\vartheta) = \hat{\pi}_n^{LS}(\vartheta) \) and \( \hat{\pi}_n = \hat{\pi}_n^{GM}(\vartheta)_0 \) then the indirect estimator \( \hat{\pi}_n^{Ind} \) is weakly consistent and

\[
\sqrt{n}(\hat{\pi}_n^{Ind} - \vartheta_0) \xrightarrow{p} \mathcal{N}(0, \Xi_{Ind}(\vartheta)_0),
\]

where

\[
\Xi_{Ind}(\vartheta)_0 = J_{Ind}(\vartheta)_0^{-1} \mathcal{J}_{Ind}(\vartheta)_0 J_{Ind}(\vartheta)_0^{-1}
\]

with

\[
J_{Ind}(\vartheta)_0 = [\nabla_{\vartheta} \pi(\vartheta)_0]^T \Omega [\nabla_{\vartheta} \pi(\vartheta)_0] \quad \text{and} \quad J_{Ind}(\vartheta)_0 = [\nabla_{\vartheta} \pi(\vartheta)_0]^T \Omega \left[ \Xi_{GM}(\vartheta)_0 + \frac{1}{s} \Xi_{LS}(\vartheta)_0 \right] \Omega [\nabla_{\vartheta} \pi(\vartheta)_0],
\]

where the matrix \( \Xi_{LS}(\vartheta) \) is defined as in (4.14) 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 \( \hat{\pi}_n^L(\vartheta) \) we require several auxiliary lemmata. The remaining of this section is devoted to that.

For the ease of notation we write in the following for the Lévy process \((L_t^\vartheta)_{t \in \mathbb{R}}\) 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 \((Y_t^\vartheta)_{t \in \mathbb{R}}\) is \((Y_t)_{t \in \mathbb{R}}\).

**Lemma 5.4.** Define for any \( \vartheta \in \Theta \) the function \( f_\vartheta(u) = e_\vartheta^T e_\vartheta^u e_p^1 \mathbb{1}_{[0, \infty)}(u) \) and

\[
G_\vartheta(u) = (f_\vartheta(u), \nabla_\vartheta f_\vartheta(u), \nabla_\vartheta^2 f_\vartheta(u)).
\]

Then \( \mathbb{P}\text{-a.s. we have} \)

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

which is strongly mixing and ergodic.

The proof is moved to Appendix A.

**Proposition 5.5.** For \( j, l \in \{0, \ldots, r\} \) define

\[
\hat{\vartheta}_{0,n}(l, j) = \frac{1}{n - r} \sum_{k=1}^{n-r} Y_{(k+j)l}(\vartheta) Y_{(k+j)j}(\vartheta).
\]

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

(a) \( \sup_{\vartheta \in \Theta} |\hat{\vartheta}_{0,n}(l, j) - \vartheta(l - j)| \xrightarrow{\mathbb{P}} 0. \)

(b) \( \sup_{\vartheta \in \Theta} \left| \frac{\partial}{\partial \vartheta_l} \hat{\vartheta}_{0,n}(l, j) - \frac{\partial}{\partial \vartheta_l} \vartheta_0(l - j) \right| \xrightarrow{\mathbb{P}} 0. \)

(c) \( \sup_{\vartheta \in \Theta} \left| \frac{\partial^2}{\partial \vartheta_l \partial \vartheta_u} \hat{\vartheta}_{0,n}(l, j) - \frac{\partial^2}{\partial \vartheta_l \partial \vartheta_u} \vartheta_0(l - j) \right| \xrightarrow{\mathbb{P}} 0. \)

**Proof.** (a) First, we prove the pointwise convergence of the sample autocovariance function and second, that \( \hat{\vartheta}_{0,n}(l, j) \) 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 5.4 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 \( n \to \infty, \)

\[
\hat{\vartheta}_{0,n}(l, j) \xrightarrow{\mathbb{P}} \vartheta_0(l - j).
\]

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

\[
U_k := \sup_{\vartheta_1, \vartheta_2 \in \Theta} \frac{|Y_{kh}(\vartheta_1) - Y_{kh}(\vartheta_2)|}{\|\vartheta_1 - \vartheta_2\|_\gamma}.
\]

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

\[
|Y_{kh}(\vartheta_1) - Y_{kh}(\vartheta_2)| \leq U_k \|\vartheta_1 - \vartheta_2\|_\gamma
\]

20
and hence,
\[ |Y_{(k+1)h}^{(0)} \hat{\vartheta}_1 Y_{(k+1)h}^{(0)} \hat{\vartheta}_2 - Y_{(k+1)h}^{(0)} \hat{\vartheta}_1 Y_{(k+1)h}^{(0)} \hat{\vartheta}_2| \]
\[ \leq \left( \sup_{\vartheta \in \Theta} \right. |Y_{(k+1)h}^{(0)} \hat{\vartheta}| + \sup_{\vartheta \in \Theta} |Y_{(k+1)h}^{(0)} \vartheta| \right) (U_{k+l} + U_{k+l}) \| \vartheta_1 - \vartheta_2 \|^7 
\]
hold. Finally,
\[ |\hat{\gamma}_{l,n}(l,j) - \hat{\gamma}_{n,l}(l,j)| \leq \frac{1}{n-r} \sum_{k=1}^{n-r} U_{k+l}^* \| \vartheta_1 - \vartheta_2 \|^7 \text{ for } \| \vartheta_1 - \vartheta_2 \| < 1 \] (5.2)
with
\[ \mathbb{E}(U_{k+l}^*) = \mathbb{E}(U_{l+1,1+l}) \leq 4 \left( \mathbb{E} \left( \sup_{\vartheta \in \Theta} |Y_{l}^{(0)} \vartheta| \right)^2 \right)^{1/2} < \infty \]
where we used Lemma A.3 to get the finite expectation.

**Step 3.** Let \( \varepsilon, \eta > 0 \). Take \( 0 < \delta < \min\{1, \eta \varepsilon / \mathbb{E}(U_{l+1,1+l})^{1/2} \} \). Then (5.2) and Markov’s inequality give
\[ \mathbb{P} \left( \sup_{0 < |\vartheta_1 - \vartheta_2| < \delta} |\hat{\gamma}_{l,n}(l,j) - \hat{\gamma}_{n,l}(l,j)| > \eta \right) \leq \mathbb{E}(U_{l+1,1+l}) \frac{\delta^7}{\eta} < \varepsilon. \]
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).

**Proposition 5.6.**

(a) \( \sup_{\vartheta \in \Theta} |\hat{\pi}_{l,n}^{(0)}(\vartheta) - \pi(\vartheta)| \xrightarrow{p} 0. \)

(b) \( \sup_{\vartheta \in \Theta} |\nabla_{\vartheta} \hat{\pi}_{l,n}^{(0)}(\vartheta) - \nabla_{\vartheta} \pi(\vartheta)| \xrightarrow{p} 0. \)

(c) \( \sup_{\vartheta \in \Theta} |\nabla_{\vartheta}^2 \hat{\pi}_{l,n}^{(0)}(\vartheta) - \nabla_{\vartheta}^2 \pi(\vartheta)| \xrightarrow{p} 0. \)

**Proof.** Define
\[ \hat{\gamma}_{l,n}^{(r-1)}(\vartheta) = \begin{pmatrix} \hat{\gamma}_{l,n}(r, r-1) \\ \vdots \\ \hat{\gamma}_{l,n}(r, 0) \end{pmatrix} \quad \text{and} \quad \hat{\gamma}_{l,n}^{(r-1)}(\vartheta) = \begin{pmatrix} \hat{\gamma}_{l,n}(r-1, r-1) \\ \vdots \\ \hat{\gamma}_{l,n}(r-1, 0) \end{pmatrix}. \]

Then
\[ \hat{\pi}_{l,n}^{(r)}(\vartheta) := \begin{pmatrix} \hat{\pi}_{l,n}^{(r)}(\vartheta) \\ \vdots \\ \hat{\pi}_{l,n}^{(r)}(\vartheta) \end{pmatrix} = \left[ \hat{\gamma}_{l,n}^{(r-1)}(\vartheta) \right]^{-1} \hat{\gamma}_{l,n}^{(r-1)}(\vartheta), \]
(5.3)
\[ \sigma_{l,n}^{(r)}(\vartheta) = \hat{\gamma}_{l,n}(r, r) - \left[ \hat{\pi}_{l,n}^{(r)}(\vartheta) \right]^{T} \hat{\gamma}_{l,n}^{(r-1)}(\vartheta). \]
A conclusion of Proposition 5.5(a) and the definition of $\Gamma^{(r-1)}(\hat{\theta})$ and $\gamma^{(r-1)}(\hat{\theta})$ in (2.5) is that
\[
\sup_{\theta \in \Theta} \| \Gamma_n^{(r-1)}(\theta) - \Gamma^{(r-1)}(\hat{\theta}) \| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\theta \in \Theta} \| \gamma_n^{(r-1)}(\theta) - \gamma^{(r-1)}(\hat{\theta}) \| \xrightarrow{p} 0. \tag{5.4}
\]
Due to the continuity and the positive definiteness of $\Gamma^{(r-1)}(\hat{\theta})$ (cf. proof of Proposition 2.2), and the compactness of $\Theta$ we receive $\sup_{\theta \in \Theta} \| \Gamma^{(r-1)}(\hat{\theta}) \| < \infty$. Hence, statement (a) is a consequence of (5.3), (5.4) and (2.5).

(b) Note that
\[
\begin{align*}
\frac{\partial}{\partial \theta_i} \pi_n^{\hat{\theta}}(\hat{\theta}) &= -[\Gamma_n^{(r-1)}(\hat{\theta})]^{-1} \left[ \frac{\partial}{\partial \theta_i} \Gamma_n^{(r-1)}(\hat{\theta}) \right] \Gamma_n^{(r-1)}(\hat{\theta})^{-1} \left[ \frac{\partial}{\partial \theta_i} \gamma_n^{(r-1)}(\hat{\theta}) \right] \\
\frac{\partial}{\partial \theta_i} \pi^{\hat{\theta}}(\hat{\theta}) &= -[\Gamma^{(r-1)}(\hat{\theta})]^{-1} \left[ \frac{\partial}{\partial \theta_i} \Gamma^{(r-1)}(\hat{\theta}) \right] \Gamma^{(r-1)}(\hat{\theta})^{-1} \left[ \frac{\partial}{\partial \theta_i} \gamma^{(r-1)}(\hat{\theta}) \right].
\end{align*}
\tag{5.5}
\]
Applying Proposition 5.5(b) we receive that
\[
\sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta_i} \pi_n^{\hat{\theta}}(\hat{\theta}) - \frac{\partial}{\partial \theta_i} \Gamma^{(r-1)}(\hat{\theta}) \right\| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta_i} \pi^{\hat{\theta}}(\hat{\theta}) - \frac{\partial}{\partial \theta_i} \gamma^{(r-1)}(\hat{\theta}) \right\| \xrightarrow{p} 0. \tag{5.6}
\]
Then the same arguments as in (a) and (5.4)-(5.6) lead to statement (b).

(c) The proof goes in analogous lines as in (a) and (b).

\[\square\]

**Corollary 5.7.** Let $\hat{\theta}_n$ be a sequence in $\Theta$ with $\hat{\theta}_n \xrightarrow{p} \hat{\theta}_0$. Then the following statements hold:

(a) $\pi_n^{\hat{\theta}_n}(\hat{\theta}_n) \xrightarrow{p} \pi(\hat{\theta}_0)$.

(b) $\nabla_\theta \pi_n^{\hat{\theta}_n}(\hat{\theta}_n) \xrightarrow{p} \nabla_\theta \pi(\hat{\theta}_0)$.

(c) $\nabla_\theta^2 \pi_n^{\hat{\theta}_n}(\hat{\theta}_n) \xrightarrow{p} \nabla_\theta^2 \pi(\hat{\theta}_0)$.

**Proof.** (a) We use the upper bound
\[
\|\pi_n^{\hat{\theta}_n}(\hat{\theta}_n) - \pi(\hat{\theta}_0)\| \leq \sup_{\theta \in \Theta} \|\pi_n^{\hat{\theta}_n}(\theta) - \pi(\hat{\theta}_0)\| + \|\pi(\hat{\theta}_n) - \pi(\hat{\theta}_0)\|.
\]
The first term converges in probability to 0 due Proposition 5.6(a) and the second term because $\pi(\hat{\theta})$ is continuous (see Lemma 2.5) and $\hat{\theta}_n \xrightarrow{p} \hat{\theta}_0$. The proof of (b,c) goes on the same way. \[\square\]

**Theorem 5.8.** For any $\hat{\theta} \in \Theta$ the LS-estimator $\pi_n^{\hat{\theta}(\hat{\theta})}$ is strongly consistent and as $n \to \infty$,
\[
\sqrt{n} \left(\pi_n^{\hat{\theta}(\hat{\theta})} - \pi(\hat{\theta})\right) \xrightarrow{d} N\left(0, \Xi_{\text{LS}}(\hat{\theta})\right).
\]

**Proof.** Due to Proposition 5.6(a) we already know that the LS-estimator $\pi_n^{\hat{\theta}(\hat{\theta})}$ is consistent. The asymptotic normality of $\pi_n^{\hat{\theta}(\hat{\theta})}$ follows in principle from Theorem 4.11 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 4.11 is that the Jacobian $\mathcal{J}_{\text{GM}}(\hat{\theta}) = \nabla_\pi \mathcal{J}_{\text{GM}}(\pi(\hat{\theta}), \hat{\theta})$ is non–singular. This we can verify
by direct calculations for the LS-estimator because

$$
\nabla_{\pi} \mathcal{L}_S(\pi, \vartheta) = \mathcal{J}_S(\vartheta) = -\frac{1}{\sigma(\vartheta)} \begin{pmatrix}
\gamma_0(0) & \gamma_0(h) & \ldots & \gamma_0((r-1)h) & 0 \\
\gamma_0(h) & \gamma_0(0) & \ldots & \gamma_0((r-2)h) & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma_0((r-1)h) & \ldots & \ldots & \gamma_0(0) & 0 \\
0 & 0 & \ldots & 0 & 2
\end{pmatrix}.$$

Hence, $\mathcal{J}_S(\vartheta)$ 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 $(Y_h(\vartheta), \ldots, Y_{rh}(\vartheta))$ which is non–singular (cf. proof of Proposition 2.3).

Still, we need to be careful because the function $\phi$ and $\chi$ do not satisfy Assumptions (E.2), (E.4) and (E.6) with respect to boundedness. However, a close inspection of the proof of Theorem 4.11 reveals that the boundedness is only used at two points. First, in Lemma 4.8 where we deduce the finiteness of the expectation in (4.12). However, for the LS-estimator

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

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

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

Therefore, inequality (4.12) follows since the Lévy process $(L_t)_{t \in \mathbb{R}}$ 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 4.9 to deduce the existence of $\nabla_{\pi} \mathcal{L}_S(\pi, \vartheta)$ and its continuity. But by the above calculations $\nabla_{\pi} \mathcal{L}_S(\pi, \vartheta)$ exists obviously and is continuous.

Proof of Theorem 5.3 [C.1] and [C.4] follow from Theorem 4.7 and Theorem 4.11 [C.2] is proven in Proposition 5.6 [C.3] is a consequence of Theorem 5.8. Finally, [C.5] is derived in Corollary 5.7.

### 6 Robustness properties of the indirect estimator

In this section we study the robustness properties of the indirect estimator for the CARMA parameters of $(Y_{mh})_{m \in \mathbb{Z}}$ where we assume that $\hat{\theta}_n = \hat{\theta}^{GM}(\vartheta_0)$ is the GM-estimator satisfying Assumptions A1 B1 [D.3] for $(Y_{mh})_{m \in \mathbb{Z}}$ and that $\pi(\vartheta_0)$ is the unique solution of (4.2) for $(Y_{mh})_{m \in \mathbb{Z}}$ (a sufficient criterion for this is given in Proposition 4.3). Moreover, we require similarly to (E.2):

(E2') $(y,u) \mapsto \phi(y,u)$ is bounded and there exists $c > 0$ such that

$$
\frac{1}{2} \| \phi(y_1, u_1)y_1 - \phi(y_2, u_2)y_2 \| \leq \frac{c}{\min\{\|y_1\|, \|y_2\|\}} \|y_1 - y_2\|, \\
\frac{1}{2} \| \phi(u_1)y_1 - \phi(u_2)y_2 \| \leq \frac{c}{\min\{|u_1|, |u_2|\}} \|u_1 - u_2\|.
$$

Under some mild assumptions on $\psi(u)$ and $w(y)$ both the Mallows estimator and the Hampel–Krasker–Welsch estimator satisfy these conditions (cf. Boente et al. (1987, p.1305)). For the simula-
tion part we take some estimator $\widehat{\pi}^2_n(\vartheta)$, e.g., the LS-estimator, such that as $n \to \infty$,
\[
\sup_{\vartheta \in \Theta} \|\widehat{\pi}^2_n(\vartheta) - \pi(\vartheta)\| \xrightarrow{P} 0
\]
holds.

### 6.1 Resistance and qualitative robustness

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 $\pi$–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 $\pi_{d_\varrho}$–robustness for time series is equivalent to Hampel’s $\pi$–robustness for i.i.d. random variables and therefore, extends Hampel’s $\pi$–robustness. They go ahead and define the term of 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 while the term of 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 while the term of 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 while the term of 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 while the term of 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 while the term of 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 while the term of 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 while

**Definition 6.1.** Let $y \in \mathbb{R}^\infty$ and let $(\widehat{\vartheta}_n)_{n \in \mathbb{N}}$ be a sequence of estimators. Denote by $\widehat{\vartheta}_n(z^n)$ the value of $\vartheta$ when it is calculated using the deterministic realization $z^n \in \mathbb{R}^n$.

(a) $(\widehat{\vartheta}_n)_{n \in \mathbb{N}}$ is called resistant at $y$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that
\[
\sup \left\{ \|\widehat{\vartheta}_n(z^n) - \widehat{\vartheta}_n(w^n)\| : z^n, w^n \in B_\delta(y^n) \right\} \leq \varepsilon \quad \forall n \in \mathbb{N},
\]
where $B_\delta(x)$ denotes an open ball with center $x$ and radius $\delta$ with respect to the metric
\[
d_n(z^n, w^n) = \inf \left\{ \varepsilon : \#\{i \in \{1, \ldots, n\} : |z^n_i - w^n_i| \geq \varepsilon \} \leq \frac{n}{\varepsilon} \right\},
\]
(b) $(\widehat{\vartheta}_n)_{n \in \mathbb{N}}$ is called asymptotically resistant at $y$ if for any $\varepsilon > 0$ there exists a $\delta > 0$ and $N_0(\varepsilon, y) \in \mathbb{N}$ such that (6.1) holds for $n \geq N_0(\varepsilon, y)$.
(c) For $Q \in \mathcal{P}(\mathbb{R}^\infty)$ we say that $(\widehat{\vartheta}_n)_{n \in \mathbb{N}}$ is strongly resistant at $Q$ if
\[
Q \left\{ y \in \mathbb{R}^\infty : (\widehat{\vartheta}_n)_{n \in \mathbb{N}} is resistant at y \right\} = 1.
\]
(d) \((\hat{\theta}_n)_{n\in\mathbb{N}}\) is called asymptotically strongly resistant at \(Q\) if
\[
Q\left(\left\{ y \in \mathbb{R}^m : (\hat{\theta}_n)_{n\in\mathbb{N}} \text{ is asymptotically resistant at } y \right\}\right) = 1.
\]

(e) \((\hat{\theta}_n)_{n\in\mathbb{N}}\) is called weakly resistant at \(Q\) if for any \(\varepsilon > 0\) there exists a \(\delta > 0\) such that
\[
Q\left(\left\{ y \in \mathbb{R}^m : \sup \left\{ \| \hat{\theta}_n(z^n) - \hat{\theta}_n(w^n) \| : z^n, w^n \in B_\delta(y^n) \right\} \leq \varepsilon \right\}\right) \geq 1 - \varepsilon \quad \forall n \in \mathbb{N}.
\]

(f) \((\hat{\theta}_n)_{n\in\mathbb{N}}\) is called asymptotically weakly resistant at \(Q\) if for any \(\varepsilon > 0\) there exist a \(\delta > 0\) and \(N(\varepsilon) \in \mathbb{N}\) such that
\[
Q\left(\left\{ y \in \mathbb{R}^m : \sup \left\{ \| \hat{\theta}_n(z^n) - \hat{\theta}_n(w^n) \| : z^n, w^n \in B_\delta(y^n) \right\} \leq \varepsilon \right\}\right) \geq 1 - \varepsilon \quad \forall n \geq N(\varepsilon).
\]

With this definition at hand, we want to study the question whether our indirect estimator for the parameters of a CARMA processes is resistant. We will make use of the fact that the indirect estimator consists of two independent parts, the GM-estimator for the parameters of the auxiliary AR representation, which deals with possible outliers in the observations, and the outlier-free estimator of the AR parameters based on simulated data.

**Theorem 6.2.** The GM-estimator \((\hat{\pi}^\text{GM}_{y(0)})_{n \in \mathbb{N}}\) is strongly resistant at \(\mathbb{P}_{y(0)}\).

**Proof.** First of all, \((\hat{\pi}^\text{GM}_{y(0)})_{n \in \mathbb{N}}\) is asymptotically strongly resistant at \(\mathbb{P}_{y(0)}\). This follows from [Boente et al. 1987, Theorem 5.1]. The theorem requires that \(\phi\) and \(\chi\) fulfill Assumption E (E2') and that the limiting equation has a unique solution, which we assumed. Moreover, \((Y_{mh})_{m \in \mathbb{Z}}\) is stationary and ergodic due to [Marquardt and Stelzer 2007, Proposition 3.34] and fulfills (D.3).

Next, by [Cox 1981, Lemma 5] \(\hat{\pi}^\text{GM}_{y(0)}\) is a continuous function of \(\theta^n\) for every \(n \in \mathbb{N}\). This in combination with the asymptotically strongly resistance of \((\hat{\pi}^\text{GM}_{y(0)})_{n \in \mathbb{N}}\) at \(\mathbb{P}_{y(0)}\) implies the strong resistance due to [Boente et al. 1987, Proposition 4.2].

**Theorem 6.3.** The indirect estimator \((\hat{\pi}^\text{Ind}_{y(0)})_{n \in \mathbb{N}}\) is weakly resistant and asymptotically weakly resistant at \(\mathbb{P}_{y(0)}\).

**Proof.** Let \(\varepsilon > 0\). Since \(\varphi_{\text{Ind}}\) has a unique minimum in \(\theta_0\)
\[
\eta := 3^{-1} \inf_{\| \theta - \theta_0 \| > \varepsilon / 4} \varphi_{\text{Ind}}(\theta) > 0.
\]

Moreover, the map \(x \mapsto x^T \Omega x\) is continuous and hence, uniformly continuous on the compact set \(\Pi' = \pi(\Theta)\). Thus, there exists an \(\varepsilon' > 0\) such that
\[
\sup_{x, x' \in \Pi'} \left| x^T \Omega x - x'^T \Omega x' \right| \leq \frac{\eta}{8}.
\]

Define
\[
\Omega_0(\varepsilon, \delta) := \bigcap_{n \in \mathbb{N}} \left\{ y \in \mathbb{R}^m : \sup \left\{ \left\| \hat{\pi}_n(z^n) - \hat{\pi}_n(w^n) \right\| : z^n, w^n \in B_\delta(y^n) \right\} \leq \varepsilon / 2 \right\}.
\]
Likewise, (6.2) and (6.4) give us that
de э(Boente et al., 1987, Proposition 4.1) there exists an δ > 0 such that for
e := \min\{\epsilon, \epsilon', \eta/(8\|\Omega\| \sup_{\pi \in \Pi} \|\pi\|)\}
\mathbb{P}(\Omega_0(\epsilon, \delta)) \geq 1 - \frac{\epsilon}{2} \geq 1 - \frac{\epsilon}{2}.

Let y \in \Omega_0(\epsilon, \delta) and \epsilon'' \in B_\delta(y^n). Then for any n \in \mathbb{N},
\sup_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, y^n) - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
= \sup_{\theta \in \Theta} |\epsilon''| \sup_{\theta \in \Theta} \mathcal{L}_\text{Ind}(\theta, y^n) - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\leq \sup_{\pi \in \Pi} \mathbb{E}(\mathcal{L}_\text{Ind}(\theta, y^n) - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\leq \frac{\eta}{4}. (6.2)

On the other hand, from Theorem 3.2 a) we know that \sup_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, y^n) - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\rightarrow 0. Hence, there exists an N(\epsilon) \in \mathbb{N} such that
\mathbb{P}\left(\sup_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, y^n) - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\leq \eta\right) > 1 - \frac{\epsilon}{2} \quad \forall n \geq N(\epsilon).

Define \Omega_n := \{y \in \mathbb{R}^n : \sup_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, y^n) - \mathcal{L}_\text{Ind}(\theta, \epsilon'')| \leq \eta\} and let y \in \Omega_n \cap \Omega_0(\epsilon, \delta) and n \geq N(\epsilon).
Then
|\mathcal{L}_\text{Ind}(\theta_0, y^n)|
= |\mathcal{L}_\text{Ind}(\theta_0, y^n) - \mathcal{L}_\text{Ind}(\theta_0, \epsilon'')|
\leq \eta, (6.3)
and with the definition of \eta we receive
\inf_{|\theta - \theta_0| \geq \frac{\epsilon}{4}} |\mathcal{L}_\text{Ind}(\theta, y^n)|
\geq \inf_{|\theta - \theta_0| \geq \frac{\epsilon}{4}} |\mathcal{L}_\text{Ind}(\theta, y^n)|
\geq \inf_{|\theta - \theta_0| \geq \frac{\epsilon}{4}} |\mathcal{L}_\text{Ind}(\theta, y^n) - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\geq 3\eta - \eta = 2\eta. (6.4)

Since \hat{\theta}_n^{\text{Ind}}(y^n) minimizes \mathcal{L}_\text{Ind}(\theta, y^n) we can deduce from (6.3) and (6.4) that
|\hat{\theta}_n^{\text{Ind}}(y^n) - \theta_0| < \frac{\epsilon}{4}. (6.5)

Let \epsilon'' \in B_\delta(y^n). Due to (6.2) and (6.3) we obtain
|\mathcal{L}_\text{Ind}(\theta_0, \epsilon'')|
\leq \sup_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, \epsilon'') - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\leq \frac{\eta}{4} + \eta = \frac{5\eta}{4}. (6.6)

Likewise, (6.2) and (6.4) give us that
\inf_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\geq \inf_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\geq \inf_{\theta \in \Theta} |\mathcal{L}_\text{Ind}(\theta, \epsilon'') - \mathcal{L}_\text{Ind}(\theta, \epsilon'')|
\geq \frac{7\eta}{4}. (6.7)

Since \hat{\theta}_n^{\text{Ind}}(\epsilon'') minimizes \mathcal{L}_\text{Ind}(\theta, \epsilon'') we can conclude from (6.6) and (6.7) that
|\hat{\theta}_n^{\text{Ind}}(\epsilon'') - \theta_0| < \frac{\epsilon}{4}. (6.8)
Finally, (6.5) and (6.8) result in
\[
\|\hat{\vartheta}_n^\text{Ind}(z^n) - \hat{\vartheta}_n^\text{Ind}(y^n)\| \leq \|\hat{\vartheta}_n^\text{Ind}(z^n) - \vartheta_0\| + \|\hat{\vartheta}_n^\text{Ind}(y^n) - \vartheta_0\| < \frac{\varepsilon}{2}.
\]
To summarize, for \(n \geq N(\varepsilon)\) we have \(\mathbb{P}_{Y(\delta)}(\Omega_0(\varepsilon, \delta) \cap \Omega_n) \geq 1 - \varepsilon\), and for \(y \in \Omega_0(\varepsilon, \delta) \cap \Omega_n\) and \(z^n, w^n \in B_\delta(y^n)\) we have
\[
\|\hat{\vartheta}_n^\text{Ind}(z^n) - \hat{\vartheta}_n^\text{Ind}(w^n)\| \leq \|\hat{\vartheta}_n^\text{Ind}(z^n) - \hat{\vartheta}_n^\text{Ind}(y^n)\| + \|\hat{\vartheta}_n^\text{Ind}(y^n) - \hat{\vartheta}_n^\text{Ind}(w^n)\| < \varepsilon.
\]
This gives the asymptotically weakly resistance of \((\hat{\vartheta}_n^\text{Ind})_{n \in \mathbb{N}}\) at \(\mathbb{P}_{Y(\delta)}\).

By definition, \(\hat{\vartheta}_n^\text{Ind}\) depends on \(\vartheta^n\) through a continuous function applied to \(\hat{\pi}_n^\text{GM}(\vartheta_0)\) and therefore, \(\hat{\vartheta}_n^\text{Ind}\) is a continuous function in \(\vartheta^n\). This and the asymptotically weakly resistance of \((\hat{\vartheta}_n^\text{Ind})_{n \in \mathbb{N}}\) at \(\mathbb{P}_{Y(\delta)}\) imply the weakly resistance at \(\mathbb{P}_{Y(\delta)}\) by Boente et al. (1987, Proposition 4.2).

**Remark 6.4.** If the stronger version \(\sup_{\vartheta \in \Theta} \|\hat{\pi}_n^\text{Ind}(\vartheta) - \pi(\vartheta)\| \to 0\) \(\mathbb{P}\)-a.s. holds then it is possible to show on a similar way that the indirect estimator \((\hat{\vartheta}_n^\text{Ind})_{n \in \mathbb{N}}\) is even strongly resistant.

As already mentioned, one could also define qualitative robustness of a sequence of estimators by demanding that the distribution of the estimator does not change too much when the data is changed slightly. To make this notion explicit, we first define a pseudometric for measures on metric spaces.

**Definition 6.5.** For a metric space \((M, d)\) with Borel sets \(\mathcal{B}(M)\), the Prokhorov distance \(\pi_\vartheta\) between two measures \(\mu, \nu\) on \(\mathcal{B}(M)\) with respect to \(d\) is defined as
\[
\pi_\vartheta(\mu, \nu) := \inf \{\varepsilon > 0 : \mu(A) \leq \nu(\{x \in M : d(x, A) < \varepsilon\}) + \varepsilon \forall A \in \mathcal{B}(M)\}.
\]
This pseudometric is a key component of the definition of qualitative robustness.

**Definition 6.6.** Let \(d_\Theta\) be a metric on \(\Theta\) and let \(\rho_n\) be a pseudometric on \(\mathcal{P}(\mathbb{R}^n), n \in \mathbb{N}\). For \(\mathbb{P} \in \mathcal{P}(\mathbb{R}^n)\) denote by \(\mathbb{P}_n\) the \(n\)-th order marginal of \(\mathbb{P}\). Finally, \(\mathbb{P}_{\hat{\vartheta}_n}\) is the distribution of the estimator \(\hat{\vartheta}_n\) under \(\mathbb{P}_n\). Then the sequence of estimators \((\hat{\vartheta}_n)_{n \in \mathbb{N}}\) is called \(\rho_n\)-robust at \(\mathbb{P}\) if for every \(\varepsilon > 0\) there exists a \(\delta > 0\) such that for every \(Q_n \in \mathcal{P}(\mathbb{R}^n)\) with \(\rho_n(\mathbb{P}_n, Q_n) < \delta\):
\[
\pi_{\rho_n}(\mathbb{P}_{\hat{\vartheta}_n}, Q_{\hat{\vartheta}_n}) \leq \varepsilon.
\]

As shown in Boente et al. (1987, Theorem 3.1), this is a direct generalization of the definition of \(\pi\)-robustness given by Hampel (1971) for i.i.d. processes.

**Theorem 6.7.** On \(\mathbb{R}^n\) we define the metric
\[
d_n(x^n, y^n) = \inf \{\varepsilon : \# \{i: |x_i - y_i| \geq \varepsilon\} / n \leq \varepsilon\}
\]
and use the Prokhorov distance with respect to \(d_n\),
\[
\pi_{d_n}(\mathbb{P}_n, Q_n) = \inf \{\varepsilon > 0 : \mathbb{P}_n(A) \leq Q_n(\{x^n \in \mathbb{R}^n : d_n(x^n, A) < \varepsilon\}) + \varepsilon \forall A \in \mathcal{B}(\mathbb{R}^n)\}
\]
on \(\mathcal{B}(\mathbb{R}^n)\). Then the sequence of estimators \((\hat{\vartheta}_n^\text{Ind})_{n \in \mathbb{N}}\) is \(\pi_{d_n}\)-robust at \(\mathbb{P}_{Y(\delta)}\).

Regarding the metric \(d_n\) two points are close if all coordinates except a small fraction are close.
We continue our investigation of robustness of \( \hat{\theta}_{\text{ind}}^{(n)} \) with the study of the influence functional of the indirect estimator. 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 the 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 contamination 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)). We work with their definition in the following.

In addition to the assumptions given at the beginning of Section 6, we assume throughout this section that there exists a unique solution \( \pi_{\text{GM}}(\theta_0^Y) \) of (4.2) for \( (Y_{mh}^Y)_{m \in \mathbb{Z}} = (Y_{mh}^Y(\theta_0))_{m \in \mathbb{Z}} \) for any \( 0 \leq \gamma \leq 1 \), and \( J(\theta_0) \) is non-singular. We denote the probability measure associated to the distribution of \( (Y_{mh}^Y)_{m \in \mathbb{Z}} \) by \( \mathbb{P}_{Y(h)}^Y \) for \( 0 \leq \gamma \leq 1 \). Note that \( \gamma = 0 \) corresponds to the case where there are no outliers, i.e., we can observe the nominal process without error and then write \( \mathbb{P}_{Y(h)}^0 = \mathbb{P}_{Y(h)}^\gamma \). Similarly \( \mathbb{P}_{Z} \) is the distribution of \((Z_m)_{m \in \mathbb{Z}}\) and \( \mathbb{P}_{W} \) is the distribution of \((W_m)_{m \in \mathbb{Z}}\). We write \( \{\mathbb{P}_{Y(h)}^Y\} := \{\mathbb{P}_{Y(h)}^{\gamma}, 0 \leq \gamma \leq 1\} \subset \mathcal{P}(\mathbb{R}^\infty) \) and introduce the statistical functional

\[
T_{\text{GM}} : \{\mathbb{P}_{Y(h)}^Y\} \to \Pi \quad \text{as} \quad \mathbb{P}_{Y(h)}^\gamma \mapsto \pi_{\text{GM}}(\theta_0^Y).
\]

Then, the definition of the influence functional for the GM-estimator is

\[
IF_{\text{GM}}(\mathbb{P}_{Z}, \{\mathbb{P}_{Y(h)}^Y\}) := \lim_{\gamma \to 0} \frac{T_{\text{GM}}(\mathbb{P}_{Y(h)}^Y) - T_{\text{GM}}(\mathbb{P}_{Y(h)}^\gamma)}{\gamma} = \lim_{\gamma \to 0} \frac{\pi_{\text{GM}}(\theta_0^Y) - \pi(\theta_0)}{\gamma} \quad (6.9)
\]

whenever this limit is well-defined. Note that the influence functional depends on the whole “arc” of contaminated measures \( \{\mathbb{P}_{Y(h)}^{\gamma}, 0 \leq \gamma \leq 1\} \). This is the most important difference to the definition used by Künsch (1984), because in that paper the approximation \( \mathbb{P}_{Y(h)}^Y = (1 - \gamma)\mathbb{P}_{Y(h)} + \gamma \mathbb{P} \) for some fixed \( \mathbb{P} \in \mathcal{P}(\mathbb{R}^\infty) \) is used (Künsch (1984, Eq. (1.11))). The influence functional measures the effect of an infinitesimal contamination of the true process by the process \((Z_m)\) on the asymptotic estimate defined via the functional \( T_{\text{GM}} \).

In a similar vein, we can define the influence functional for the estimation of the parameter \( \theta_0 \) of our CARMA process. Analogous to \( T_{\text{GM}} \), we first define a suitable statistical functional

\[
T_{\text{ind}} : \{\mathbb{P}_{Y(h)}^Y\} \to \Theta \quad \text{as} \quad \mathbb{P}_{Y(h)}^Y \mapsto \tilde{\theta}_{\text{ind}}(\gamma) := \arg\min_{\tilde{\theta} \in \Theta} [\pi(\tilde{\theta}) - \pi_{\text{GM}}(\theta_0^Y)^T \Omega [\pi(\tilde{\theta}) - \pi(\theta_0^Y)].
\]

This is the analog of \( \theta_0 = \arg\min_{\theta \in \Theta} [\pi(\theta) - \pi(\theta_0)]^T \Omega [\pi(\theta) - \pi(\theta_0)] \) in the uncontaminated case (cf. (5.1)). With this and \( \tilde{\theta}_{\text{ind}}(0) = \theta_0 \) due to \( \pi_{\text{GM}}(\theta_0) = \pi(\theta_0) \) the definition of the influence
functional of the indirect estimator is
\[
IF_{\text{Ind}}(P_Z, \{P_{Y(h)}^{T}\}) = \lim_{\gamma \to 0} \frac{T_{\text{Ind}}(P_{Y(h)}^{T}) - T_{\text{Ind}}(P_{Y(h)})}{\gamma} = \lim_{\gamma \to 0} \frac{\partial T_{\text{Ind}}(\gamma) - \partial_0}{\gamma}.
\]

We are interested in properties of this functional, in particular, if it is bounded. Boundedness of the influence functional implies that the estimate arising from the contaminated process cannot move too far away from the one in the uncontaminated case if the rate of contamination is infinitesimal. This property is well-known for the influence functional for GM-estimators of AR processes. Since these estimators are an integral building block of the indirect estimator, one can hope that it carries over to our scenario and indeed it does, since the two functionals are proportional.

**Proposition 6.8.** Assume that \(\nabla_\theta \pi(\theta_0)\) has full column rank \(N(\Theta)\). Then if \(IF_{\text{GM}}\) exists then \(IF_{\text{Ind}}\) exists and
\[
IF_{\text{Ind}}(P_Z, \{P_{Y(h)}^{T}\}) = \mathcal{H}(T_{\text{Ind}}(P_{Y(h)}^{T}))IF_{\text{GM}}(P_Z, \{P_{Y(h)}^{T}\}),
\]
where
\[
\mathcal{H}(T_{\text{Ind}}(P_{Y(h)}^{T})) = \mathcal{H}(\theta_0) = [\nabla_\theta \pi(\theta_0)^T \Omega \nabla_\theta \pi(\theta_0)]^{-1} \nabla_\theta \pi(\theta_0)^T \Omega.
\]

**Proof.** This follows from de Luna and Genton (2000, Theorem 1) as special case.

From this theorem we see that the question of the boundedness of the influence functional for the indirect estimator of a discretely sampled CARMA process reduces to the question of the boundedness of the influence functional for the GM-estimator of the auxiliary AR representation of the sampled CARMA process.

**Theorem 6.9.** Let the additive outlier model be given and let [Assumption D] hold. Then there exists a constant \(K > 0\) such that
\[
\|IF_{\text{GM}}(P_W, \{P_{Y(h)}^{T}\})\| \leq 2(r + 1)K \| J_{\text{GM}}(\theta_0)^{-1} \|.
\]

**Proof.** The plan is to apply Martin and Yohai (1986, Theorem 4.3). Therefore, we have to check that (Martin and Yohai, 1986, Eq. (4.6)) hold. Sufficient conditions for this equation are given in Martin and Yohai (1986, Theorem 4.2) which are obviously satisfied in our case due to (E.2) and due to the fact that \(\pi^{GM}(\theta_0^T)\) depends only on the distribution of the finite random vector \((Y_h^T, \ldots, Y_{(r+1)h}^T)\). For the same reasons and with our assumption that \(J(\theta_0)\) is non-singular the other conditions in Martin and Yohai (1986, Theorem 4.3) are satisfied as well.

As well due to Martin and Yohai (1986, Theorem 4.3) it is possible to give an explicit (but not a very handy) representation of \(IF_{\text{GM}}(P_W, \{P_{Y(h)}^{T}\})\) and hence, though Proposition 6.8 for \(IF_{\text{Ind}}(P_W, \{P_{Y(h)}^{T}\})\).

### 6.3 The breakdown point

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.
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 $\hat{\pi}_n(\theta)$ is applied to a simulated outlier–free sample.

7 Simulation study

We simulate a CARMA process on the interval $[0,1000]$ and choose a sampling distance of $h = 1$, resulting in $n = 1000$ observations of the discrete–time process. The simulated processes are driven either by a standard Brownian motion or by a univariate NIG Lévy process. For the NIG Lévy process we use the parameters $\alpha = 3$, $\beta = 1$, $\Delta = 1$, $\delta = 2.5145$ 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 Brownian motion case.

The indirect estimator is defined as in Section 5. We take $\hat{\pi}_n$ as GM-estimator $\hat{\pi}_n^{GM}(\theta_0)$ using the S–Plus software which provides a pre–built function ar.gm for applying GM-estimators to AR processes. This function uses a Mallows estimator as in Example 4.3(a). The weight function $w(y)$ is the Tukey bisquare function from Example 4.3(b) applied to $\|y\|$, for the function $\psi(u)$ the user can choose between the Huber $\psi_k$–function and the bisquare function. The function is implemented as an iterative least squares procedure as described by Martin (1980, p. 231ff.) and therefore also allows to use first some iteration steps with the Huber $\psi_k$–function and then some steps with the bisquare function. As advocated by Martin (1980), our experience is that doing 6 iterations using the Huber function and then 2 steps with the bisquare function works well. In our experiments we use $k = 4$ for the tuning constant of the $\psi_k$–function. In general, we set $s = 75$ to obtain the simulation–based observations $Y_{S}^{sn}(\theta) = (Y_{h}^{S}(\theta),\ldots,Y_{snh}^{S}(\theta))$ in the simulation part of the indirect procedure. The type of Lévy process used for the simulation is of the same type as the Lévy process driving the CARMA process. For the estimator $\hat{\pi}_n(\theta)$ we apply the least squares estimator and as weighting matrix $\Omega$ we take the identity matrix for convenience reasons. Some experiments in which we first estimated the asymptotic covariance matrix of the GM-estimator by the empirical covariance matrix of a suitable number of independent realizations of $\hat{\pi}_n$ and set $\Omega$ to be the inverse of that estimate did not significantly affect the procedure positively or negatively so that the use of the convenient identity matrix seems justified. For the outlier model we choose the process $(V_m)_{m \in \mathbb{Z}}$ as i.i.d. Bernoulli random variables where $P(V_1 = 1) = \gamma$ varies. In all but one of the studies, the process $(Z_m)_{m \in \mathbb{Z}}$ is chosen to be constant $Z_m = \xi$ for $m \in \mathbb{Z}$. We use varying values of $\xi$. In the experiment where $(Z_m)_{m \in \mathbb{Z}}$ has a different structure, this will be mentioned explicitly. In each experiment, we calculate the indirect estimator and, for comparison purposes, the QMLE as defined in Schlemm and Stelzer (2012) in 50 independent replications and report on the average estimated value, the bias relative to the true parameters and the empirical variance of the parameter estimates.
7.1 CARMA processes driven by a Brownian motion

In a first experiment, we use as true process a CARMA(1,0) process with parameter \( \theta_0^{(1)} = -2 \). This process is of particular interest because its discretely sampled version admits an AR(1) representation. For this reason, one would expect the procedure to work very well here as the auxiliary representation is actually exact. Naturally, we use \( r = 1 \) in this case. We consider three different scenarios of outlier contamination. In the first case, we set \( \xi = 5 \) and \( \gamma = 0.1 \). In the second, we set \( \xi = 10 \) and \( \gamma = 0.1 \), while for the last one we choose \( \xi = 5 \) and \( \gamma = 0.15 \). Note that already \( \xi = 5 \) represents quite large outliers, since for a sample path in this situation we typically observe that the values of the discretely sampled process lie between \(-3.5\) and \(3.5\). The results of the simulation studies are given in Table 1.

<table>
<thead>
<tr>
<th>( \xi = 5, \gamma = 0.1 )</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_1 )</td>
<td>-2.4497</td>
<td>-0.4497</td>
</tr>
<tr>
<td>( \xi = 10, \gamma = 0.1 )</td>
<td>MLE</td>
<td>Indirect</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>-5.9097</td>
<td>-3.9097</td>
</tr>
<tr>
<td>( \xi = 5, \gamma = 0.15 )</td>
<td>MLE</td>
<td>Indirect</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>-2.7873</td>
<td>-0.7873</td>
</tr>
</tbody>
</table>

Table 1: Results for \( \theta_0^{(1)} \).

As we can see, already in the case \( \xi = 5 \) and \( \gamma = 0.1 \), the indirect estimator performs vastly better than the QMLE giving a much less biased estimate at a similar variance. In the situation where \( \xi \) is increased to 10 the QMLE has lost all its information about the true parameter and provides no useful estimate anymore. On the other hand, the indirect estimator stays close to the true value (we explain the even smaller bias in comparison to the situation with \( \xi = 5 \) as caused by the relatively small number of 50 iterations. Of course one should not expect the bias to systematically decrease when \( \xi \) increases), while the variance has increased only slightly. Increasing \( \gamma \) to 0.15 but keeping \( \xi = 5 \) shows that both estimators perform worse than in the situation with \( \gamma = 0.1 \), which is to be expected. But once again, the indirect estimator deals much better with the higher outlier percentage than the QMLE. Comparing these studies, we see that for the indirect estimator the percentage of outliers has a bigger effect on the estimates than the actual size of the outliers.

In our next 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

\[
\theta_0^{(2)} = (-1 \quad -2 \quad -2 \quad 0 \quad 1).
\]

For this process, we choose \( r = 5 \), which is also the minimum order of the auxiliary AR representation to satisfy Assumption B. We do four experiments in this setup. We estimate \( \theta_0^{(2)} \) for each of the following contamination configurations: \( \xi = 5 \) and \( \gamma = 0.1 \), \( \xi = 10 \) and \( \gamma = 0.1 \), \( \xi = 5 \) and \( \gamma = 1/6 \), and \( \xi = 5 \) and \( \gamma = 0.25 \). Remember that in this situation, the breakdown point has an upper bound of \( 1/6 \) since we have \( r = 5 \). Hence, \( \gamma = 0.25 \) lies above the breakdown point and we would expect to encounter problems in the estimation procedure, while for \( \gamma \leq 1/6 \) these problems should not occur.

31
We will see that this is indeed the case. The results of the four experiments are given in Table 2.

<table>
<thead>
<tr>
<th>ξ = 5, γ = 0.1</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ₁</td>
<td>−0.6876</td>
<td>0.3124</td>
</tr>
<tr>
<td>φ₂</td>
<td>−2.6307</td>
<td>−0.6307</td>
</tr>
<tr>
<td>φ₃</td>
<td>−3.3831</td>
<td>−1.3831</td>
</tr>
<tr>
<td>φ₄</td>
<td>2.5467</td>
<td>2.5467</td>
</tr>
<tr>
<td>φ₅</td>
<td>0.5621</td>
<td>−0.4379</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ξ = 10, γ = 0.1</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ₁</td>
<td>−0.4614</td>
<td>0.5386</td>
</tr>
<tr>
<td>φ₂</td>
<td>−1.4955</td>
<td>0.5045</td>
</tr>
<tr>
<td>φ₃</td>
<td>−1.8424</td>
<td>0.1576</td>
</tr>
<tr>
<td>φ₄</td>
<td>3.3178</td>
<td>3.3178</td>
</tr>
<tr>
<td>φ₅</td>
<td>2.6317</td>
<td>1.6317</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ξ = 5, γ = 1/6</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ₁</td>
<td>−0.4445</td>
<td>0.5555</td>
</tr>
<tr>
<td>φ₂</td>
<td>−2.3450</td>
<td>−0.3450</td>
</tr>
<tr>
<td>φ₃</td>
<td>−3.5119</td>
<td>−1.5119</td>
</tr>
<tr>
<td>φ₄</td>
<td>3.1446</td>
<td>3.1446</td>
</tr>
<tr>
<td>φ₅</td>
<td>0.7903</td>
<td>−0.2097</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ξ = 5, γ = 0.25</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ₁</td>
<td>−0.1022</td>
<td>0.8978</td>
</tr>
<tr>
<td>φ₂</td>
<td>−1.5663</td>
<td>0.4337</td>
</tr>
<tr>
<td>φ₃</td>
<td>−4.2751</td>
<td>−2.2751</td>
</tr>
<tr>
<td>φ₄</td>
<td>4.2080</td>
<td>4.2080</td>
</tr>
<tr>
<td>φ₅</td>
<td>0.8238</td>
<td>−0.1762</td>
</tr>
</tbody>
</table>

Table 2: Results for θ_0^{(2)}.

For the first two experiments where γ = 0.1, we immediately recognize the maximum likelihood estimate as basically worthless being severely biased and very far from the true parameter value. 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. On the other hand, the indirect estimator is still very close to the true parameter value in all components including the zero component. Increasing ξ to 10 while keeping γ = 0.1 results in a slightly worse performance of the indirect estimator. However, the increase in the bias is not very substantial and the estimates are still reasonably close to the true values. This reaffirms that the indirect estimation procedure works in practical scenarios and the results in the former experiments were not due to the use of the CARMA(1,0) process.

The increase of γ from 0.1 to 1/6 also affects the performance of the indirect estimator. For all components of θ_0^{(2)}, except the first one, the bias and the variance of the indirect estimator increase. However, the loss in quality of the indirect estimator is manageable and the calculated estimates still
resemble the true parameter. This means that even at the breakdown point of $1/6$, the performance of the indirect estimator is satisfying, although of course not as good as for lower contamination probabilities.

The situation is vastly different in the experiment with $\gamma = 0.25$ where $\gamma$ is above the breakdown point. Here, we see not surprisingly that the indirect estimator, too, gives estimates which are severely biased and are quite far away from the true parameters. We also observe that the numerical procedure used to obtain the parameter estimates quite often fails to deliver a result at all because the algorithm terminates with an error. The error occurs when the estimated value of $\theta_0$ is not an element of $\Theta$ anymore. The results in the table are averaged over experiments in which the algorithm did deliver a result, the failed attempts were discarded. The ratio of successful to unsuccessful experiments was roughly equal to 1:2, i.e., the algorithm failed about twice as often as it succeeded. In this sense, we can say that the estimator has broken down: for a given outlier–contaminated sample, it either does not return an admissible estimate at all, or, if it does, the estimate is far away from the true parameter. The latter statement is also evident from the fact that the variances of the indirect estimates are far smaller in this case than in the other experiments which intuitively means that the algorithm typically returns very similar bad estimates if it returns a result at all.

Lastly, we use a CARMA(2,1) process with true parameter

$$\vartheta_0^{(3)} = (-0.5, -1, -2).$$

We report on two simulation studies, which are a bit different than the ones conducted before. In the first study, we take $\gamma = 0$ and compare the performance of the indirect estimator to that of the QMLE in the uncontaminated case. In the next study, the process $(Z_m)_{m \in \mathbb{Z}}$ is a sequence of i.i.d. random variables with $P(Z_1 = \xi) = P(Z_1 = -\xi) = \frac{1}{2}$, i.e., every time an outlier appears the sign is chosen randomly with equal probability. This change in model serves to study the performance of the procedure in circumstances that are a bit more complicated than the simple additive outlier model with fixed size $\xi$. We choose $\gamma = 0.1$ and $\xi = 10$ in this study. In both studies $r = 3$. The results are given in Table 3.

<table>
<thead>
<tr>
<th>$\gamma = 0$</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vartheta_1$</td>
<td>Mean</td>
<td>-0.5112</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.0071</td>
</tr>
<tr>
<td>$\vartheta_2$</td>
<td>Mean</td>
<td>-0.9969</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.0056</td>
</tr>
<tr>
<td>$\vartheta_3$</td>
<td>Mean</td>
<td>-2.0427</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.0246</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\gamma = 0.1$</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vartheta_1$</td>
<td>Mean</td>
<td>-0.3796</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.0585</td>
</tr>
<tr>
<td>$\vartheta_2$</td>
<td>Mean</td>
<td>-2.0699</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.2484</td>
</tr>
<tr>
<td>$\vartheta_3$</td>
<td>Mean</td>
<td>-4.0399</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.1408</td>
</tr>
</tbody>
</table>

Table 3: Results for $\vartheta_0^{(3)}$.

In the situation without outliers, both estimators are very close to the true parameter values. The differences in the bias and the variance are not systematic and probably due to the approximations in the numerical procedure and only 50 iterations. We see that both estimators can be used to achieve satisfying results. For the parameters $\vartheta_0^{(1)}$ and $\vartheta_0^{(2)}$ we conducted the same study (i.e. in the outlier–free
situation) and observed basically the same results. Using the indirect estimator also yields satisfying results under the more complicated outlier model. This is not surprising since the GM-estimator, which controls the robustness properties of \( \hat{\vartheta}_{\text{Ind}} \), is per construction not sensitive to the sign of the outlier, only to the absolute size.

7.2 CARMA processes driven by a NIG Lévy process

In the following experiments, the driving Brownian motion is replaced by the NIG Lévy process. We repeat some of the experiments of Table 1 and Table 2 using the same outlier configurations and auxiliary AR orders as in those experiments. The results are given in Table 4 and Table 5.

<table>
<thead>
<tr>
<th>ζ = 5, γ = 0.1</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Bias</td>
<td>Variance</td>
</tr>
<tr>
<td>( \varphi_1 )</td>
<td>-0.6669</td>
<td>0.3331</td>
</tr>
<tr>
<td>( \varphi_2 )</td>
<td>-2.5668</td>
<td>-0.5668</td>
</tr>
<tr>
<td>( \varphi_3 )</td>
<td>-3.3917</td>
<td>-1.3917</td>
</tr>
<tr>
<td>( \varphi_4 )</td>
<td>0.5077</td>
<td>-0.4923</td>
</tr>
<tr>
<td>( \varphi_5 )</td>
<td>2.5503</td>
<td>2.5503</td>
</tr>
</tbody>
</table>

Table 4: Results for \( \vartheta^{(1)}_0 \), \( \xi = 5 \), \( \gamma = 0.1 \), driving NIG Lévy process

<table>
<thead>
<tr>
<th>ζ = 10, γ = 0.1</th>
<th>MLE</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Bias</td>
<td>Variance</td>
</tr>
<tr>
<td>( \varphi_1 )</td>
<td>-0.4691</td>
<td>0.5309</td>
</tr>
<tr>
<td>( \varphi_2 )</td>
<td>-1.5286</td>
<td>0.4714</td>
</tr>
<tr>
<td>( \varphi_3 )</td>
<td>-1.8365</td>
<td>0.1635</td>
</tr>
<tr>
<td>( \varphi_4 )</td>
<td>2.6651</td>
<td>1.6651</td>
</tr>
<tr>
<td>( \varphi_5 )</td>
<td>3.3668</td>
<td>3.3668</td>
</tr>
</tbody>
</table>

Table 5: Results for \( \vartheta^{(2)}_0 \), \( \gamma = 0.1 \), driving NIG Lévy process

We see that the results do not substantially differ from those in the Brownian motion case. In all three experiments with the NIG Lévy process, the QMLE, just like in the Brownian motion case, ceases to be a meaningful estimator in the presence of outliers in the data. The indirect estimator, on the other hand, continues to provide good estimates.

8 Conclusion

In this paper 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 CARMA parameters and the AR 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 \( \hat{\pi}_n = \hat{\pi}_n^{GM}(\theta_0) \) and \( \hat{\pi}_n^S(\theta) = \hat{\pi}_n^{LS}(\theta) \). Moreover, we investigated different robustness properties and showed that the indirect estimator is weakly resistant, \( \pi_{dc} \)-robust and 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. As soon as outliers are present the use of the indirect estimator instead of the QMLE seems advisable, since there was no situation in which the performance of the QMLE came close to being satisfying in contrast to the indirect estimator. Of course, it is clear that the indirect estimator has its bounds as well. We especially saw that both an increase of \( \gamma \), the proportion of outliers, and the size of the outliers affect the performance. Increasing \( \gamma \) too far eventually causes the estimator to break down because we get above the breakdown point. Unfortunately, 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 Appendix: Proof of Lemma 5.4

Before we state the proof we require some auxiliary results.

**Lemma A.1.** For any \( i, j, l \in \{1, \ldots, N(\Theta)\} \) the following conditions hold:

(a) \( \sup_{\theta \in \Theta} |f_0(u)| \leq Ce^{-pu} \) and \( \int_0^\infty \sup_{\theta \in \Theta} f_0(u)^2 \, du < \infty. \)

(b) \( \sup_{\theta \in \Theta} \left| \frac{\partial^2 f_0(u)}{\partial \theta_i \partial \theta_j} \right| \leq C u^2 e^{-pu} \) and \( \int_0^\infty \sup_{\theta \in \Theta} \left( \frac{\partial^2 f_0(u)}{\partial \theta_i \partial \theta_j} \right)^2 \, du < \infty. \)

(c) \( \sup_{\theta \in \Theta} \left| \frac{\partial^3 f_0(u)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right| \leq C u^3 e^{-pu} \) and \( \int_0^\infty \sup_{\theta \in \Theta} \left( \frac{\partial^3 f_0(u)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right)^2 \, du < \infty. \)

(d) \( \sup_{\theta \in \Theta} \left| \frac{\partial^3 f_0(u)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right| \leq C u^3 e^{-pu} \) and \( \int_0^\infty \sup_{\theta \in \Theta} \left( \frac{\partial^3 f_0(u)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right)^2 \, du < \infty. \)

**Proof.** (a) Due to Remark 2.1(iii) we have that \( \sup_{\theta \in \Theta} \| e^{A_\theta u} \| \leq Ce^{-pu} \) and hence, \( \sup_{\theta \in \Theta} |f_0(u)| \leq \sup_{\theta \in \Theta} |c_\theta| \sup_{\theta \in \Theta} \| e^{A_\theta u} \| \leq Ce^{-pu} \) using the continuity of \( c_\theta \) on the compact set \( \Theta \). Finally, \( \int_0^\infty \sup_{\theta \in \Theta} f_0(u)^2 \, du < \infty. \)

(b) A consequence of [Wilcox (1967)] is that

\[
\frac{\partial e^{A_\theta u}}{\partial \theta_j} = \int_0^u e^{A_\theta (u-s)} \left( \frac{\partial}{\partial \theta_j} A_\theta \right) e^{A_\theta s} \, ds,
\]

and hence, \( \sup_{\theta \in \Theta} \| \frac{\partial e^{A_\theta u}}{\partial \theta_j} \| \leq C u e^{-pu} \). From this and

\[
\frac{\partial f_0(u)}{\partial \theta_j} = \left( \frac{\partial c_\theta}{\partial \theta_j} \right) e^{A_\theta u} e_p + c_\theta \left( \frac{\partial e^{A_\theta u}}{\partial \theta_j} \right) e_p
\]

we receive

\[
\int_0^\infty \sup_{\theta \in \Theta} \left( \frac{\partial f_0(u)}{\partial \theta_j} \right)^2 \, du \leq C \int_0^\infty u^2 e^{-2pu} \, du < \infty.
\]

(c,d) can be proven similarly to (a) and (b).
Lemma A.2. Let \((L_t)_{t \in \mathbb{R}}\) be a Lévy process with \(\mathbb{E}|L_1|^{2N^*} < \infty\) for some \(N^* \in \mathbb{N}\) and \(\mathbb{E}(L_1) = 0\). Furthermore, let \(\phi : \mathbb{R} \to \mathbb{R}\) be a measurable map with \(\phi \in \bigcap_{j=1}^{N^*} L^2_j(\mathbb{R})\). Then \(\mathbb{E} \left| \int_{-\infty}^{\infty} \phi(u) \, dL_u \right|^{2N^*} < \infty\) and there exist finite constants \(c_{j_1, \ldots, j_k}\) such that

\[
\mathbb{E} \left( \int_{-\infty}^{\infty} \phi(u) \, dL_u \right)^{2N^*} = \sum_{k=1}^{N^*} \sum_{j_1 + \cdots + j_k = N^*} \sum_{j_1, \ldots, j_k \in \mathcal{C}_0} c_{j_1, \ldots, j_k} \left( \int_{-\infty}^{\infty} \phi^{2j_1}(u) \, du \right) \cdots \left( \int_{-\infty}^{\infty} \phi^{2j_k}(u) \, du \right).
\]

Proof. For \(N^* = 2\) the result was already derived in Cohen and Lindner (2013, Lemma 3.2), the proof for general \(N^*\) uses the same ideas. Let \(\nu\) be the Lévy measure of \((L_t)_{t \in \mathbb{R}}\) and \(\nu\) its Gaussian parameter. Define

\[
\psi(x) = -\frac{1}{2} V^2 x^2 \int_{-\infty}^{\infty} \phi^2(u) \, du + \int_{-\infty}^{\infty} \left[ e^{ix\phi(u)x} - 1 - i\phi(u)x \right] \nu(dx) \, ds,
\]

and write \(\psi^{(i)}(s)\) for the \(i\)-th derivative of \(\psi(s)\). Due to the Lévy-Khintchine formula we get

\[
\xi(s) := \mathbb{E} \left( \exp \left( is \int_{-\infty}^{\infty} \phi(u) \, dL_u \right) \right) = \exp(\psi(s)).
\]

Then \(\mathbb{E} \left( \int_{-\infty}^{\infty} \phi(u) \, dL_u \right)^{2N^*}\) is obtained as \((2N^*)\)-th derivative of \(\xi(s)\) at \(s = 0\) times \((-1)^{N^*}\). Straightforward calculations yield that the \((2N^*)\)-th derivative \(\xi^{(2N^*)}(s)\) of \(\xi(s)\) has the form

\[
\xi^{(2N^*)}(s) = \sum_{k=1}^{2N^*} \sum_{j_1 + \cdots + j_k = 2N^*} \hat{c}_{j_1, \ldots, j_k} \psi^{(2j_1)}(s) \cdots \psi^{(2j_k)}(s) \exp(\psi(s)).
\]

Plugging in for \(s = 0\) and taking into account that \(\psi(0) = 1\) and \(\psi^{(1)}(0) = 0\) gives

\[
\mathbb{E} \left( \int_{-\infty}^{\infty} \phi(u) \, dL_u \right)^{2N^*} = (-1)^{N^*} \xi^{(2N^*)}(0) = \sum_{k=1}^{N^*} \sum_{j_1 + \cdots + j_k = N^*} \sum_{j_1, \ldots, j_k \in \mathcal{C}_0} c_{j_1, \ldots, j_k} \psi^{(2j_1)}(0) \cdots \psi^{(2j_k)}(0).
\]

Finally, with

\[
\psi^{(2)}(0) = \left[ -V^2 - \int_{-\infty}^{\infty} x^2 \, \nu(dx) \right] \left[ \int_{-\infty}^{\infty} \phi^2(u) \, du \right],
\]

\[
\psi^{(2i)}(0) = (-1)^i \left[ \int_{-\infty}^{\infty} x^{2i} \, \nu(dx) \right] \left[ \int_{-\infty}^{\infty} \phi^{2i}(u) \, du \right], \quad i \geq 2,
\]

we obtain the result.

Lemma A.3. Let \(N^* \in \mathbb{N}\) be such that \(2N^*>N(\Theta)\) and \(\mathbb{E}|L_1|^{2N^*} < \infty\). For any \(i, j, k \in \{1, \ldots, N(\Theta)\}\) the following maps are \(\mathbb{P}\)-a.s. Hölder-continuous of order \(\gamma \in [0, 1 - N(\Theta)/(2N^*)] \):

(a) \(\vartheta \mapsto \int_{0}^{\infty} f_\vartheta(u) \, dL_u =: Z(\vartheta)\),

(b) \(\vartheta \mapsto \int_{0}^{\infty} \frac{\partial}{\partial u} f_\vartheta(u) \, dL_u =: Z^{(j)}(\vartheta)\),

(c) \(\vartheta \mapsto \int_{0}^{\infty} \frac{\partial}{\partial u \partial u} f_\vartheta(u) \, dL_u =: Z^{(j, j)}(\vartheta)\).
Moreover, \( \mathbb{E}(\sup_{\vartheta \in \Theta} |Z(\vartheta)|^{2N^*}) < \infty \) and for \( U := \sup_{\vartheta_1, \vartheta_2 \in \Theta; \vartheta_1 - \vartheta_2 < C} \frac{|Z(\vartheta_1) - Z(\vartheta_2)|}{\|\vartheta_1 - \vartheta_2\|} \) we have \( \mathbb{E}U^{2N^*} < \infty \). The same is true for \( Z^{(j)}(\vartheta) \) and \( Z^{(j)}(\vartheta) \).

**Proof.**

(a) Let \( \vartheta_1, \vartheta_2 \in \Theta \) and define \( \phi(u) := f_{\vartheta_1}(u) - f_{\vartheta_2}(u) \). Due to a Taylor expansion we obtain

\[
\phi(u) = f_{\vartheta_1}(u) - f_{\vartheta_2}(u) = \nabla \vartheta f_{\vartheta(u)}(\vartheta_1 - \vartheta_2)
\]

for some \( \tilde{\vartheta}(u) \in \Theta \) with \( \|\tilde{\vartheta}(u) - \vartheta_2\| \leq \|\vartheta_1 - \vartheta_2\| \). Hence, we receive by Lemma A.1

\[
|\phi(u)| \leq \sup_{\vartheta \in \Theta} \|\nabla \vartheta f_{\vartheta(u)}\| ||\vartheta_1 - \vartheta_2|| \leq C e^{-p\vartheta} ||\vartheta_1 - \vartheta_2||.
\]

Plugging this into Lemma A.2 gives

\[
\mathbb{E}(Z(\vartheta_1) - Z(\vartheta_2))^{2N^*} \\
\leq C \sum_{k=1}^{N^*} \sum_{j_1, \ldots, j_k} \left| {c_{j_1, \ldots, j_k}} \right| \left( \int_{-\infty}^{\infty} ||\vartheta_1 - \vartheta_2||^{2j_1} u^{2j_2} e^{-2j_3} \, du \right) \cdots \left( \int_{-\infty}^{\infty} ||\vartheta_1 - \vartheta_2||^{2j_1} u^{2j_2} e^{-2j_3} \, du \right)
\leq C ||\vartheta_1 - \vartheta_2||^{2N^*}.
\]

Then, an application of the Hölder continuity and \( \mathbb{E}U^{2N^*} < \infty \). Since \( \Theta \) is compact, some straightforward calculations yield \( \mathbb{E}(\sup_{\vartheta \in \Theta} |Z(\vartheta)|^{2N^*}) < \infty \). The proofs of (b)-(c) are similarly to the proof of (a) and thus, skipped.

**Lemma A.4.** Let \( [a, b] \subseteq \mathbb{R} \) be a bounded interval, \( (L_t)_{t \in \mathbb{R}} \) be a Lévy process with finite second moments and let \( g : [a, b] \times \mathbb{R} \rightarrow \mathbb{R} \) be differentiable in the first component with derivative \( \frac{\partial g(\vartheta, u)}{\partial \vartheta} \). Moreover, assume

(a) \( \frac{\partial g(\vartheta, u)}{\partial \vartheta} \) is bounded \( \mathcal{B}([a, b]) \otimes \mathcal{B}([-u_1, u_2]) \)-measurable for all \( u_1, u_2 > 0 \).

(b) \( \lim_{M \to \infty} \sup_{\vartheta \in [a, b]} \int_{|u| > M} \left| \frac{\partial g(\vartheta, u)}{\partial \vartheta} \right| \, du = 0 \) and \( \lim_{|u| \to \infty} \sup_{\vartheta \in [a, b]} \left| \frac{\partial g(\vartheta, u)}{\partial \vartheta} \right| = 0 \).

(c) \( \vartheta \mapsto \int_{-\infty}^{\infty} g(\vartheta, u) \, dL_u \) is \( \mathbb{P} \)-a.s. continuous.

(d) \( \vartheta \mapsto \int_{-\infty}^{\infty} \frac{\partial g(\vartheta, u)}{\partial \vartheta} \, dL_u \) is \( \mathbb{P} \)-a.s. continuous.

Then, outside a \( \mathbb{P} \)-zero set, \( Z(\vartheta, \omega) := \int_{-\infty}^{\infty} g(\vartheta, u) \, dL_u(\omega) \) is continuous differentiable over the interval \( (a, b) \) and

\[
\frac{\partial}{\partial \vartheta} \int_{-\infty}^{\infty} g(\vartheta, u) \, dL_u(\omega) = \int_{-\infty}^{\infty} \frac{\partial g(\vartheta, u)}{\partial \vartheta} \, dL_u(\omega).
\]
Proof. An application of Fubini’s Theorem for Lévy-integrals (see Brockwell and Schlemm, 2013, Theorem 2.4)) gives that for \( \varrho \in [a,b] \)

\[
\int_a^b \int_{-\infty}^{\infty} \frac{\partial g(\varrho, u)}{\partial \varrho} dL_u d\varrho = \int_{-\infty}^{\infty} \int_a^b \frac{\partial g(\varrho, u)}{\partial \varrho} d\varrho dL_u \quad \mathbb{P}\text{-a.s.}
\]

The remaining of the proof follows the same line as Hutton and Nelson (1984, Theorem 2.2).

Proof of Lemma 5.4. A combination of Lemma A.1-Lemma A.4 result in the MA representation. A conclusion of the MA representation and Fuchs and Stelzer (2013, Theorem 3.5) is that the process is mixing and therefore, in particular, ergodic.

References


