INAUGURAL-DISSERTATION

zur Erlangung der Doktorwürde

der Naturwissenschaftlich-Mathematischen Gesamtfakultät

der Ruprecht-Karls-Universität Heidelberg

vorgelegt von

Stefanie Feiler
aus Backnang

Tag der mündlichen Prüfung: 7. Dezember 2005

Parameter Estimation
in Panels
of Intercorrelated Time Series

Gutachter: Prof. Dr. Rainer Dahlhaus
Prof. Dr. Enno Mammen

Abstract

We consider parameter estimation in panels of intercorrelated time series. By a factorisation of the conditional log-likelihood function we obtain a new estimator $\hat{a}_{n,T}$ for panels of intercorrelated autoregressive time series. We generalise this model to a factor model, where a single underlying background process is responsible for the common behaviour of the time series in the panel, and derive the corresponding conditional maximum likelihood estimators. Consistency and asymptotic normality are proved for the estimators in both models. It turns out that $\hat{a}_{n,T}$ is asymptotically equivalent to the estimator $\hat{a}_{HT}$ given in Hjellvik and Tjøstheim (1999a) if the number of time series in the panel tends to infinity. It is more efficient if only the length of the time series increases. Furthermore the mean squared errors of the dominant terms in the stochastic expansions of these estimators have the ratio $(n-1)/n$, which indicates that already the small sample bias of $\hat{a}_{n,T}$ is smaller than that of $\hat{a}_{HT}$. These properties are confirmed in the simulations.

The second part of the thesis is concerned with robust estimation in panels of autoregressive time series. We investigate three different approaches. Firstly we robustify the above estimators in a direct way. Furthermore we generalise the robust autocovariance estimator of Ma and Genton (2000) to the panel case. We define a panel breakdown point for time series in two ways depending on the type of outliers assumed and compute its value for the panel autocovariance estimator. The estimated autocovariances are then used for the robust parameter estimation. Finally we propose an outlier test based upon the phase space representation of the time series in the panel, which can be used for eliminating outliers from the data set before using a non-robust method of estimation. We derive the asymptotic distribution of the test statistic and define a robust version of the test. For comparison we include other estimators in the analysis. The performance of the proposed robust procedures is investigated in a simulation study.

For assessing the applicability of the above methods we analyse two sets of empirical data.

Kurzfassung

Die vorliegende Arbeit befasst sich mit Parameterschätzung in Panels interkorrelierter Zeitreihen. Durch eine Faktorisierung der bedingten Log-Likelihood-Funktion erhalten wir einen Schätzer $\hat{a}_{n,T}$ in Panels von interkorrelierten autoregressiven Zeitreihen. Dieses Modell wird zu einem Faktormodell verallgemeinert, in dem ein einzelner im Hintergrund ablaufender Prozess für das gemeinsame Verhalten der Zeitreihen im Panel verantwortlich ist. Hierfür entwickeln wir den zugehörigen Maximum-Likelihood-Schätzer. Für die Schätzer in beiden Modellen werden Konsistenz und asymptotische Normalität bewiesen. Es stellt sich heraus, dass $\hat{a}_{n,T}$ asymptotisch äquivalent zu dem Schätzer $\hat{a}_{HT}$ aus Hjellvik and Tjøstheim (1999a) ist, wenn die Zahl der Zeitreihen im Panel gegen Unendlich strebt. Wenn nur die Länge der Zeitreihen wächst, ist $\hat{a}_{n,T}$ effizienter. Zudem stehen die quadratischen Fehler der Hauptterme in der Entwicklung dieser Schätzer im Verhältnis $(n-1)/n$, was nahelegt, dass schon der Bias von $\hat{a}_{n,T}$ kleiner als derjenige von $\hat{a}_{HT}$ ist. Diese Eigenschaften werden durch die Simulationen bestätigt.
Der zweite Teil der Arbeit beschäftigt sich mit robuster Schätzung für Panels von autoregressiven Zeitreihen. Wir untersuchen drei unterschiedliche Ansätze. Zunächst robustifizieren wir die obigen Schätzer direkt. Des weiteren verallgemeinern wir den robusten Autokovarianzschätzer von Ma and Genton (2000) auf die Panel-Situation. Wir definieren einen Breakdown Point für Zeitreihen in Abhängigkeit von der Art der angenommenen Ausreißer und berechnen seinen Wert für den Panel-Autokovarianzschätzer. Die geschätzten Autokovarianzen werden dann für die robuste Parameterschätzung eingesetzt. Zuletzt schlagen wir einen Test für Ausreißer vor, der auf der Phasenraumdarstellung der Zeitreihen im Panel beruht. Dieser kann dazu verwandt werden, Ausreißer vor Anwendung einer nicht robusten Schätzmethode aus dem Datensatz zu entfernen. Wir bestimmen die asymptotische Verteilung der Teststatistik und definieren eine robuste Version des Tests. Zum Vergleich schließen wir weitere Schätzer in die Untersuchung mit ein. Das Verhalten der vorgeschlagenen robusten Verfahren wird in einer Simulationssstudie untersucht.
Um die Anwendbarkeit der obigen Methoden zu beurteilen, analysieren wir zwei Datensätze aus empirischen Studien.

Contents

Introduction

Panel data analysis has a wide range of applications, including econometrics, the social sciences, population dynamics or medical studies. In contrast to repeated measurements of cross-sectional data, panel methods are used for analysing repeated measurements on the same individuals. Here the term “individuals” stands for example for workers, countries, regions or patients. If a longer period is covered, the focus is often on the individual development, and the data is also called longitudinal data. However panels also may consist of a small number of large cross-sectional samples. Thus two kinds of asymptotic behaviour are of interest: increase in the length of the measurement period or increase in the size of the cross-section.

In the present thesis we consider panels of intercorrelated time series. This means that we assume the individual measurements to be serially correlated. Furthermore we do not exclude correlation across the panel. This double structure of correlation implies that standard methods for panel data analysis are not directly applicable. Nevertheless such models are of interest in practice: the initial motivation for this thesis came from a study conducted at the University Hospital of Heidelberg, Department of Internal and Psychosomatic Medicine. The aim of the study was to investigate the therapy process in a multimodal therapy for fibromyalgia syndrome patients. This is a chronic pain disease which is characterised by widespread pain and a reduced pain threshold. The therapy’s main focus is on helping patients to cope better in their daily life. Based on a bio-psycho-social approach, the distinct modules combine information, medication, physical therapy and a psychotherapeutic group therapy. Thus the question arises whether the therapy processes of different patients still can be modelled as independent when they participate in the same therapy group. More general, it can be asked whether undergoing the same treatment may already cause a dependency.

As data collected in an experiment always may contain outliers, we were furthermore led to investigate robust methods for panel data. One source of contamination lies of course in the recording of the data. However in the above study the data was collected by the patients themselves using a handheld computer. Thus the patients filled in their questionnaires without being able to see previous values. Moreover retrospective entries could easily be identified and excluded from the data. Therefore we focus on a second type of panel data outliers, namely those where one complete time series is generated by a different model. Such a situation arises for example if a patient has been wrongly assigned to a therapy group which otherwise is homogeneous.

Intercorrelation in panels of time series

A quite general linear dynamic model for panel observations is given in (Hsiao 1986, p. 71): Let $X_t^{(i)}$, $t = 1, \ldots, T$, $i = 1, \ldots, n$, be a panel of time series observations, where t denotes time and i the individual series in the panel. Then the observations are modelled as

$$X_t^{(i)} = \sum_{k=1}^{p} a_k \, X_{t-k}^{(i)} + \eta_t + \lambda_i + \beta' \, W_t^{(i)} + \varepsilon_t^{(i)} \, .$$

Here $W_t^{(i)}$ is a possibly vector series of explanatory variables. The random variable η_t denotes a cross-sectional effect influencing all series in the panel simultaneously and λ_i stands for the individual effects not taken into account by the explanatory variables. Finally the individual error terms $\varepsilon_t^{(i)}$, $t = 1, \ldots, T$, $i = 1, \ldots, n$, are assumed to be independently and identically distributed.

In the analyses η_t often is excluded. Hsiao writes on the following page that “for ease of exposition, we assume that the time specific effects, η_t, do not appear”. Other standard textbooks, e.g. Diggle et al. (1994), Arellano (2003), and the collection of Mátyás and Sevestre (1992) do also not include this term. In the book of Baltagi (2001), interindividual correlation is only considered briefly for regression models, not for dynamic models. Maddala (1971) discusses random time effects but concludes that his estimators are biased in the presence of lagged dependent variables. A variable corresponding to η_t already is ignored in the basic papers on dynamic models by Holtz-Eakin et al. (1988) and Nerlove (1971). In the special framework of a large number of small samples, Cruddas et al. (1989) investigate approximate conditional likelihood estimation for short first-order autoregressive processes; Cox and Solomon (1988) test for serial correlation and Karioti and Caroni (2002) give a method for detecting outlying time series characterised by a level shift. Kiviet (1995) derives an approximate small sample bias for various estimators in dynamic models containing exogenous variables. Still, in each of these cases the time series are assumed to be independent.

To our knowledge, parameter estimation in a dynamic model including η_t has first been investigated by Sethuraman and Basawa (1994). They regard a panel of autoregressive processes with mean zero. In the analyses it is treated as a multivariate time series where the covariance structure of the innovations is accordingly restricted. The asymptotic distributions of the estimators are derived under the assumption that the length of the time series tends to infinity. Hjellvik and Tjøstheim (1999a,b) essentially consider the same model but distinguish η_t and the individual error terms $\varepsilon_t^{(i)}$, $i = 1, \ldots, n$. In Hjellvik and Tjøstheim (1999a) they discuss parameter estimation for this model and derive asymptotic distributions for $n\,T \to \infty$. This also includes the case that only the number of time series tends to infinity, whereas their length remains fixed. In the subsequent paper (Hjellvik and Tjøstheim 1999b) they consider estimation of the variances and order determination. Their line of research has been continued by Fu et al. (2002) who are concerned with model selection criteria.

Forni et al. (2000, 2001) propose a so-called “generalised dynamic factor model” which includes the above model as a special case. They are mainly concerned with determining the number of common factors in a panel model, but their method also allows estimating the parameters in a second step. The underlying idea is to investigate the behaviour of the $(n \times n)$-covariance matrix if the number n of time series in the panel

tends to infinity.
Tests for intercorrelation can be obtained in various ways. Brillinger (1973,1980) proposes a test for intercorrelation derived from the frequency domain representation of the processes. Frees' nonparametric test (Frees 1995), which is based on a U-statistic, is also valid in the case of short time series.
We have not included Bayesian approaches to parameter estimation in the above overview. However it seems that also in the Bayesian framework dynamic models with a common time effect are not commonly used (Congdon 2004, Bauwens et al. 1999).

Robust Methods

It is generally acknowledged that real data always may contain outliers. Hampel (1973), for example, states that their proportion reaches 10 – 15%. Thus we need robust procedures which permit inference even in the presence of outliers. In the case of single time series, many different methods have been proposed for robustly estimating autoregressive parameters. An overview of the classical estimators can be found in Martin and Yohai (1985). Generalised M-estimators (Denby and Martin 1979, Bustos 1982, Künsch 1984) are commonly used, see also Martin and Yohai (1991). Here the estimator is defined in an indirect way and has to be obtained through numerical minimisation. Rousseeuw and Leroy suggest using a robustified least squares procedure, the least median of squares. A similar method is the least trimmed squares estimator (Rousseeuw and Leroy 1987). Because of the computational complexity of these estimators, a subsampling algorithm is needed for computation. They are examples of S-estimators which have been introduced in Rousseeuw and Yohai (1984). The above estimators have been implemented in software packages such as R (Gentleman and Ihaka 2004) and can therefore be employed directly. Furthermore we want to mention R-estimators for parameter estimation in autoregressive models which were discussed in Koul and Saleh (1993) and generalised in Koul and Ossiander (1994). Ferretti et al. (1991) introduce RAR-estimators which are also rank-based in nature. A more recent generalisation are the so-called weighted Wilcoxon estimators (Terpstra et al. 2001). Depending on the weight used, they e.g. correspond to Jaeckel's dispersion function (Jaeckel 1972) with Wilcoxon scores, or in the AR(1) case to the median of pairwise slopes (Theil 1950, Sen 1968). The RA-estimators of Bustos and Yohai (1986) are obtained by modifying the residuals used in the conditional maximum likelihood estimation equations. For the computation an iterative procedure has to be used.
A more direct strategy is to use robust estimators of the covariance matrix $\hat{\Gamma}$ and the corresponding vector of autocovariances $\hat{\gamma}$ in the the least squares or Yule-Walker equations $\hat{\theta} = \hat{\Gamma}^{-1}\hat{\gamma}$. In the first case each element of the matrix and vector is estimated separately, whereas in the second case $\hat{\Gamma} = (\hat{\gamma}(i-j))_{i,j=1,\ldots,p}$ and $\hat{\gamma} = (\hat{\gamma}(1), \ldots, \hat{\gamma}(p))'$. The autocovariances in the above equation may be replaced by autocorrelations. It is also possible to estimate the covariance matrix directly in a robust way. Famous examples are the generalised M-estimators proposed by Maronna (1974) and Tyler (1987) or the projection method advocated in Maronna et al. (1992). Also the minimum volume ellipsoid or the minimum covariance determinant estimators yield robust covariance estimates (Rousseeuw and Leroy 1987). A simulation study comparing the performance of various estimators of these types can be found in Lo and Li (1990).

A third, entirely different possibility for obtaining robust estimators is to identify outliers in a first step. After deleting these data from the sample, non-robust methods can be used for the estimation. One example is the above mentioned reweighted least squares procedure (Rousseeuw and Leroy 1987). It is often stressed (Rousseeuw and Leroy 1987, Huber 1981) that for the diagnostic step a robust estimator should be chosen since otherwise masking effects (Becker and Gather 1999) cannot be excluded.
As far as we know, there exist no methods which have been designed explicitely for robust parameter estimation in panels of time series. In particular the case that one or more time series are generated by a different autoregressive model, whereas the panel otherwise is homogeneous, has never been investigated. The standard procedure is to test for homogeneity first (Hsiao 1986, p. 11). If this assumption is rejected, the data are modeled as heterogeneous.

Outline of the Thesis

We consider panels of dependent time series. More specifically we assume that the individual time series have an autoregressive structure, but that the innovations allow for a common random shock. This is also the model investigated by Hjellvik and Tjøstheim (1999a,b). Their method is to treat the common shocks $\{\eta_t\}_{t\in\mathbb{Z}}$ as a nuisance parameter, which allows them to derive a conditional maximum likelihood estimator for the autoregressive parameters. However this results in a loss of information since only the deviations from the mean process are taken into account. By a factorisation of the conditional likelihood function we obtain a new estimator which also includes the information of the mean process. As it is based on a weighted average of two separate terms, we propose a recursive algorithm for its calculation.
Furthermore the factorisation allows us to generalise our results. We assume that the panel is generated by a single underlying process and that the individual time series are fluctuating around it. To be more specific, we assume that $X_t^{(i)} = Y_t + Z_t^{(i)}$ for $t \in \mathbb{Z}$, $i = 1, \ldots, n$, where $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{Z_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, are independent autoregressive processes. It turns out that the generalised process is a special case of the factor model proposed by Forni et al. (2000).
For proving asymptotic normality of the parameter estimators, we have to distinguish the cases of $n \to \infty$, T fixed, and $T \to \infty$. In the first case, we can use the standard central limit theorem for independently and identically distributed observations. For $T \to \infty$ we however have to employ a central limit theorem for martingale arrays. It is shown that in the case of a finite number of time series the new estimator is more efficient than the one of Hjellvik and Tjøstheim. Moreover we derive the rates of convergence of the estimators. We also briefly discuss the bias terms.

In the second part of the thesis we investigate robust parameter estimation for panels of time series. As mentioned in the beginning, we are especially interested in the case that entire time series are outliers. Concentrating on some basic robust methods which can easily be generalised to the panel case, we analyse three different approaches.
The first one is to robustify an estimator by replacing all non-robust parts with a robust method in a way similar to Haddad (2000). We use this method for the parameter estimator discussed in the previous chapter. To enhance numerical stability we propose an iterative procedure for averaging over matrices. Since bootstrap methods can be

used to assess the empirical bias of parameter estimators, we discuss two versions of time series bootstrap exemplarily for this estimator.

Secondly, autocovariances may be estimated using the identity

$$\mathrm{cov}(X, Y) = \frac{1}{4ab}\left(\mathrm{var}(aX + bY) - \mathrm{var}(aX - bY)\right)$$

which is valid for any square integrable random variables X and Y. Here the variance can be replaced with a robust alternative, see Huber (1981). We generalise the estimator proposed by Ma and Genton (2000), which is based on the robust scale estimator Q_n (Rousseeuw and Croux 1993), to the panel case. A panel breakdown point is defined in two ways depending on the type of outliers assumed. We compute its value for the robust panel autocovariance estimator. The estimated autocovariances are used as the components of the covariance matrix and the autocovariance vector. In contrast to this elementwise robustification, we then study the behaviour of the parameter estimator derived from another method where the covariance matrix is estimated directly using the minimum covariance determinant (MCD) estimator (Rousseeuw and Leroy 1987). Next we treat as a reference two methods designed for robust regression: an M-estimator proposed by Huber (1996) and the least trimmed squares procedure (Rousseeuw and Leroy 1987). These methods are investigated as alternatives to the robustified version of the parameter estimator derived in the first chapter.

Finally we discuss two methods for outlier detection in panels of time series. We focus on the case that entire time series may be generated by another model. The first one is derived from a likelihood ratio test for panel homogeneity which has been proposed by Basawa et al. (1984). We include it in order to illustrate how outliers affect a non-robust test in our setting. The second method is based on a phase space representation of the time series in the panel. It generalises the procedure for fast outlier detection developed by Gather, Imhoff and Fried (2002). All of the proposed robust panel estimators are compared in a simulation study.

As the thesis has been motivated by a medical study, applicability is an important aspect for us. We thus use our methods for analysing two empirical data sets. First we consider the grey-sided voles data which already served Hjellvik and Tjøstheim (1999a) as an example. Then we analyse the data from the fibromyalgia syndrome study mentioned at the beginning. This chapter also illustrates the behaviour of the parameter estimators depending on the strength of the intercorrelation.

Thus the thesis is structured as follows: First we introduce our notation and summarise some basic results. Then we regard the theoretical properties of our parameter estimators in the intercorrelated model and its generalisation. We prove asymptotic normality in both cases and derive the rates of convergence. For numerical simulations we refer to the Appendix. The third chapter is concerned with robust estimation in the panel case. We propose several methods based on the different concepts and investigate their behaviour with simulated data. In the last chapter, we apply our methods to the empirical data. Each chapter concludes with a discussion. The Appendix contains additionally some basic calculations which we include for reference.

Acknowledgements

At this point I want to express my gratitude to all those who have accompanied me through the last years.
First of all I thank my supervisor Prof. Dr. Rainer Dahlhaus for his suggestions, his encouragement and his patience. I am furthermore indebted to my colleagues Suhasini, Michael and Jan for many helpful discussions and proof-reading, and to Christina, who proof-read the first english version of the thesis. All remaining mistakes are my responsibility alone. Thank you to Prof. Dahlhaus and all present and past members of the stats group for the welcoming atmosphere at the department which made working there very agreeable.
During the first years of my PhD thesis I participated in the sessions of the fibromyalgia syndrome research group at the Medical University Hospital of Heidelberg, Department of Internal and Psychosomatic Medicine, under Prof. Dr. Wolfgang Eich. I am grateful for this experience which allowed me to gain insight into the research practice in another field of science. Thank you to all of the team for what I have learned there and also for permitting me to use the fibromyalgia data set in this thesis.
Moreover I am indebted to Prof. Dr. Dag Tjøstheim of the University of Bergen and Dr. Vidar Hjellvik of the Institute of Marine Research, Bergen, for generously providing me with the grey-sided voles data set.
Finally I want to thank my parents and my friends, especially my very best friend and husband Carsten, for constantly supporting me throughout all the ups and downs in writing a PhD thesis - without you I would never have managed!

I tell ya a guy gets too lonely an' he gets sick.

(Steinbeck 2000, p. 72)

Figure 1: Grey-sided vole *(Clethrionomys rufocanus)*.
Photograph©Roar Solheim, Norsk Naturreportasje.

Notation

General notations

I_n	identity matrix of dimension $n \times n$
$\mathbb{1}_n$	matrix of dimension $n \times n$ consisting of ones
$\mathrm{tr}(A)$	trace of the matrix A
$\|A\|$	determinant of the matrix A
$\|\|X\|\| = \left(\sum_{k=1}^n X_k^2\right)^{1/2}$	euclidian norm of the vector $X = (X_1, \ldots, X_n)'$
$\mathbb{R}_0^+ = \{x \in \mathbb{R} \mid x \geq 0\}$	set of non-negative real numbers
$\lfloor x \rfloor$	largest $z \in \mathbb{Z}$ such that $z \leq x$
δ_{ij}	Kronecker delta
$a(\mathrm{L})$	backward shift operator (section 1.1)
causal AR(p) process	process for which the stationarity condition $a(z) \neq 0$ for all z with $\|z\| \leq 1$ (ass. 1.1.1) is fulfilled
$X_n \Rightarrow X$	convergence in distribution
X_n is $\mathrm{AN}(0, \Sigma_n)$	X_n is asymptotically normal (see e.g. Brockwell and Davis (1991))

Panels of time series

$\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$	panel of n time series
$\bar{X}_t = \frac{1}{n}\sum_{i=1}^n X_t^{(i)}$, $t \in \mathbb{Z}$	mean process
$\mathring{X}_t^{(i)} = X_t^{(i)} - \bar{X}_t$, $t \in \mathbb{Z}$	ith residual process, $i = 1, \ldots, n$
$\mathbf{x}_{t-1} = (X_{t-1}, \ldots, X_{t-p})'$	vector of past values derived from an AR(p) process $\{X_t\}_{t\in\mathbb{Z}}$
$\mathcal{L}_{n,T}(\theta)$	cond. log-likelihood function in the ICM (theorem 2.4.2)
$\mathcal{L}_n(\theta) = \lim_{T\to\infty} \mathcal{L}_{n,T}(\theta)$	pointwise limit of $\mathcal{L}_{n,T}(\theta)$ for $T \to \infty$, n fixed (def. 2.5.3)
$\mathcal{L}(\theta) = \lim_{n\to\infty} \mathcal{L}_{n,T}(\theta)$	pointwise limit of $\mathcal{L}_{n,T}(\theta)$ for $n \to \infty$ (def. 2.5.3)

The intercorrelation model (ICM) is defined in section 2.2 and its generalised version (GICM) in section 2.3. For $t \in \mathbb{Z}$, $i = 1, \ldots, n$, we have

- in the ICM:

  $X_t^{(i)} = \sum_{u=0}^\infty \psi_u \left(\varepsilon_{t-u}^{(i)} + \eta_{t-u}\right) = \sum_{u=0}^\infty \psi_u \left(\mathring{\varepsilon}_{t-u}^{(i)} + \xi_{t-u}\right)$

  with $\varepsilon_t^{(i)} \sim \mathrm{N}(0, \sigma^2)$, $\eta_t \sim \mathrm{N}(0, \tau^2)$, $\xi_t = \bar{\varepsilon}_t + \eta_t \sim \mathrm{N}(0, \omega_n^2)$, where $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n}$

- in the GICM:

  $X_t^{(i)} = Z_t^{(i)} + Y_t = \sum_{u=0}^\infty \psi_u \zeta_{t-u}^{(i)} + \sum_{u=0}^\infty \varphi_u \upsilon_{t-u}$,

  where $\zeta_t^{(i)} \sim \mathrm{N}(0, \sigma_n^2)$, $\upsilon_t \sim \mathrm{N}(0, \omega_n^2)$;

  $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$, $\omega^2 = \lim_{n\to\infty} \omega_n^2$, $\sigma_n^{ij} = \mathrm{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right)$, $i \neq j$; $\tilde{\sigma}_n^2 = \sigma_n^2 - \sigma_n^{ij}$

Always let $\sum_{u=0}^\infty |\psi_u| < \infty$ and $\sum_{u=0}^\infty |\varphi_u| < \infty$

and denote $\Psi(h) = \sum_{u=0}^\infty \psi_u \psi_{u+|h|}$ and $\Phi(h) = \sum_{u=0}^\infty \varphi_u \varphi_{u+|h|}$.

Chapter 1

Preliminaries

The topic of this thesis are special panels of autoregressive time series. Depending on the context, we use different notations for representing the processes. These are introduced in the first section. Furthermore we present some fundamental properties of stationary autoregressive time series here and state our basic assumptions.
A tool employed several times throughout the thesis is the panel covariance estimator $\hat{\gamma}_{n,T}(h) = \frac{1}{n(T-h)} \sum_{i=1}^{n} \sum_{t=h+1}^{T} X_t^{(i)} X_{t-h}^{(i)}$, where n denotes the number of time series in the panel and T their length. Under some regularity conditions on the correlation structure of the time series in the panel it is mean-square consistent for $n\,T \to \infty$. We prove this result in section 1.2.

1.1 Basic Results

In order to represent autoregressive time series, we often use backward shift operators for ease of notation. The one-step backward shift operator L is defined by

$$\mathrm{L}\big(Z_t\big) = Z_{t-1}$$

for an arbitrary process $\{Z_t\}_{t\in\mathbb{Z}}$. Then the AR(p) process with innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$,

$$X_t = \sum_{k=1}^{p} a_k\, X_{t-k} + \varepsilon_t\,, \quad t \in \mathbb{Z},$$

can be written in the form

$$a(\mathrm{L})\, X_t = \sum_{k=0}^{p} \alpha_k\, \mathrm{L}^k\big(X_t\big) = \varepsilon_t \quad \text{for all } t \in \mathbb{Z}\,.$$

Here $\alpha_0 = 1$ and $\alpha_k = -a_k$ for all $k = 1, \ldots, p$. We refer to the linear operator $a(\mathrm{L})$ as the *backward shift operator*.

Throughout this thesis we impose the following assumptions (see e.g. Brockwell and Davis 1991).

1.1.1 ASSUMPTION
Let $\{X_t\}_{t\in\mathbb{Z}}$ be a zero mean *causal* AR(p) process. This means that

$$X_t = \sum_{k=1}^{p} a_k\, X_{t-k} + \varepsilon_t \qquad \text{for all } t \in \mathbb{Z},$$

where $a_p \neq 0$ and the $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ form a white noise process with $\mathbb{E}\,(\varepsilon_t^2) = \sigma^2$ for all $t \in \mathbb{Z}$. Furthermore the coefficients fulfil

$$a(z) = 1 - a_1 z - \cdots - a_p z^p \neq 0 \quad \text{for all } z \in \mathbb{C} \text{ such that } |z| \leq 1.$$

The above condition implies that the process is (weakly) stationary (see e.g. Shiryayev 1984, p. 392). A second consequence is that in this case the autoregressive process can be written as an MA(∞) process with absolutely summable coefficients.

1.1.2 LEMMA
Under assumption 1.1.1 the process $\{X_t\}_{t\in\mathbb{Z}}$ admits a MA(∞) representation, i.e. for all $t \in \mathbb{Z}$ we have $X_t = \sum_{u=0}^{\infty} \psi_u\, \varepsilon_{t-u}$, where $\sum_{u=0}^{\infty} |\psi_u| < \infty$.

PROOF:
See e.g. Lütkepohl (1991). □

The proof of this well-known fact is based upon the representation of the univariate AR(p) process as a vector autoregressive process of order one which is obtained as follows. If $a_1, \ldots, a_p$ are the coefficients of the autoregressive process $\{X_t\}_{t\in\mathbb{Z}}$, let

$$A = \begin{pmatrix} a_1 & \cdots & \cdots & a_p \\ & I_{p-1} & & \underline{0}_{p-1} \end{pmatrix},$$

where I_{p-1} is the identity matrix of order $(p-1)$ and $\underline{0}_{p-1} = (0, \ldots, 0)'$. Denote $\mathbf{x}_t = (X_t, \ldots, X_{t-p+1})'$ and $\underline{\varepsilon}_t = (\varepsilon_t, 0, \ldots, 0)'$. Then for $t \in \mathbb{Z}$ the process can be written as

$$\mathbf{x}_t = A\,\mathbf{x}_{t-1} + \underline{\varepsilon}_t\,.$$

Since it can easily be seen that $\det(I_p - zA) = 1 - \sum_{k=1}^{p} a_k z^k$, the condition in assumption 1.1.1 implies that all eigenvalues λ of A fulfil $|\lambda| < 1$. This leads to a stronger result. The proof follows the reasoning of Künsch (1995).

1.1.3 PROPOSITION
The MA(∞) coefficients $\{\psi_u\}_{u\geq 0}$ of an AR(p) process as in assumption 1.1.1 fulfil

$$|\psi_u| \leq c\,\rho^u \quad \textit{for } u \geq 0\,,$$

with constants $c > 0$ and $\rho < 1$.

PROOF:
For any matrix M let $\lambda_{max}(M) = \max\{|\lambda|;\, \lambda \text{ eigenvalue of } M\}$.
It is known that for every matrix M and every $\varepsilon > 0$ there exists a matrix norm $||.||$ such that

$$\lambda_{max}(M) \;\leq\; ||M|| \;\leq\; \lambda_{max}(M) + \varepsilon \tag{1.1}$$

(see e.g. Lütkepohl 1996, 8.4.1 (15)).
Let A be the matrix in the vector autoregressive representation of $\{X_t\}_{t\in\mathbb{Z}}$. As all eigenvalues of A have modulus less than 1, we can choose $\varepsilon > 0$ and $\rho < 1$ such that $\lambda_{max}(A) + \varepsilon \leq \rho$. Denote the matrix norm which fulfils (1.1) by $||.||_M$. Then $||A||_M \leq \rho < 1$.
Moreover, for any two matrix norms $||.||_a$ and $||.||_b$ there exists a positive constant $c \in \mathbb{R}$ such that

$$||A||_a \leq c\,||A||_b$$

for all $(m \times m)$ matrices A (see e.g. Lütkepohl 1996, 8.3 (15)). Therefore we get

$$|\psi_u| = |\big(A^u\big)_{1,1}| \leq ||A^u||_1 \leq c\,||A^u||_M \leq c\,||A||_M^u \leq c\,\rho^u\,,$$

where $c > 0$ is constant. Here $(A^u)_{k,l}$ denotes the (k,l)th component of the matrix A^u and $||A^u||_1 = \max\{(A^u)_{k,l};\ k,l = 1,\ldots,p\}$. □

This means that the autocovariance function of a Gaussian AR(p) process is square summable.

1.1.4 LEMMA
Let $\{X_t\}_{t\in\mathbb{Z}}$ be a causal AR(p) process with $\varepsilon_t \sim \mathrm{N}(0,\sigma^2)$ for all $t \in \mathbb{Z}$. Then the autocovariance function $\gamma_X(h) = \mathrm{cov}\,(X_t, X_{t+h})$, $h \in \mathbb{Z}$, fulfils $\sum_{h=0}^{\infty} |\gamma_X(h)| < \infty$, which implies that $\sum_{h=0}^{\infty} \gamma_X(h)^2 < \infty$.

PROOF:
We have seen in lemma 1.1.2 that the process $\{X_t\}_{t\in\mathbb{Z}}$ has a MA(∞) representation such that $X_t = \sum_{u=0}^{\infty} \psi_u\,\varepsilon_{t-u}$ for all $t \in \mathbb{Z}$, where $\sum_{u=0}^{\infty} |\psi_u| < \infty$. Because of the orthogonality properties of the innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ we thus have for $h \in \mathbb{Z}$ that

$$\gamma_X(h) = \mathrm{cov}\,(X_t, X_{t+h}) = \sum_{u=0}^{\infty} \psi_u\,\psi_{u+|h|}\,\sigma^2 < \infty\,.$$

Furthermore the preceding lemma shows that the coefficients $\{\psi_u\}_{u\geq 0}$ fulfil $|\psi_u| < c\,\rho^u$ for all $u \geq 0$, where $c > 0$ and $0 < \rho < 1$. Therefore for all $h \in \mathbb{Z}$

$$|\gamma_X(h)| \leq \sum_{u=0}^{\infty} |\psi_u|\,|\psi_{u+|h|}|\,\sigma^2 \leq \sum_{u=0}^{\infty} \rho^{2u+|h|}\,c^2\,\sigma^2 = \rho^{|h|}\,\frac{1}{1-\rho^2}\,c^2\,\sigma^2\,.$$

The result follows directly. □

Finally we want to emphasise the important relation between the MA(∞) coefficients or the autocovariance function of an autoregressive process and its autoregressive parameter.

1.1.5 REMARK
Let $\{\psi_u\}_{u\geq 0}$ be the coefficients in the MA(∞) representation of an autoregressive process $\{X_t\}_{t\in\mathbb{Z}}$ fulfilling assumption 1.1.1 and denote $\Psi(h) = \sum_{u=0}^{\infty} \psi_u\,\psi_{u+|h|}$ for $h \in \mathbb{Z}$. Due to the absolute summability we get that $\gamma_X(h) = \mathrm{cov}\,(X_t, X_{t+h}) = \Psi(h)\,\sigma^2$ for

$h \in \mathbb{Z}$, where $\sigma^2 = \operatorname{var}(\varepsilon_t)$. Thus (see e.g. Brockwell and Davis 1991, p. 93) the autoregressive parameter $a = (a_1, \dots, a_p)'$ of the process fulfils, if we denote $a_0 = -1$,

$$\sum_{l=0}^{p} a_l \, \Psi(k-l) = 0 \quad \text{for all } k > 0 \quad \text{and} \qquad \sum_{l=0}^{p} a_l \, \Psi(-l) = \sum_{l=0}^{p} a_l \, \Psi(l) = -1 \, ,$$

which is just another form of the Yule-Walker equations. This implies in particular that

$$\sum_{k,l=0}^{p} a_k \, a_l \, \Psi(k-l) = -\sum_{l=0}^{p} a_l \, \Psi(-l) = 1 \, .$$

Because of the structure of the autocovariance function, these statements can directly be transferred to autocovariance functions of any stationary autoregressive process with the same autoregressive parameters.

The basic properties of the preceding remark are used frequently in this thesis. Furthermore we can employ the following result for calculating higher order mixed moments, as these can be reduced to products of covariances if the underlying processes are Gaussian.

1.1.6 PROPOSITION
Let $\{X_t\}_{t \in \mathbb{Z}}$ be a causal autoregressive process as in assumption 1.1.1 with autoregressive parameter $a = (a_1, \dots, a_p)'$ and autocovariance function $\gamma(h) = \Psi(h)\,\sigma^2$, $h \in \mathbb{Z}$. Denote $a_0 = -1$. Then for any $z \in \mathbb{Z}$,

$$\sum_{k,l=0}^{p} a_k a_l \sum_{s,t=p+1}^{T} \gamma(s-t-k+l)\,\gamma(s-t+z) = (T-p)\,\sigma^2\,\gamma(z)$$

and for any $z_1, z_2 \in \mathbb{Z}$ such that $z_1 + z_2 > 0$,

$$\sum_{k,l=0}^{p} a_k a_l \sum_{s,t=p+1}^{T} \gamma(s-t+z_1-k)\,\gamma(s-t-z_2+l) = 0 \, .$$

PROOF:
We have mentioned in the preceding remark that the Yule-Walker equations for autocovariances lead to $\sum_{l=0}^{p} a_l \, \gamma(k-l) = 0$ for all $k > 0$ and $\sum_{l=0}^{p} a_l \, \gamma(-l) = -\sigma^2$. Thus for any $z \in \mathbb{Z}$

$$\begin{aligned}
\sum_{k,l=0}^{p} a_k a_l & \sum_{s,t=p+1}^{T} \gamma(s-t-k+l)\,\gamma(s-t+z) \\
&= \sum_{h=-(T-p-1)}^{T-p-1} ((T-p)-|h|) \sum_{k,l=0}^{p} a_k a_l \, \gamma(h-k+l)\,\gamma(h+z) \\
&= \sum_{h=0}^{T-p-1} ((T-p)-h) \sum_{l_1=0}^{p} a_l \left(\sum_{k=0}^{p} a_k \, \gamma(h+l-k) \right) \gamma(h+z) \\
&\quad + \sum_{h=1}^{T-p-1} ((T-p)-h) \sum_{k=0}^{p} a_k \left(\sum_{l_1=0}^{p} a_l \, \gamma(h+k-l) \right) \gamma(h-z)
\end{aligned}$$

$$
\begin{aligned}
&= (T-p)\, a_0 \left(\sum_{k=0}^{p} a_k\, \gamma(-k) \right) \gamma(z) \\
&= (T-p)\, \sigma^2\, \gamma(z)\,,
\end{aligned}
$$

since with exception of $h = l = 0$ all terms vanish.
Moreover, for any z_1 and z_2 such that $z_1 + z_2 > 0$, the same reasoning leads to

$$
\begin{aligned}
&\sum_{k,l=0}^{p} a_k a_l \sum_{s,t=p+1}^{T} \gamma(s-t+z_1-k)\, \gamma(s-t-z_2+l) \\
&\quad = \sum_{s,t=p+1}^{T} \left(\sum_{k=0}^{p} a_k\, \gamma\left((s-t+z_1)-k\right) \right) \times \left(\sum_{l_1=0}^{p} a_l\, \gamma\left((s-t-z_2)+l\right) \right) \\
&\quad = \sum_{\substack{s,t=p+1 \\ s-t+z_1 \le 0}}^{T} \left(\sum_{k=0}^{p} a_k\, \gamma\left((s-t+z_1)-k\right) \right) \times \left(\sum_{l_1=0}^{p} a_l\, \gamma\left((t-s+z_2)-l\right) \right) \\
&\quad = \sum_{\substack{s,t=p+1 \\ s-t+z_1 \le 0 \\ t-s+z_2 \le 0}}^{T} \left(\sum_{k=0}^{p} a_k\, \gamma\left((t-s-z_1)+k\right) \right) \times \left(\sum_{l_1=0}^{p} a_l\, \gamma\left((t-s+z_2)-l\right) \right) \\
&\quad = \ 0\,,
\end{aligned}
$$

as the last sum is empty. □

1.2 The Panel Autocovariance Estimator

In this thesis we are concerned with identically distributed but dependent time series. In order to investigate the asymptotic behaviour of the parameter estimators, we need an estimator of the autocovariance function of the time series which is consistent if $n\,T \to \infty$. Indeed it is not necessary that the time series are independent. We only have to assume that the cross-sectional correlation is bounded (if $n, T \to \infty$) or tends to zero (if $n \to \infty$, T fixed). More precisely, we impose the following:

1.2.1 ASSUMPTION
Let $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $\;i = 1,\dots,n$, be a panel of identically distributed weakly stationary time series such that

$$
\begin{aligned}
X_t^{(i)} &\sim \mathrm{N}(0, \sigma_n^2)\,, \\
\gamma_n^{ii}(h) &= \operatorname{cov}(X_t^{(i)}, X_{t+h}^{(i)}) = \gamma_n(h) \text{ for all } t \in \mathbb{Z}; \\
\text{and for } i \neq j \quad \gamma_n^{ij}(h) &= \operatorname{cov}(X_t^{(i)}, X_{t+h}^{(j)}) = u_n \gamma_n(h)\,,
\end{aligned}
$$

where $\gamma_n(h)$, $h \in \mathbb{Z}$, and $u_n \in \mathbb{R}$ are independent of i and j.
Furthermore assume that the autocovariance function is square summable, i.e. that

$$
\sum_{h=0}^{\infty} \gamma_n^2(h) < \infty\,.
$$

This is motivated by the intercorrelation model ("*ICM*") we will investigate in this thesis (definition 2.2.2). In this model the time series in the panel are not independent. By subtracting the pointwise sample mean from each time series $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, we obtain residual processes $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$. The covariance function for two residual processes $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$ and $\{\mathring{X}_t^{(j)}\}_{t\in\mathbb{Z}}$ is in the example of the ICM

$$\mathring{\gamma}_n^{ij}(h) = \operatorname{cov}\left(\mathring{X}_t^{(i)}, \mathring{X}_{t+h}^{(j)}\right) = \left(\delta_{ij} - \frac{1}{n}\right) \Psi(h)\, \sigma^2 \,,$$

where σ^2 and $\Psi(h)$, $h \in \mathbb{Z}$, (which is independent of i and j) do not depend on n.

1.2.2 REMARK

1. In this section, the assumption of a centred process is mainly for notational convenience. The subsequent proposition remains valid if we suppose $\mu \neq 0$ and change the atuocovariance estimator accordingly, using the overall mean as an estimator of μ. However, the models investigated later on are always panels formed of autoregressive processes with zero mean or linear combinations of such processes.

2. The intercorrelation coefficient u_n could also be defined as a function of the lag h. However, this would complicate the notations in the following lemma. As can be seen from its proof, we would have to assume a common upper bound for the $u_n(h)$ in the case of $n, T \to \infty$. And if $n \to \infty$, T fixed, the convergence should be uniform in h. In practice, u_n will mostly be chosen independent of h. Very often we moreover assume that $u_n = O\left(\frac{1}{n}\right)$. In the example of the ICM, $u_n(h) = -1/(n-1)$ for all lags h.

We always consider panel autocovariance estimators of the following form.

1.2.3 DEFINITION

Let $X_t^{(i)}$*,* $t = 1, \dots, T$*,* $i = 1, \dots, n$*, be observations from a panel of time series as in assumption 1.2.1. For* $h \geq 0$ *we define the* panel autocovariance estimator $\hat{\gamma}_{n,T}(h)$ *as the estimator of* $\gamma_n(h)$ *obtained by*

$$\hat{\gamma}_{n,T}(h) = \frac{1}{n\,(T-h)} \sum_{i=1}^{n} \sum_{t=h+1}^{T} X_t^{(i)}\, X_{t-h}^{(i)} \,.$$

For $h < 0$*, let* $\hat{\gamma}_{n,T}(h) = \hat{\gamma}_{n,T}(-h)$*.*

Now we prove mean-square consistency for this estimator. If the intercorrelation coefficient u_n decreases fast enough, it even holds if only the number of time series tends to infinity. As the conditions of the lemma are fulfilled by the residual processes in the ICM, the result is used several times in the subsequent proofs. It is the main tool for establishing the asymptotic properties of our parameter estimators.

1.2.4 LEMMA
Let $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$*,* $i = 1, \ldots, n$*, be a panel of time series as in assumption 1.2.1. Then we have for* $|h| \le T$ *that*

$$\mathbb{E}\left(\hat{\gamma}_{n,T}(h) - \gamma_n(h)\right)^2 = O\left(\frac{1}{n\,T}\right) + O\left(\frac{u_n^2}{T}\right) .$$

PROOF:
Let $n, T \in \mathbb{N}$, $0 \le h < T$, and $I = \{h+1, \ldots, T\} \times \{1, \ldots, n\}$. Due to the linearity of the expectation we have $\mathbb{E}\left(\hat{\gamma}_{n,T}(h)\right) = \gamma_n(h)$. Since by assumption $X_t^{(i)} \sim \mathrm{N}(0, \sigma_n^2)$ for all $i = 1, \ldots, n$, all cumulants of third and higher order are zero. Therefore we get analogously to (Shiryayev 1984, p. 290) that

$$\begin{aligned}
&\mathbb{E}\left(\hat{\gamma}_{n,T}(h) - \gamma_n(h)\right)^2 \\
&= \frac{1}{n^2\,(T-h)^2} \sum_{(t_1,i),(t_2,j)\in I} \Big(\gamma_n^{ij}(t_1-t_2)^2 + \gamma_n^{ij}(t_1-t_2+h)\,\gamma_n^{ij}(t_1-t_2-h) \\
&\qquad\qquad + \gamma_n^{ii}(h)\,\gamma_n^{jj}(h)\Big) - \gamma_n(h)^2 \\
&= \frac{1}{n\,(T-h)^2} \sum_{t_1,t_2=h+1}^{T} \left(1 + (n-1)\,u_n^2\right) \\
&\qquad\qquad \times \left(\gamma_n(t_1-t_2)^2 + \gamma_n(t_1-t_2+h)\,\gamma_n(t_1-t_2-h)\right) \\
&\le \frac{1}{n\,(T-h)^2} \sum_{s=-(T-h-1)}^{T-h-1} (T-h-|s|)\,\left(1 + (n-1)\,u_n^2\right) \\
&\qquad\qquad \times \left(\gamma_n(s)^2 + \gamma_n(s+h)^2 + \gamma_n(s-h)^2\right) .
\end{aligned}$$

The last inequality is just an application of the second binomial formula. Since by assumption 1.2.1 we have that $\sum_{h=0}^{T-1} \gamma_n^2(h) = O(1)$, this concludes the proof. □

1.2.5 REMARK

1. The above result illustrates in particular the important role of the strength of intercorrelation induced by u_n. If n is fixed, u_n obviously is a constant. However in the case of $n \to \infty$ the lemma only yields mean-square convergence if $\lim_{n\to\infty} |u_n| \le c < \infty$ (if also $T \to \infty$) or if $\lim_{n\to\infty} u_n = 0$ (if T is fixed).

2. The first case, $T \to \infty$, n fixed, could also be obtained directly using the Minkowski inequality. It is a direct consequence of the mean-square convergence of the original autocovariance estimator.

3. Note that the statement of the theorem remains unchanged if we replace the factor $1/(T-h)$ in the definition of the panel autocovariance estimator by $1/(T-p)$, where $p < T$ is fixed, and start the summation with $t = p+1$. This is the form of the lemma used from now on, as we focus on AR(p) processes.

4. For proving convergence of the panel autocovariance estimator it is not necessary that the autocovariance function is square summable. As all processes are

Gaussian by assumption, they admit continuous spectral functions. This is equivalent to the fact that $\frac{1}{T}\sum_{h=0}^{T}\gamma_n^2(h) \to 0$ for $T \to \infty$ (Shiryayev 1984, p. 414), which is sufficient for proving the mean-square convergence. However, we have shown in lemma 1.1.4 that in the case of Gaussian AR(p) processes, $\gamma_n(h)$ is always square summable, and these are the models we regard in this thesis. The assumption allows the direct computation of the rate of convergence.

As we see from the proof of the above lemma, the rate of convergence for $n \to \infty$ depends on the behaviour of the intercorrelation coefficient $\{u_n\}_{n\geq 0}$. Thus we get a $\sqrt{n}$-rate of convergence under restrictions on u_n. As in particular the interest is on convergence to a fixed limit autocovariance function, we state the result as follows.

1.2.6 COROLLARY

If in the above setting there exists an autocovariance function γ such that for all $h \in \mathbb{Z}$ $|\gamma_n(h) - \gamma(h)| = O\left(\frac{1}{n}\right)$ and if furthermore $u_n^2 = O\left(\frac{1}{n}\right)$, we get that

$$\mathbb{E}\left(\hat{\gamma}_{n,T}(h) - \gamma(h)\right)^2 = O\left(\frac{1}{nT}\right).$$

PROOF:

This is an direct conclusion from the mean-square convergence of $\hat{\gamma}_{n,T}(h) - \gamma_n(h)$, which has been proved in the preceding lemma 1.2.4. □

Chapter 2

The Intercorrelation Model

2.1 Motivation

There are applications in which the hypothesis that the time series in a panel are independent is artificial. The following data illustrates this nicely. It also served Hjellvik and Tjøstheim (1999a,b) as an example of intercorrelated time series.
Figure 2.1 shows the yearly catches of grey-sided voles at 41 different locations on Hokkaido, Japan, on a logarithmic scale. The measurements cover the span of 31 years, from 1962 to 1992. We can see that there are years where most of the time series simultaneously attain a local minimum or maximum respectively; this suggests a strong intercorrelation of the time series.

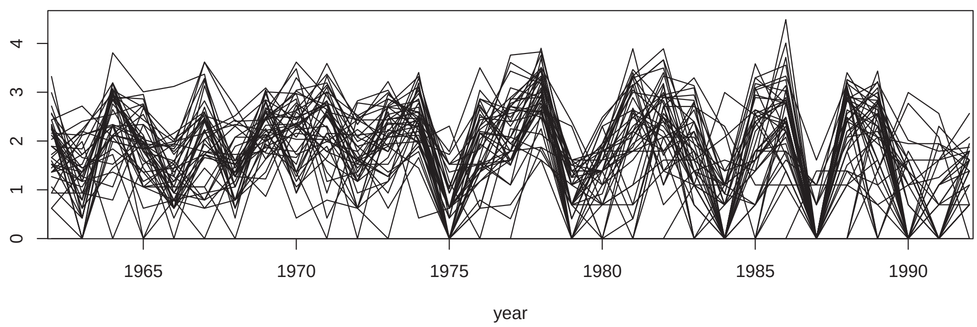

Figure 2.1: Vole data: $\log(V_t^{(i)} + 1)$, where $\{V_t^{(i)}, 1962 \leq t \leq 1992, 1 \leq i \leq 41\}$ is the number of grey-sided voles captured each year from 1962 to 1992 in 41 different locations in Hokkaido, Japan.

If we take the number of trapped voles as an indicator for the size of the population, we can, for example, think of climatic influences such as exceptionally hot summers or cold winters invoking this pattern. Another possibility is the existence of some predator which hunts the voles and is more mobile than they are. It is easy to imagine even more complex settings. Often it may be difficult to find the right covariates. Moreover, these data might not be available. This indeed is the case for the voles data where we were not provided with further information. Thus our approach is to model the common effects as random influences.

We now first introduce one specific model for a panel of intercorrelated time series, the ICM. Then we investigate a generalisation thereof, where the background process and the individual processes are allowed to have different autoregressive coefficients $a = (a_1, \dots, a_p)'$ and $b = (b_1, \dots, b_q)'$. We call this model the GICM. In section 2.4 we derive conditional maximum likelihood estimators for both models. It turns out that the estimator developed by Hjellvik and Tjøstheim (1999a) is the same as the parameter estimator $\hat{\theta}_a$ obtained from the general model under the restriction that $a = b$ (remark 2.4.8). Subsequently, we prove asymptotic normality for these estimators. We show that if $n \to \infty$, the estimator of the autoregressive parameters in the ICM and the estimator obtained from $\hat{\theta}_a$ are asymptotically equivalent. However, if n is fixed and $T \to \infty$, the ICM estimator has a higher relative efficiency. Finally, we discuss the rates of convergence and the bias. It is shown that in the case of n fixed, the mean squared error of the dominating term in the stochastic expansion is smaller for the ICM estimator than for the estimator of Hjellvik and Tjøstheim (1999a). The chapter concludes with an evaluation of the obtained results and an outlook on possible extensions. Simulations illustrating the performance of the estimators can be found in the Appendix A.

2.2 The Model (ICM)

We consider a panel of n intercorrelated time series

$$X_t^{(i)} = \sum_{k=1}^{p} a_k X_{t-k}^{(i)} + \varepsilon_t^{(i)} + \eta_t, \quad i = 1, \dots, n,\ t \in \mathbb{Z}\,,$$

where p denotes the order of the autoregressive process. Here $\varepsilon_t^{(i)}$ is a random shock specific for the time series i, while η_t denotes the common cross sectional influence. We only investigate the case of real valued time series. Moreover we assume that all time series admit the same dynamical structure, i.e. that the coefficients a_k, $k = 1, \dots, p$, are independent of i. This is the model also treated by Hjellvik and Tjøstheim (1999a,b).

Altogether, we assume the following:

2.2.1 ASSUMPTION

(i) The processes $\{\varepsilon_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, and $\{\eta_t\}_{t\in\mathbb{Z}}$ are independent Gaussian white noise processes with

$$\varepsilon_t^{(i)} \sim \mathrm{N}(0, \sigma^2) \quad \text{for } t \in \mathbb{Z},\ i = 1, \dots, n,$$

and

$$\eta_t \sim \mathrm{N}(0, \tau^2) \quad \text{for } t \in \mathbb{Z}\,.$$

(ii) The processes $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, are given by

$$X_t^{(i)} = \sum_{k=1}^{p} a_k\, X_{t-k}^{(i)} + \varepsilon_t^{(i)} + \eta_t \quad \text{for } t \in \mathbb{Z},\ i = 1, \dots, n\,.$$

They are causal (see assumption 1.1.1) with MA(∞) representation

$$X_t^{(i)} = \sum_{u=0}^{\infty} \psi_u \, \xi_{t-u}^{(i)} \qquad \text{for all } t \in \mathbb{Z},\ i = 1, \ldots, n,$$

$$\text{where} \quad \xi_t^{(i)} = \varepsilon_t^{(i)} + \eta_t \quad \text{and} \quad \sum_{u=0}^{\infty} |\psi_u| < \infty.$$

(iii) The parameter $\theta = (a_1, \ldots, a_p, \sigma^2, \tau^2)' \in \Theta$, where $\Theta \subset \mathbb{R}^{p+2}$ is a compact parameter space. Furthermore we have for all $\theta = (\alpha', \sigma_\theta^2, \tau_\theta^2)' \in \theta$ that $\tau^2 \geq 0$ and that there exists a $c > 0$ such that $\sigma^2 \geq c$ for all $\theta \in \Theta$.

Thus the model is a panel of identically distributed autoregressive time series sharing a common intercorrelation factor.

2.2.2 DEFINITION
If assumption 2.2.1 is fulfilled, we call the panel of time series described above the intercorrelation model ("ICM"). *From the ICM we derive the* mean process

$$\bar{X}_t = \frac{1}{n} \sum_{i=1}^{n} X_t^{(i)}, \quad t \in \mathbb{Z}, \tag{2.1}$$

and the n residual processes

$$\mathring{X}_t^{(i)} = X_t^{(i)} - \bar{X}_t, \quad t \in \mathbb{Z}, i = 1, \ldots, n. \tag{2.2}$$

2.2.3 REMARK

1. Using the backward shift operator $a(L)$ (section 1.1), equations (2.1) and (2.2) can be written for $t \in \mathbb{Z}$, $i = 1, \ldots, n$, as

$$a(\mathrm{L})\, \bar{X}_t = \eta_t + \bar{\varepsilon}_t, \qquad \text{where } \bar{\varepsilon}_t = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_t^{(i)},$$

$$\text{and} \quad a(\mathrm{L})\, \mathring{X}_t^{(i)} = \mathring{\varepsilon}_t^{(i)}, \qquad \text{where } \mathring{\varepsilon}_t^{(i)} = \varepsilon_t^{(i)} - \bar{\varepsilon}_t\,.$$

2. $\bar{\varepsilon}_s$ and $\mathring{\varepsilon}_t^{(i)}$ are independent for all $s, t \in \mathbb{Z}$ since the processes $\varepsilon_t^{(i)}$, $i = 1, \ldots, n$, are independent and Gaussian. Thus also $\{\bar{X}_t\}_{t \in \mathbb{Z}}$ and $\{\mathring{X}_t^{(i)}\}_{t \in \mathbb{Z}}$, $i = 1, \ldots, n$, are independent Gaussian processes.

3. As the processes $\{X_t^{(i)}\}_{t \in \mathbb{Z}}$, $i = 1, \ldots, n$, are causal, this is also the case for $\{\bar{X}_t\}_{t \in \mathbb{Z}}$ and $\{\mathring{X}_t^{(i)}\}_{t \in \mathbb{Z}}$, $i = 1, \ldots, n$. They admit representations as $\mathrm{MA}(\infty)$ processes with the same coefficients $\{\psi_u\}_{u \geq 0}$ (see lemma 1.1.2).

   So for all $t \in \mathbb{Z}$, $i = 1, \ldots, n$,

$$X_t^{(i)} = \sum_{u=0}^{\infty} \psi_u \left(\eta_{t-u} + \varepsilon_{t-u}^{(i)}\right), \qquad \bar{X}_t = \sum_{u=0}^{\infty} \psi_u \left(\eta_{t-u} + \bar{\varepsilon}_{t-u}\right)$$

and $\quad \mathring{X}_t^{(i)} = \sum_{u=0}^{\infty} \psi_u \, \mathring{\varepsilon}_{t-u}^{(i)} \, .$

This means that the process $X_t^{(i)} = \mathring{X}_t^{(i)} - \bar{X}_t$ can be viewed as a sum of two MA(∞) processes having the same coefficients.

We can easily derive the autocovariance functions. The next lemma serves as reference as we use these representations throughout the entire thesis.

2.2.4 LEMMA
Let $\Psi(h) = \sum_{u=0}^{\infty} \psi_u \, \psi_{u+|h|}$.
The autocovariance functions in the ICM are given for $h \in \mathbb{Z}$, $i, j = 1, \ldots, n$, *by*

$$\gamma_n(h) = \operatorname{cov}\left(X_t^{(i)}, X_{t+h}^{(i)}\right) = \Psi(h)\,(\tau^2 + \sigma^2)\,,$$

$$\bar{\gamma}_n(h) = \operatorname{cov}\left(\bar{X}_t, \bar{X}_{t+h}\right) = \Psi(h)\,\omega_n^2\,, \ \textit{where}\ \omega_n^2 = \operatorname{var}(\eta_t + \bar{\varepsilon}_t) = \tau^2 + \frac{\sigma^2}{n},$$

$$\textit{and}\ \mathring{\gamma}_n(h) = \operatorname{cov}\left(\mathring{X}_t^{(i)}, \mathring{X}_{t+h}^{(i)}\right) = \Psi(h)\left(\frac{n-1}{n}\right)\sigma^2\,.$$

For $i \neq j$, $\mathring{\gamma}_n^{ij}(h) = \operatorname{cov}(\mathring{X}_t^{(i)}, \mathring{X}_{t+h}^{(j)}) = \Psi(h)\,\left(-\frac{\sigma^2}{n}\right)$.
If $\{Z_t^{(i)}\}_{t\in\mathbb{Z}}$ *is the processes generated by* $\{\varepsilon_t^{(i)}\}_{t\in\mathbb{Z}}$, *where* $i \in \{1, \ldots, n\}$, *i.e. if*

$$Z_t^{(i)} = \sum_{u=0}^{\infty} \psi_u \, \varepsilon_{t-u}^{(i)} \qquad \textit{for all}\ t \in \mathbb{Z},$$

its autocovariance functions is given by $c(h) = \operatorname{cov}\left(Z_t^{(i)}, Z_{t+h}^{(i)}\right) = \Psi(h)\,\sigma^2$.

PROOF:
The assertions can be derived directly from the MA(∞) representations of the processes as the coefficients $\{\psi_u\}_{u\geq 0}$ are absolutely summable (lemma 1.1.2). □

2.2.5 REMARK

1. Note that the processes $\{Z_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, are not observable. However they are used e.g. as a tool in the proof of asymptotic normality of the parameter estimators in the ICM in the case of $n \to \infty$, T fixed. We can also represent $\mathring{\gamma}_n$ by $\mathring{\gamma}_n(h) = \frac{n-1}{n}\, c(h)$ for all $h \in \mathbb{Z}$.

2. In lemma 1.2.4 we have shown mean-square convergence of the autocovariance parameter estimator. The rate is $O\left(\frac{1}{n\,T}\right)$ if $u_n^2 = O\left(\frac{1}{n}\right)$ (see corollary 1.2.6). This is fulfilled by the processes $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$ in the ICM: as can be seen from the above considerations, there

$$u_n = \frac{\mathring{\gamma}_n^{ij}(h)}{\mathring{\gamma}_n(h)} = -\frac{1}{n-1} \quad \text{for all}\ h \in \mathbb{Z}\,.$$

3. The processes $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, themselves do not fulfil the condition of $u_n^2 = O\left(\frac{1}{n}\right)$. For the mean process $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ we however get with the methods of lemma 1.2.4 that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}\bar{X}_t\,\bar{X}_{t-h}-\bar{\gamma}_n(h)\right)^2 \\ &= \frac{1}{(T-p)^2}\sum_{s,t=p+1}^{T}\big(\Psi(s-t)^2+\Psi(s-t-h)\,\Psi(s-t+h)\big)\;\omega_n^4 \\ &= O\left(\frac{\omega_n^4}{T}\right)\end{aligned}$$

since all higher order cumulants are zero as the process $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ is Gaussian. As $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n}$, we thus even have mean-square convergence of order $O\left(\frac{1}{n\sqrt{T}}\right)$ if $\tau^2 = \operatorname{var}\eta_t = 0$, i.e. in the degenerate case of no intercorrelation.

2.3 Generalisation (GICM)

Up to here, we have divided the processes $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, into a mean process and n residual processes following the same dynamics. A more general class of models is given by decompositions of the form $X_t^{(i)} = Z_t^{(i)} + Y_t$, where $\{Z_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, and $\{Y_t\}_{t\in\mathbb{Z}}$ are stationary autoregressive processes: we now assume that the "mean" or "background" process $\{Y_t\}_{t\in\mathbb{Z}}$ is responsible for the common structure of the panel, and that the time series $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, fluctuate around $\{Y_t\}_{t\in\mathbb{Z}}$.
More specifically, our assumptions are as follows:

2.3.1 ASSUMPTION

(i) The background process $\{Y_t\}_{t\in\mathbb{Z}}$ is a causal Gaussian autoregressive process (assumption 1.1.1), such that

$$b(\mathrm{L})\,Y_t = \upsilon_t \qquad \text{for all } t \in \mathbb{Z},$$

where L is the backward shift operator and $b(\mathrm{L}) = 1 - b_1\,\mathrm{L} - \cdots - b_q\,\mathrm{L}^q$.

(ii) For $i = 1, \dots, n$, $t \in \mathbb{Z}$, let $X_t^{(i)} = Z_t^{(i)} + Y_t$, where the residuals $\{Z_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, are causal and obey

$$a(\mathrm{L})\,Z_t^{(i)} = \zeta_t^{(i)} \qquad \text{for all } t \in \mathbb{Z},\ i = 1, \dots, n,$$

with $a(\mathrm{L}) = 1 - a_1\,\mathrm{L} - \cdots - a_p\,\mathrm{L}^p$.

(iii) The innovations $\{\upsilon_t\}_{t\in\mathbb{Z}}$ and $\{\zeta_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, are Gaussian white noise processes such that

$$\upsilon_t \sim \mathrm{N}(0, \omega_n^2) \qquad \text{for all } t \in \mathbb{Z},$$

where $\lim_{n\to\infty}\omega_n^2 = \omega^2 \geq 0$, and

$$\zeta_t^{(i)} \sim \mathrm{N}(0, \sigma_n^2) \qquad \text{for all } t \in \mathbb{Z},\ i = 1, \dots, n,$$

where $\lim_{n\to\infty} \sigma_n^2 = \sigma_0^2 > 0$.

$\zeta_s^{(i)}$ and υ_t are independent for all $s, t \in \mathbb{Z},\ i = 1, \ldots, n$.

(iv) For $i \neq j$, let $\sigma_n^{ij} = \operatorname{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right)$. We assume that $\lim_{n\to\infty} \sigma_n^{ij} = 0$.

(v) Moreover assume that $\theta_a = (a_1, \ldots, a_p, \sigma_n^2) \in \Theta_a \subset \mathbb{R}^p \times \mathbb{R}_0^+$ and that analogously $\theta_b = (b_1, \ldots, b_q, \omega_n^2) \in \Theta_b \subset \mathbb{R}^q \times \mathbb{R}_0^+$, where Θ_a and Θ_b are compact parameter spaces.

Based on these assumptions, we define a generalised model of intercorrelated time series.

2.3.2 DEFINITION

If assumption 2.3.1 is fulfilled, we call the panel of time series described above the generalised intercorrelation model ("GICM"). *From the GICM we derive the* mean processes

$$\bar{X}_t = \frac{1}{n}\sum_{i=1}^{n} X_t^{(i)} \quad \textit{and} \quad \bar{Z}_t = \frac{1}{n}\sum_{i=1}^{n} Z_t^{(i)}, \quad t \in \mathbb{Z},$$

and the residual processes

$$\mathring{X}_t^{(i)} = \mathring{Z}_t^{(i)} = X_t^{(i)} - \bar{X}_t, \quad t \in \mathbb{Z},\ i = 1, \ldots, n.$$

Moreover we let $\tilde{\sigma}_n^2 = \sigma_n^2 - \sigma_n^{ij}$, $\bar{\zeta}_t = \frac{1}{n}\sum_{i=1}^{n} \zeta_t^{(i)}$ *and* $\mathring{\zeta}_t^{(i)} = \zeta_t^{(i)} - \bar{\zeta}_t$ *for* $t \in \mathbb{Z}$, $i = 1, \ldots, n$.

2.3.3 REMARK

1. Note that σ_n^{ij} does not depend on i and j as the innovations are assumed to be identically distributed. However, we have not assumed $\zeta_t^{(i)}$ and $\zeta_t^{(j)}$ to be independent for $i \neq j$. We just have to guarantee that the intercorrelation is "not too large", because we want to use the mean-square consistency of the panel covariance estimator (lemma 1.2.4) for proving asymptotic normality of our parameter estimators. Usually we moreover assume that $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$. We will see in lemma 2.3.6 that then Y_t can be approximated by $\bar{X}_t$ for all $t \in \mathbb{Z}$ since $\mathbb{E}\left(\bar{X}_t - Y_t\right)^2 = O\left(\frac{1}{n}\right)$. This approximation is used in section 2.4.3 for the estimation of θ_b.

2. Secondly, as all processes are causal, we can represent them as MA(∞) processes (lemma 1.1.2):

$$Z_t^{(i)} = \sum_{u=0}^{\infty} \psi_u\, \zeta_{t-u}^{(i)} \quad \text{and} \quad Y_t = \sum_{u=0}^{\infty} \varphi_u\, \upsilon_{t-u},$$

where $\{\psi_u\}_{u\geq 0}$ and $\{\varphi_u\}_{u\geq 0}$ are absolutely summable. This means that the panel $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, is a special case of a factor model as investigated in Forni et al. (2000):

$$X_t^{(i)} = \sum_{u=0}^{\infty} \psi_u\, \zeta_{t-u}^{(i)} + \sum_{u=0}^{\infty} \varphi_u\, \upsilon_{t-u}, \quad \text{for } t \in \mathbb{Z},\ i = 1, \ldots, n,$$

where $\sum_{u=0}^{\infty} \varphi_u \, \upsilon_{t-u}$ is the common and $\sum_{u=0}^{\infty} \psi_u \, \zeta_{t-u}^{(i)}$ is the idiosyncratic component. Forni et al. (2000) show that their estimator of the common component is consistent for $n, T \to \infty$. In the present case of only one common factor, their method yields $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ as the estimator of the common factor. However, their focus is on estimating the common components and in particular their number, whereas here the main interest is on parameter estimation. Nevertheless, the consistency result for the estimator $\hat{\theta}_b$ of θ_b obtained in theorem 2.4.15 reflects the convergence properties of $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ to the common component (see remark 2.4.16): the convergence behaviour of the estimator of the common factor is discussed in Forni et al. (2001). In particular the authors show that if both n and T tend to infinity, the estimator is consistent, even if the length of the time series grows arbitrarily slow. This behaviour can also be observed in theorem 2.4.15; in the special case treated here the result can be proved directly.

The above assumptions allow for a much broader modelling as we can see in the following examples:

2.3.4 Examples

1. Obviously, we obtain the ICM described in the last section as a special case:

   let $Y_t = \bar{X}_t$ for all $t \in \mathbb{Z}$ and $b(\mathrm{L}) = a(\mathrm{L})$. In the notation of the GICM, we have $\zeta_t^{(i)} = \mathring{\varepsilon}_t^{(i)}$ and $\upsilon_t = \eta_t + \bar{\varepsilon}_t$, i.e. here $Z_t^{(i)} = \mathring{X}_t^{(i)}$. In particular, this implies that $Z_t^{(i)} = \mathring{Z}_t^{(i)}$ because in the ICM $\bar{Z}_t = \frac{1}{n}\sum_{i=1}^{n} \mathring{X}_t^{(i)} = 0$. Moreover it can be seen that $\omega_n^2 = \operatorname{var} \upsilon_t = \tau^2 + \frac{\sigma^2}{n}$, where $\tau^2 = \operatorname{var} \eta_t$ and $\sigma^2 = \operatorname{var} \varepsilon_t^{(i)}$. Thus the notation is consistent. The variance $\sigma_n^2 = \operatorname{var} \zeta_t^{(i)}$ in the GICM corresponds to $\operatorname{var} \mathring{\varepsilon}_t^{(i)} = \frac{n-1}{n}\,\sigma^2$ in the ICM, whereas for $i \neq j$ we have that $\sigma_n^{ij} = \operatorname{cov}\left(\mathring{\zeta}_t^{(i)}, \mathring{\zeta}_t^{(j)}\right)$ in the GICM corresponds to $\operatorname{cov}\left(\mathring{\varepsilon}_t^{(i)}, \mathring{\varepsilon}_t^{(j)}\right) = -\frac{1}{n}\,\sigma^2$. This shows in particular that $\sigma_0^2 = \lim_{n\to\infty} \sigma_n^2 = \sigma^2 = \operatorname{var} \varepsilon_t^{(i)}$.

2. Starting from the ICM, where $a(\mathrm{L})\, X_t^{(i)} = \varepsilon_t^{(i)} + \eta_t$, for $t \in \mathbb{Z}$, $i = 1, \dots, n$, we can derive the simplest form of the GICM by setting $b(\mathrm{L}) = a(\mathrm{L})$, $a(\mathrm{L})Y_t = \eta_t$ and $a(\mathrm{L})Z_t^{(i)} = \varepsilon_t^{(i)}$ for $t \in \mathbb{Z}$. Here, we cannot derive $\{Y_t\}_{t\in\mathbb{Z}}$ directly from the data. However, if n is large, we can approximate $\{Y_t\}_{t\in\mathbb{Z}}$ by $\{\bar{X}_t\}_{t\in\mathbb{Z}}$. Because the $\{\varepsilon_t^{(i)}\}_{t\in\mathbb{Z}}$ are independent for $i = 1, \dots, n$, in this case $\sigma_n^{ij} = 0$.

3. By assumption we always have

   $$a(\mathrm{L})\, X_t^{(i)} = a(\mathrm{L})\, Z_t^{(i)} + a(\mathrm{L})\, Y_t = \zeta_t^{(i)} + a(\mathrm{L})\, Y_t\,.$$

   Thus η_t in the ICM corresponds to $a(\mathrm{L})\, Y_t$ in the GICM.

   We can additionally assume that the processes are linked. Let for example $q > p$ and

   $$b(\mathrm{L}) = c(\mathrm{L})\, a(\mathrm{L})$$

   with $c(\mathrm{L}) = 1 - c_1\, \mathrm{L} - \cdots - c_{q-p}\, \mathrm{L}^{q-p}$, $c(\mathrm{L})$ invertible.
   Then, since

   $$a(\mathrm{L})\, X_t^{(i)} = c(\mathrm{L})^{-1}\, \upsilon_t + \zeta_t^{(i)} \qquad \text{for } t \in \mathbb{Z},\ i = 1, \dots, n,$$

$c(\mathrm{L})^{-1}\,\upsilon_t$ corresponds to η_t and $\zeta_t^{(i)}$ to $\varepsilon_t^{(i)}$. Hence this case is a generalisation of the ICM allowing η_t to be an autoregressive process.

Of course we can also regard the "inverse" linking $a(\mathrm{L}) = c(\mathrm{L})\,b(\mathrm{L})$. We need $p > q$ and get

$$a(\mathrm{L})\,X_t^{(i)} = c(\mathrm{L})\,\upsilon_t + \zeta_t^{(i)} \qquad \text{for } t \in \mathbb{Z},\ i = 1, \ldots, n, .$$

Here $\{\eta_t\}_{t\in\mathbb{Z}}$ corresponds to a finite moving average process.

This last class of examples allows for a large variety of common shocks $\{\eta_t\}_{t\in\mathbb{Z}}$, as autoregressive processes can be used to model very different data. Therefore, also the GICM is very flexible. Since in general $a(L) \neq b(L)$, estimation in the GICM is done separately for $\theta_a = (a_1, \ldots, a_p, \sigma_n^2)$ and $\theta_b = (b_1, \ldots, b_q, \omega_n^2)$. The first parameter is estimated using the residuals $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$; the second one is obtained from $\{Y_t\}_{t\in\mathbb{Z}}$.

2.3.5 REMARK

1. If we are only interested in the parameter of the individual processes, the structure of $\eta_t = a(\mathrm{L})\,Y_t$ does not play a role in the estimation procedure. It is eliminated by the transformation $\mathring{X}_t^{(i)} = X_t^{(i)} - \bar{X}_t$. This will be discussed in more detail in section 2.4.3 (remark 2.4.8), where we derive the conditional log-likelihood functions. We however want to infer about the structure of $\{Y_t\}_{t\in\mathbb{Z}}$, too; thus the assumption of Y_t, $t \in \mathbb{Z}$, being a causal autoregressive process. Furthermore, in the special case of the ICM, including $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ into the analysis leads to an improvement of the estimators. We discuss this effect at the end of section 2.6, which is concerned with the asymptotic properties of the different estimators, in remark 2.6.10.

2. For $\mathring{X}_t^{(i)} = X_t^{(i)} - \bar{X}_t = \mathring{Z}_t^{(i)}$ we now obtain for $t \in \mathbb{Z}$, $i = 1, \ldots, n$, that

$$a(\mathrm{L})\,\mathring{X}_t^{(i)} = a(\mathrm{L})\,Z_t^{(i)} - a(\mathrm{L})\,\frac{1}{n}\sum_{i=1}^{n} Z_t^{(i)} = \zeta_t^{(i)} - \bar{\zeta}_t = \mathring{\zeta}_t^{(i)}\,.$$

3. The variance of $\bar{\zeta}_t$, $t \in \mathbb{Z}$, is

$$\operatorname{var}\bar{\zeta}_t = \frac{1}{n}\,\sigma_n^2 + \frac{n-1}{n}\,\sigma_n^{ij}\,,$$

where $\sigma_n^2 = \operatorname{var}\zeta_t^{(i)}$ and $\sigma_n^{ij} = \operatorname{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right)$, $i \neq j$.

For $\mathring{\zeta}_t^{(i)} = \zeta_t^{(i)} - \bar{\zeta}_t$, $t \in \mathbb{Z}$, $i = 1, \ldots, n$, we have since $\tilde{\sigma}_n^2 = \sigma_n^2 - \sigma_n^{ij}$ that

$$\operatorname{cov}\left(\mathring{\zeta}_t^{(i)}, \mathring{\zeta}_t^{(j)}\right) = \delta_{ij}\,\sigma_n^2 + (1-\delta_{ij})\,\sigma_n^{ij} - \frac{1}{n}\,\sigma_n^2 - \frac{n-1}{n}\,\sigma_n^{ij} = \left(\delta_{ij} - \frac{1}{n}\right)\tilde{\sigma}_n^2,$$

$$\text{and} \quad \mathring{\gamma}_n^{ij}(h) = \operatorname{cov}\left(\mathring{X}_t^{(i)}, \mathring{X}_{t+h}^{(j)}\right) = \Psi(h)\operatorname{cov}\left(\mathring{\zeta}_t^{(i)}, \mathring{\zeta}_t^{(j)}\right)\,,$$

where $\Psi(h) = \sum_{u=0}^{\infty}\,\psi_u\psi_{u+|h|}$.

Note that due to Hölder's inequality we get for $i \neq j$ that

$$\sigma_n^{ij} = \operatorname{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right) = \mathbb{E}\left(\zeta_t^{(i)}\,\zeta_t^{(j)}\right) \leq \sqrt{\mathbb{E}\,\zeta_t^{(i)\,2}}\,\sqrt{\mathbb{E}\,\zeta_t^{(j)\,2}} = \sigma_n^2\,.$$

Thus $\tilde{\sigma}_n^2 \geq 0$ for all n.

4. The autocovariance function of the process $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ depends on ω_n^2 as well. As we have assumed that $\{Y_t\}_{t\in\mathbb{Z}}$ is a stationary autoregressive process, we have $Y_t = \sum_{u=0}^{\infty} \varphi_u\, \upsilon_{t-u}$ for all $t \in \mathbb{Z}$, with $\sum_{u=0}^{\infty} |\varphi_u| < \infty$ (see lemma 1.1.2). Due to the independence of $\{\bar{Z}_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ we obtain with the notation $\Phi(h) = \sum_{u=0}^{\infty} \varphi_u\varphi_{u+|h|}$ that

$$\begin{aligned}\bar{\gamma}_n(h) &= \operatorname{cov}\left(\bar{X}_t, \bar{X}_{t+h}\right) = \Psi(h)\operatorname{var}(\bar{\zeta}_t) + \Phi(h)\,\omega_n^2\\ &= \gamma_{\bar{Z}}(h) + \gamma_Y(h)\,,\end{aligned}$$

where $\gamma_{\bar{Z}}(h)$ and $\gamma_Y(h)$, $h \in \mathbb{Z}$, are the autocovariance functions of $\{\bar{Z}\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$.

If $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$, we can approximate $\{Y_t\}_{t\in\mathbb{Z}}$ by $\{\bar{X}_t\}_{t\in\mathbb{Z}}$. We conclude this section with two results which illustrate the nature of this approximation.

2.3.6 LEMMA
Let $\{\zeta_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1,\ldots,n$, as in assumption 2.3.1. If the covariances fulfil

$$\sigma_n^{ij} = \operatorname{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right) = O\left(\frac{1}{n}\right) \qquad \text{for } i \neq j\,,$$

then

$$\mathbb{E}(\bar{X}_t - Y_t)^2 = \mathbb{E}\,\bar{Z}_t^2 = O\left(\frac{1}{n}\right)\,.$$

PROOF:
Because $\{\bar{X}_t - Y_t\}_{t\in\mathbb{Z}} = \{\bar{Z}_t\}_{t\in\mathbb{Z}}$ is a causal autoregressive process, we can represent it as a MA(∞) process. Using the notations of the preceding remark we get that

$$\bar{X}_t - Y_t = \sum_{u=0}^{\infty} \psi_u\,\bar{\zeta}_{t-u} \qquad \text{for all } t \in \mathbb{Z},$$

where $\{\psi_u\}_{u\geq 0}$ are absolutely summable.
By assumption $\zeta_s^{(i)}$ and $\zeta_t^{(j)}$ are independent for $s \neq t$, $\mathbb{E}\,\zeta_t^{(i)} = 0$ for all $t \in \mathbb{Z}$, $i = 1,\ldots,n$, and $\lim_{n\to\infty}\sigma_n^2 = \sigma^2$.
If $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$, we thus obtain

$$\begin{aligned}\mathbb{E}(\bar{X}_t - Y_t)^2 = \mathbb{E}\,\bar{Z}_t^2 &= \sum_{u,v=0}^{\infty} \psi_u\psi_v\,\frac{1}{n^2}\sum_{i,j=1}^{n} \operatorname{cov}\left(\zeta_{t-u}^{(i)}, \zeta_{t-v}^{(j)}\right)\\ &= \sum_{u=0}^{\infty}\psi_u^2\left(\frac{1}{n}\,\sigma_n^2 + \frac{n-1}{n}\,\sigma_n^{ij}\right) = O\left(\frac{1}{n}\right)\,.\end{aligned}$$

□

We can use this approximation in many important models. The condition is trivially fulfilled if $\zeta_t^{(i)}$ and $\zeta_t^{(j)}$ are independent for $i \neq j$. But it is also possible to approximate a variety of cases where the intercorrelation between different time series in the panel is small enough, e.g. in the ICM. There (see example 2.3.4) σ_n^{ij} corresponds to $\operatorname{cov}\left(\mathring{\varepsilon}_t^{(i)}, \mathring{\varepsilon}_t^{(j)}\right) = -\frac{1}{n} \operatorname{var} \varepsilon_t^{(i)} = O\left(\frac{1}{n}\right)$ for $i \neq j$ and in this case even $\bar{Z}_t = 0$ for all $t \in \mathbb{Z}$.

We also can compute the variance of $b(\mathrm{L})\, \bar{X}_t$ in the GICM.

2.3.7 PROPOSITION

Under the assumptions of the GICM (2.3.1), we obtain

$$\operatorname{var}\left(b(\mathrm{L})\, \bar{X}_t\right) = \omega_n^2 + \left(\frac{\sigma_n^2}{n} + \frac{n-1}{n}\, \sigma_n^{ij}\right) \frac{1}{2\,\pi} \int_{-\pi}^{\pi} \frac{|b(\exp(-i\lambda))|^2}{|a(\exp(-i\lambda))|^2}\, d\lambda\,,$$

where $\sigma_n^2 = \operatorname{var} \zeta_t^{(i)}$, $\sigma_n^{ij} = \operatorname{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right)$ *and* $\omega_n^2 = \operatorname{var} \upsilon_t$.

PROOF:

As we have assumed $\{Z_t^{(i)}\}_{t\in\mathbb{Z}} = \{X_t^{(i)} - Y_t\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, to be causal, $a(\mathrm{L})$ is invertible. Therefore, we get for $\{\bar{X}_t\}_{t\in\mathbb{Z}} = \{Y_t + \bar{Z}_t\}_{t\in\mathbb{Z}}$ that

$$b(\mathrm{L})\, \bar{X}_t = \upsilon_t + b(\mathrm{L})\, a(\mathrm{L})^{-1} \bar{\zeta}_t \qquad \text{for all } t \in \mathbb{Z}.$$

υ_t and $\bar{\zeta}_t$ are independent, thus

$$\begin{aligned} \mathbb{E}(b(\mathrm{L})\, \bar{X}_t)^2 &= \omega_n^2 + \operatorname{var}(b(\mathrm{L})\, a(\mathrm{L})^{-1} \bar{\zeta}_t) \\ &= \omega_n^2 + \operatorname{var}(\bar{\zeta}_t)\, \frac{1}{2\,\pi} \int_{-\pi}^{\pi} \frac{|b(\exp(-i\lambda))|^2}{|a(\exp(-i\lambda))|^2}\, d\lambda \end{aligned}$$

(see for example Brockwell and Davis 1991, p. 123).

From remark 2.3.5 above we know that $\operatorname{var} \bar{\zeta}_t = \frac{1}{n}\, \sigma_n^2 + \frac{n-1}{n}\, \sigma_n^{ij}$. □

2.3.8 REMARK

1. The backward shift operators $a(\mathrm{L})$ and $b(\mathrm{L})$ do not depend on the number of time series n. Denoting $\omega_{\bar{X}}^2 = \operatorname{var}\left(b(\mathrm{L})\, \bar{X}_t\right)$, the proposition thus shows that $\omega_{\bar{X}}^2 - \omega_n^2 = O\left(\frac{1}{n}\right)$ if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$. In this case we therefore can approximate $\omega_{\bar{X}}^2 \approx \omega_n^2 = \operatorname{var} \upsilon_t$ if $n \to \infty$. So we have an explicit expression of the error term in the approximation which we will use for the parameter estimation in the GICM (see section 2.4.3).

2. In the ICM, where $a = b$, we have $\omega_n^2 = \operatorname{var} \upsilon_t = \operatorname{var}(\eta_t + \bar{\varepsilon}_t) = \tau^2 + \frac{\sigma^2}{n}$, where $\tau^2 = \operatorname{var} \eta_t$ and $\sigma^2 = \operatorname{var} \varepsilon_t^{(i)}$ (example 2.3.4). Thus the notation is consistent with the previous section where we have denoted $\tau^2 + \frac{\sigma^2}{n} = \omega_n^2$. As in this case both $\frac{n-1}{n}\, \sigma_n^{ij} = \frac{n-1}{n} \times \left(-\frac{\sigma^2}{n}\right)$ and $-\frac{\sigma_n^2}{n} = \frac{n-1}{n} \times \left(-\frac{\sigma^2}{n}\right)$, the second term in the above representation of $\operatorname{var}\left(b(\mathrm{L})\, \bar{X}_t\right)$ cancels out.

2.4 Conditional Maximum Likelihood Estimation

2.4.1 Factorisation of the Log-Likelihood in the ICM

In the ICM $\bar{\varepsilon}_t$ and $\mathring{\varepsilon}_t^{(i)}$ and therefore also $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ and $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$ are independent for $i = 1, \ldots, n$ (remark 2.2.3). This implies a possible factorisation of the conditional likelihood function. We indeed obtain a closed form of the conditional log-likelihood, which is one of our main results. It allows including the information contained in the mean process $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ into the estimation procedure, and thus to improve the estimator in the setting of the ICM. The original procedure used in Hjellvik and Tjøstheim (1999a) is based only on the residual processes $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$. There the estimators are obtained by minimising the conditional log-likelihood function $\mathcal{L}^{\circ}_{n,T}(\theta)$, which we use for estimation in the GICM (see section 2.4.3). The differences between these estimators are discussed in remarks 2.4.8 and 2.6.10.

2.4.1 NOTATIONS
In order to facilitate the notation, let

$$\mathbf{X}_t = \left(X_t^{(1)}, \ldots, X_t^{(n)}\right)' \quad \text{for } t \in \mathbb{Z}$$

and denote the parameter of the ICM by $\theta = (a_1, \ldots, a_p, \sigma^2, \tau^2)$. We study the conditional log-likelihood

$$\mathcal{L}_{n,T}(\theta) = -\frac{2}{n(T-p)} \log \mathcal{L}(\mathbf{X}_{p+1}, \ldots, \mathbf{X}_T \mid \mathbf{X}_1, \ldots, \mathbf{X}_p)$$

derived from $\mathcal{L}(\mathbf{X}_{p+1}, \ldots, \mathbf{X}_T | \mathbf{X}_1, \ldots, \mathbf{X}_p)$, the conditional likelihood function given $\mathbf{X}_1, \ldots, \mathbf{X}_p$.

We can obtain the factorisation of the conditional likelihood function in the following way; the proof is based upon an idea of Dahlhaus (1999).

2.4.2 THEOREM
Under assumption 2.2.1 and using the above notations, we obtain the conditional log-likelihood function depending on the parameter $\theta = (a_1, \ldots, a_p, \sigma^2, \tau^2)'$ of the intercorrelation model as

$$\begin{aligned}\mathcal{L}_{n,T}(\theta) = \frac{n-1}{n} \log \sigma^2 &+ \frac{1}{\sigma^2} \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \left(a(\mathrm{L})\,\mathring{X}_t^{(i)}\right)^2 \\ &+ \frac{1}{n} \log \omega_n^2 + \frac{1}{\omega_n^2} \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \left(a(\mathrm{L})\,\bar{X}_t\right)^2 \\ &+ \frac{1}{n} \log n + \log(2\pi)\ ,\end{aligned}$$

where $\omega_n^2 = \tau^2 + \frac{1}{n}\sigma^2$.

PROOF:
In the trivial case of $n = 1$, we have $\mathring{X}_t^{(1)} = 0$, $\bar{X}_t = X_t^{(1)}$ and $\mathrm{var}(\mathbf{X}_t) = \sigma^2 + \tau^2 = \omega_1^2$. It is easily seen that $\mathcal{L}_{n,T}(\theta)$ can be written in the above form.

For $n > 1$ we regard the $(n-1)$-dimensional vector $\mathring{\mathbf{X}}_t = (\mathring{X}_t^{(1)}, \ldots, \mathring{X}_t^{(n-1)})$, since

$$\sum_{i=1}^{n-1} \mathring{X}_t^{(i)} = -\mathring{X}_t^{(n)} .$$

Furthermore observe that

$$(\mathring{X}_t^{(1)}, \ldots, \mathring{X}_t^{(n-1)}, \bar{X}_t) = S\,(X_t^{(1)}, \ldots, X_t^{(n)})$$

where the transformation matrix S fulfils $|S| = \frac{1}{n}$.
The transformation theorem therefore leads to

$$\begin{aligned}
&\mathcal{L}(\mathbf{X}_{p+1}, \ldots, \mathbf{X}_T \mid \mathbf{X}_1, \ldots, \mathbf{X}_p) = \prod_{t=p+1}^{T} \mathrm{f}_X(\mathbf{X}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}) \\
&= \prod_{t=p+1}^{T} \frac{1}{n}\, \mathrm{f}_{\bar{X}_t, \mathring{X}_t}(\bar{X}_t, \mathring{\mathbf{X}}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}) \\
&= \prod_{t=p+1}^{T} \frac{1}{n}\, \mathrm{f}_{\bar{X}_t}(\bar{X}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1})\, \mathrm{f}_{\mathring{X}_t}(\mathring{\mathbf{X}}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}) \\
&= \prod_{t=p+1}^{T} \frac{1}{n}\, \frac{1}{\sqrt{2\pi}\sqrt{\omega_n^2}}\, \exp\left(-\frac{1}{2\,\omega_n^2} \left(\bar{X}_t - \sum_{k=1}^{p} a_k \bar{X}_{t-k} \right)^2 \right) \\
&\times \frac{1}{\sqrt{(2\pi)^{(n-1)}}\sqrt{|\tilde{\Sigma}|}}\, \exp\left(-\frac{1}{2} \left(\mathring{\mathbf{X}}_t - \sum_{k=1}^{p} a_k \mathring{\mathbf{X}}_k \right)' \tilde{\Sigma}^{-1} \left(\mathring{\mathbf{X}}_t - \sum_{k=1}^{p} a_k \mathring{\mathbf{X}}_k \right) \right) ,
\end{aligned}$$

since

$$\mathbb{E}(\bar{X}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}) = \sum_{k=1}^{p} a_k \bar{X}_{t-k} \quad \text{and} \quad \mathbb{E}(\mathring{\mathbf{X}}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}) = \sum_{k=1}^{p} a_k \mathring{\mathbf{X}}_k .$$

Here, $\omega_n^2 = \mathrm{var}(\bar{X}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}) = \tau^2 + \frac{1}{n}\sigma^2$ and $\tilde{\Sigma} = \mathrm{var}(\mathring{\mathbf{X}}_t \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1})$ is the conditional covariance matrix of $\mathring{\mathbf{X}}_t$. The factorisation of the conditional densities is due to the independence of $\bar{X}_t$ and $\mathring{X}_t^{(i)}$ for all $t \in \mathbb{Z}$, $i = 1, \ldots, n$.
It is easily seen that $\mathrm{cov}(\mathring{X}_t^{(i)}, \mathring{X}_t^{(j)} \mid \mathbf{X}_1, \ldots, \mathbf{X}_{t-1}) = \mathrm{cov}\left(\mathring{\varepsilon}_t^{(i)}, \mathring{\varepsilon}_t^{(j)}\right) = \left(\delta_{ij} - \frac{1}{n}\right)\sigma^2$.
Therefore we get, if we denote the $(n-1)\times(n-1)$-matrix consisting of ones by $\mathbb{1}_{n-1}$,

$$\tilde{\Sigma} = (I_{n-1} - \frac{1}{n}\mathbb{1}_{n-1})\,\sigma^2 ; \qquad \text{and thus} \quad \tilde{\Sigma}^{-1} = (I_{n-1} + \mathbb{1}_{n-1})\,\frac{1}{\sigma^2} .$$

By recursively calculating the determinant of $\tilde{\Sigma}$ we furthermore obtain

$$|\tilde{\Sigma}| = \left(\left(1 - \frac{1}{n}\right) - (n-2)\frac{1}{n} \right) \sigma^{2(n-1)} = \frac{1}{n}\,\sigma^{2(n-1)} .$$

Taking the logarithm leads to the stated form of

$$\mathcal{L}_{n,T}(\theta) = -\frac{2}{n(T-p)} \log \mathcal{L}(\mathbf{X}_{p+1}, \ldots, \mathbf{X}_T \mid \mathbf{X}_1, \ldots, \mathbf{X}_p) . \qquad \square$$

Minimising $\mathcal{L}_{n,T}$, we obtain the conditional maximum likelihood estimator $\hat{\theta}_{n,T}$. In the subsequent sections, we present an algorithm for the computation and derive its asymptotic properties.

2.4.2 The Minimisation Algorithm

We want to estimate the parameters of the ICM by minimising $\mathcal{L}_{n,T}$. This cannot be achieved directly, as we can see from its derivatives.

2.4.3 REMARK
Following the notations of the preceding theorem 2.4.2, denote

$$A_{n,T}(a) = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \Big(a(L) \mathring{X}_t^{(i)} \Big)^2$$

$$\text{and} \qquad B_{n,T}(a) = \frac{1}{T-p} \sum_{t=p+1}^{T} \big(a(L) \bar{X}_t \big)^2 \,.$$

Since $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n}$, the partial derivatives of $\mathcal{L}_{n,T}$ are for $l = 1, \ldots, p$

$$\frac{\partial}{\partial a_l} \mathcal{L}_{n,T}(\theta) = -\frac{2}{\sigma^2} \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \Big(a(L)\, \mathring{X}_t^{(i)} \Big)\, \mathring{X}_{t-l}^{(i)} - \frac{2}{\omega_n^2} \frac{1}{n(T-p)} \sum_{t=p+1}^{T} \big(a(L)\, \bar{X}_t \big)\, \bar{X}_{t-l}\,,$$

$$\frac{\partial}{\partial \tau^2} \mathcal{L}_{n,T}(\theta) = \frac{1}{n\,\omega_n^2} - \frac{1}{n\,\omega_n^4} B_{n,T}(a)$$

$$\text{and} \quad \frac{\partial}{\partial \sigma^2} \mathcal{L}_{n,T}(\theta) = \frac{n-1}{n\,\sigma^2} - \frac{1}{\sigma^4} A_{n,T}(a) + \frac{1}{n^2\,\omega_n^2} \left(1 - \frac{1}{\omega_n^2} B_{n,T}(a) \right) .$$

If $a = (a_1, \ldots, a_p)'$ is given, minimising leads to the estimator

$$\hat{\omega}_n^2 = B_{n,T}(a)\,,$$

and by plugging in $\hat{\omega}_n^2$ for ω_n^2, we can calculate

$$\hat{\sigma}^2 = \frac{n}{n-1} A_{n,T}(a)\,.$$

$\hat{\tau}^2$ then can be obtained from $\hat{\sigma}^2$ and $\hat{\omega}^2$. Note that if $\hat{\sigma}^2$ is fixed, choosing $\hat{\tau}^2$ such that $\mathcal{L}_{n,T}(\theta)$ is minimal corresponds to choosing $\hat{\omega}^2$ such that $\mathcal{L}_{n,T}(\theta)$ is minimal. We then can calculate $\hat{a}$ conditional on $\hat{\sigma}^2$ and $\hat{\omega}^2$ as

$$\hat{a} = \hat{a}(\hat{\sigma}^2, \hat{\omega}^2) = \Big(\frac{1}{\hat{\sigma}^2} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{\mathbf{x}}_{t-1}^{(i)}\, \mathring{\mathbf{x}}_{t-1}^{(i)\,\prime} + \frac{1}{n\,\hat{\omega}_n^2} \sum_{t=p+1}^{T} \bar{\mathbf{x}}_{t-1}\, \bar{\mathbf{x}}_{t-1}' \Big)^{-1} \times \left(\frac{1}{\hat{\sigma}^2} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{X}_t^{(i)}\, \mathring{\mathbf{x}}_{t-1}^{(i)} + \frac{1}{n\,\hat{\omega}_n^2} \sum_{t=p+1}^{T} \bar{X}_t\, \bar{\mathbf{x}}_{t-1} \right) ,$$

where $\mathring{\mathbf{x}}_{t-1}^{(i)} = (\mathring{X}_{t-1}^{(i)}, \ldots, \mathring{X}_{t-p}^{(i)})'$, $i = 1, \ldots, n$, and $\bar{\mathbf{x}}_{t-1} = (\bar{X}_{t-1}, \ldots, \bar{X}_{t-p})'$.

For the estimation we thus use the following recursive algorithm, which is similar to the multistep procedure suggested in (Hsiao 1986, p. 55). As initial values for the variances we set $\hat{\sigma}_0^2 = \hat{\omega}_{n,0}^2 = 1$ and calculate estimates of the coefficients a_k, $k = 1, \ldots, p$. From the values we get we can in turn derive new estimates $\hat{\sigma}_1^2$ and $\hat{\omega}_{n,1}^2$ of the variances. This procedure then is iterated.

2.4.4 ALGORITHM
Let $\Theta \subset \mathbb{R}^{p+2}$ be a compact parameter space such that $\sigma^2 \geq c > 0$ and $\tau^2 \geq c_2 \geq 0$ for all $\theta = (a_1, \ldots, a_p, \sigma^2, \tau^2)' \in \Theta$. Denote $a = (a_1, \ldots, a_p)'$. Furthermore let the conditional log-likelihood function $\mathcal{L}_{n,T}(\theta)$ for $\theta \in \Theta$ be as defined in theorem 2.4.2. Then $\hat{\theta}_{n,T} = \text{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta)$ can be obtained as follows:

1. Let $\nu = 0$, $\hat{\sigma}_\nu^2 = \hat{\omega}_{n,\nu}^2 = 1$ and $\hat{\tau}^2 = \hat{\omega}_{n,\nu}^2 - \frac{\hat{\sigma}_\nu^2}{n}$.
2. Let $\hat{a}_{\nu+1}$ such that $\hat{\theta}_{\nu+1} = (\hat{\alpha}'_{\nu+1}, \hat{\sigma}_\nu^2, \hat{\tau}_\nu)' = \text{argmin}_{\{\theta\in\Theta|\sigma^2=\hat{\sigma}_\nu^2, \tau^2=\hat{\tau}_\nu^2\}}\mathcal{L}_{n,T}(\theta)$.
3. Let $\hat{\omega}_{n,\nu+1}^2 = \hat{\tau}_{\nu+1}^2 + \frac{\hat{\sigma}_\nu^2}{n}$ such that
   $(\hat{\alpha}'_{\nu+1}, \hat{\sigma}_\nu^2, \hat{\tau}_{\nu+1})' = \text{argmin}_{\{\theta\in\Theta|a=\hat{a}_{\nu+1}, \sigma^2=\hat{\sigma}_\nu^2\}}\mathcal{L}_{n,T}(\theta)$,
   i.e. if $B_{n,T}(\hat{a}_{\nu+1}) \geq c_2 + \frac{\hat{\sigma}_\nu^2}{n}$, then $\hat{\omega}_{n,\nu+1}^2 = B_{n,T}(\hat{a}_{\nu+1})$.
4. Let $\hat{\sigma}_{\nu+1}^2$ such that $(\hat{\alpha}'_{\nu+1}, \hat{\sigma}_{\nu+1}^2, \hat{\tau}_{\nu+1})' = \text{argmin}_{\{\theta\in\Theta|a=\hat{a}_{\nu+1}, \tau^2=\hat{\tau}_{\nu+1}^2\}}\mathcal{L}_{n,T}(\theta)$,
   i.e. if $A_{n,T}(\hat{a}_{\nu+1}) \geq c$, we have $\hat{\sigma}_{\nu+1}^2 = \frac{n}{n-1}\, A_{n,T}(\hat{a}_{\nu+1})$.
5. Iterate step 2) to 4) until convergence is attained.
6. Compute $\hat{\tau}^2$ as $\hat{\tau}^2 = \hat{\omega}_{n,\nu}^2 - \frac{\hat{\sigma}_\nu^2}{n}$ and denote the obtained conditional maximum likelihood estimator by $\hat{\theta}_{n,T} = (\hat{a}'_\nu, \hat{\sigma}_\nu^2, \hat{\tau}^2)' = \text{argmin}_{\theta\in\Theta}\, \mathcal{L}_{n,T}(\theta)$.

2.4.5 REMARK
1. Our criterion for stopping the algorithm is that the distance between two consecutive estimates becomes small. To be more specific, we use the conditions $||\hat{a}_{\nu+1} - \hat{a}_\nu|| < \delta$ and $||(\hat{\sigma}_{\nu+1}^2, \hat{\omega}_{n,\nu+1}^2)' - (\hat{\sigma}_\nu^2, \hat{\omega}_{n,\nu}^2)'|| < \varepsilon$ for some $\delta > 0$ and $\varepsilon > 0$. In the simulations we have set $\delta = \varepsilon = 10^{-3}$. If this condition is not fulfilled after a fixed number of iterations, the algorithm stops with an error message. Note that due to the restriction of the parameter space, we have $\hat{\sigma}^2 \geq c$, $\hat{\tau}^2 \geq 0$ and $\hat{\omega}_n^2 \geq \frac{c}{n}$.
2. The algorithm is based on a successive minimising of the log-likelihood function, thus $\mathcal{L}_{n,T}(\hat{\theta}_\nu)$ decreases monotonically for $\nu \to \infty$. In order to eliminate the risk that the algorithm stops in a saddle point $\tilde{\theta}$, we could perform more iterations with slightly pertubed parameters $\tilde{\theta} + \delta$. This guarantees that we reach a (local) minimum $\hat{\theta}_{n,T}$, but it cannot easily be excluded by theoretical arguments that the algorithm oszillates between several local minima (Drton and Eichler 2004). We prove however in section 2.5.3 that asymptotically the minimum is unique if the parameter space is chosen appropriately.

3. Numerical simulations have shown that this algorithm works well. Even without taking saddle points into account, the algorithm usually converged in our simulations after 6 or 7 iterations to the true value. The simulation results are discussed in the Appendix in section A.

2.4.3 Parameter Estimation in the GICM

In the case of the generalised model, where the coefficients of the common factor and the residuals are not necessarily identical, the situation is different. Here we can calculate the likelihood functions separately. $\theta_a = (a_1, \ldots, a_p, \tilde{\sigma}_n^2)'$ can be estimated directly from the transformed processes $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}} = \{\mathring{Z}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$. If the background process $\{Y_t\}_{t\in\mathbb{Z}}$ is not observable, we must use $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ as an approximation to $\{Y_t\}_{t\in\mathbb{Z}}$ in order to estimate $\theta_b = (b_1, \ldots, b_q, \omega_n^2)'$. We show that in this case we get consistent estimators if, besides their length T, the number n of the time series in the panel tends to infinity and if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$. Of course, if $n = 1$, $X_t = \bar{X}_t = Z_t + Y_t$, and parameter estimation just makes sense if $(a_1, \ldots, a_p) = (b_1, \ldots, b_q)$. Thus we assume throughout this section that $n \geq 2$.

Estimation of θ_a

First, we consider the parameters of the individual effects. Here, since $\mathring{X}_t^{(i)} = \mathring{Z}_t^{(i)}$, we have $a(\mathrm{L})\, \mathring{X}_t^{(i)} = \mathring{\zeta}_t^{(i)} = \zeta_t^{(i)} - \bar{\zeta}_t$ for all $t \in \mathbb{Z}$, $i = 1, \ldots, n$. As in section 2.4.1 denote $\mathring{\mathbf{X}}_t = (\mathring{X}_t^{(1)}, \ldots, \mathring{X}_t^{(n-1)})'$. We can derive the conditional log-likelihood if n is large enough.

2.4.6 PROPOSITION
Under the assumptions of the GICM (2.3.1), there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the conditional log-likelihood of the $\mathring{\mathbf{X}}_t$, $p < t \leq T$, given $\mathring{\mathbf{X}}_1, \ldots, \mathring{\mathbf{X}}_p$ is

$$\begin{aligned}
\mathcal{L}^{\circ}_{n,T}(\theta_a) &= -\frac{2}{n\,(T-p)} \log \mathcal{L}(\mathring{\mathbf{X}}_{p+1}, \ldots, \mathring{\mathbf{X}}_T \mid \mathring{\mathbf{X}}_1, \ldots, \mathring{\mathbf{X}}_p) \\
&= \frac{n-1}{n} \log \tilde{\sigma}_n^2 + \frac{1}{\tilde{\sigma}_n^2} \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \left(a(\mathrm{L})\, \mathring{X}_t^{(i)} \right)^2 \\
&\quad + \frac{n-1}{n} \log(2\pi) - \frac{1}{n} \log n\,,
\end{aligned}$$

where $\tilde{\sigma}_n^2 = \sigma_n^2 - \sigma_n^{ij}$.

PROOF:
Since $\mathrm{cov}\left(\mathring{\zeta}_t^{(i)}, \mathring{\zeta}_t^{(j)}\right) = \left(\delta_{ij} - \frac{1}{n}\right)\tilde{\sigma}_n^2$ (remark 2.3.5), the conditional covariance matrix of $\mathring{\mathbf{X}}_t$, given $\mathring{\mathbf{X}}_1, \ldots, \mathring{\mathbf{X}}_{t-1}$, is $\tilde{\Sigma} = (I_{n-1} - \frac{1}{n}\mathbb{1})\,\tilde{\sigma}_n^2$. As we have shown in remark 2.3.5 that $\tilde{\sigma}_n^2 \geq 0$ for all $n \in \mathbb{N}$ and as by assumption $\sigma_n^2 \to \sigma^2 > 0$ and $\sigma_n^{ij} \to 0$ for $n \to \infty$, there exists a n_0 such that $\tilde{\sigma}_n^2 > 0$ for all $n \geq n_0$.

Analogously to the proof of proposition 2.4.2, we therefore obtain that

$$\mathcal{L}\Big(\mathring{\mathbf{X}}_{p+1},\ldots,\mathring{\mathbf{X}}_T \mid \mathring{\mathbf{X}}_1,\ldots,\mathring{\mathbf{X}}_p\Big) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{(2\pi)^{(n-1)}}\sqrt{\frac{1}{n}\,\tilde{\sigma}_n^{2\,(n-1)}}} \exp\left(-\frac{1}{2\tilde{\sigma}_n^2}\sum_{i=1}^{n}\Big(a(\mathrm{L})\mathring{X}_t^{(i)}\Big)^2\right).$$

It is easily seen that this yields the form of $\mathcal{L}^\circ_{n,T}(\theta_a)$ stated above. □

Minimising $\mathcal{L}^\circ_{n,T}$ leads to consistent estimators of a and $\tilde{\sigma}_n^2$.

2.4.7 PROPOSITION
In the setting of the preceding proposition, $\mathcal{L}^\circ_{n,T}$ is minimised by $\hat{\theta}_a = (\hat{a}_1,\ldots,\hat{a}_p,\hat{\sigma}_n^2)'$, where $\hat{a} = (\hat{a}_1,\ldots,\hat{a}_p)'$ is given by

$$\hat{a} = \left(\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{\mathbf{x}}_{t-1}^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)\,\prime}\right)^{-1}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{X}_t^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)},$$

denoting $\mathring{\mathbf{x}}_{t-1}^{(i)} = (\mathring{X}_{t-1}^{(i)},\ldots,\mathring{X}_{t-p}^{(i)})'$, $i = 1,\ldots,n$.
The variance is obtained as

$$\hat{\sigma}_n^2 = \frac{1}{(n-1)(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\Big(\hat{a}(\mathrm{L})\,\mathring{X}_t^{(i)}\Big)^2.$$

Then the estimator $\hat{\theta}_a = (\hat{a}',\hat{\sigma}_n^2)$ is consistent:

$$\hat{a} - a = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$$

$$\text{and} \qquad \hat{\sigma}_n^2 - \tilde{\sigma}_n^2 = O_P\left(\frac{1}{\sqrt{n\,T}}\right).$$

PROOF:
The above stated form of the estimators is directly obtained by minimising the conditional log-likelihood function $\mathcal{L}^\circ_{n,T}(\theta)$. Thus it remains to show consistency. By assumption $\operatorname{cov}\Big(\mathring{X}_t^{(i)},\mathring{X}_{t+h}^{(j)}\Big) = \big(\delta_{ij} - \frac{1}{n}\big)\,\Psi(h)\,\tilde{\sigma}_n^2$ (see remark 2.3.5). Therefore $\sigma_n^{ij} = \operatorname{cov}\Big(\mathring{X}_t^{(i)},\mathring{X}_{t+h}^{(j)}\Big)$ for $i \neq j$ fulfils $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$. We thus get due to the mean-square convergence of the panel covariance estimator (lemma 1.2.4) that

$$\frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(i)} - \frac{n-1}{n}\,\Psi(k-l)\,\tilde{\sigma}_n^2 = O_P\left(\frac{1}{\sqrt{n\,T}}\right).$$

For ease of notation let

$$\hat{A} = \frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{X}_t^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)}, \qquad \hat{B} = \sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{\mathbf{x}}_{t-1}^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)\,\prime},$$

$$A = \mathbb{E}\,\hat{A} = \frac{n-1}{n}\;(\Psi(k))_{k=1,\dots,p}\;\tilde{\sigma}_n^2$$
and
$$B = \mathbb{E}\,\hat{B} = \frac{n-1}{n}\;(\Psi(k-l))_{k,l=1,\dots,p}\;\tilde{\sigma}_n^2\,.$$

Then the above implies that $\hat{A} - A = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$ and $\hat{B} - B = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$. Standard theory (e.g. Brockwell and Davis (1991)) yields that the true parameter a fulfils the Yule-Walker equation $B\,a = A$. As $\hat{B}\,\hat{a} = \hat{A}$ and moreover $\hat{a} = O_P(1)$ due to the compactness of Θ, we thus get that

$$B\,(\hat{a}-a) = B\,\hat{a} - A + \hat{A} - \hat{B}\,\hat{a} = \big(B - \hat{B}\big)\,\hat{a} + \hat{A} - A = O_P\left(\frac{1}{\sqrt{n\,T}}\right)\,.$$

Thus $\hat{a} - a = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$. From the consistency of $\hat{a} - a$ we can in turn conclude that, choosing $a_0 = -1 = \hat{a}_0$,

$$\hat{\sigma}^2 - \tilde{\sigma}_n^2 = \frac{1}{(n-1)\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \sum_{k,l=0}^{p} \hat{a}_k\,\hat{a}_l \left[\mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(i)} - \frac{n-1}{n}\,\Psi(k-l)\tilde{\sigma}_n^2\right] + \sum_{k,l=0}^{p} (\hat{a}_k\,\hat{a}_l - a_k\,a_l)\,\Psi(k-l)\,\tilde{\sigma}_n^2 = O_P\left(\frac{1}{\sqrt{n\,T}}\right)\,,$$

which completes the proof. □

2.4.8 REMARK

1. In general, using $\mathcal{L}_{n,T}^{\circ}$ we cannot estimate $\operatorname{var}\zeta_t^{(i)} = \sigma_n^2$, but $\tilde{\sigma}_n^2 = \sigma_n^2 - \sigma_n^{ij}$, which for $n \to \infty$ tends to $\sigma^2 > 0$. However, in the special case of the ICM, $\tilde{\sigma}_n^2 = (1 - \frac{1}{n})\,\sigma^2 - (-\frac{1}{n})\,\sigma^2 = \sigma^2 > 0$ for all $n \in \mathbb{N}$, where $\sigma^2 = \operatorname{var}\varepsilon_t^{(i)}$ (see example 2.3.4(i)). Thus here $\hat{\theta}_a$ is a consistent estimator of the true parameter $\theta_a = (a_1,\dots,a_p,\sigma^2)'$ even if $T \to \infty$ and n is fixed. $\hat{a} = (\hat{a}_1,\dots,\hat{a}_p)' = \hat{a}_{HT}$ is the estimator $\tilde{a}$ of Hjellvik and Tjøstheim (1999a). Their model is the ICM, but they obtain their estimator by treating η_t as a nuisance parameter which they eliminate using the transformation $\mathring{X}_t^{(i)} = X_t^{(i)} - \bar{X}_t$. This leads to the conditional likelihood function $\mathcal{L}_{n,T}^{\circ}$. The advantage of their procedure is that no assumptions on the structure of η_t have to be made. It even can be deterministic, which corresponds to the case of $\tau^2 = 0$. If we are interested in the structure of the background process, assumptions on the distribution of $\{\eta_t\}_{t\in\mathbb{Z}}$ are however needed. Hjellvik and Tjøstheim furthermore assume that the individual innovation processes are independent, which means that the cross-correlation is entirely induced by $\{\eta_t\}_{t\in\mathbb{Z}}$. This assumption is relaxed in the GICM.

2. Asymptotic normality of $\hat{\theta}_a$ is studied in section 2.5.3. If we restrict our model to the ICM, asymptotic normality can be directly obtained by employing the independence of the processes $\{\varepsilon_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1,\dots,n$. If $n \to \infty$, T fixed, the proof is an application of the standard central limit theorem for independently and identically distributed observations. If $T \to \infty$, asymptotic normality is due

to the α-mixing property of autoregressive processes. Hjellvik and Tjøstheim (1999a) have already stated these properties. For proving asymptotic normality in the general case, we however use a similar method as for the ICM. This is necessary due to the more complicated intercorrelation structure. We here cannot write the random variables $\mathring{\zeta}_t^{(i)}$ as $\mathring{\zeta}_t^{(i)} = \mathring{\varepsilon}_t^{(i)} = \varepsilon_t^{(i)} - \bar{\varepsilon}_t$, $i = 1, \ldots, n$, where $\varepsilon_t^{(i)}$ and $\varepsilon_t^{(j)}$ are independent for $i \neq j$. Thus it is impossible to exploit the independence property in the proof. In the present work, the asymptotic distribution of the parameter estimators in the special case of the ICM is then deduced from the general case in corollary 2.5.35.

Estimation of θ_b

The background process $\{Y_t\}_{t\in\mathbb{Z}}$ is not observable in general. In this case, inference about its parameters is based on approximations. The conditional log-likelihood of the background process $\{Y_t\}_{t\in\mathbb{Z}}$ itself can be derived easily and the asymptotic properties follow from standard theory.

2.4.9 LEMMA
The conditional log-likelihood of the process $Y_t = \sum_{u=0}^{\infty} \varphi_u \upsilon_{t-u}$, $t \in \mathbb{Z}$, (assumption 2.3.1) is

$$\begin{aligned}\mathcal{L}_{n,T}^{Y}(\theta_b) &= -\frac{2}{T-q} \log \mathcal{L}(Y_{q+1}, \ldots, Y_T \mid Y_1, \ldots Y_q) \\ &= \log \omega_n^2 + \log(2\pi) + \frac{1}{\omega_n^2\,(T-q)} \sum_{t=q+1}^{T} \left(b(\mathrm{L})\, Y_t\right)^2\end{aligned}$$

and $\hat{\theta}_b = (\hat{b}_1, \ldots, \hat{b}_q, \hat{\omega}_n^2) = \operatorname{argmin}_{\theta\in\Theta_b} \mathcal{L}_{n,T}^{Y}(\theta)$ is a consistent estimator of the true parameter $\theta_b = (b_1, \ldots, b_q, \omega_n^2)' = (b', \omega_n^2)'$.
Furthermore $\hat{b}_Y = (\hat{b}_1, \ldots, \hat{b}_q)'$ is asymptotically normal with

$$\sqrt{T-q}\,\left(\hat{b}_Y - b\right) \Rightarrow N_Y \quad \textit{for } T \to \infty,$$

where $N_Y \sim \mathrm{N}(0, \Sigma_Y)$ with $\Sigma_Y = (\Phi(k-l))_{k,l=1,\ldots,q}$.

PROOF:
Minimising $\mathcal{L}_{n,T}^{Y}$, we get as estimator of $b = (b_1, \ldots, b_q)'$

$$\hat{b}_Y = \left(\sum_{t=p+1}^{T} \mathbf{y}_{t-1}\, \mathbf{y}_{t-1}'\right)^{-1} \sum_{t=p+1}^{T} Y_t\, \mathbf{y}_{t-1}\,,$$

where $\mathbf{y}_{t-1} = (Y_{t-1}, \ldots, Y_{t-p})'$; and as estimator of ω_n^2

$$\hat{\omega}_n^2 = \frac{1}{T-q} \sum_{t=q+1}^{T} \left(\hat{b}_Y(\mathrm{L}) Y_t\right)^2 .$$

As $\{Y_t\}_{t\in\mathbb{Z}}$ is causal by assumption, its autocovariance function $\gamma_Y(h)$, $h \in \mathbb{Z}$, fulfils $\sum_{h=-\infty}^{\infty} |\gamma_Y(h)| < \infty$ (lemma 1.1.4). Thus

$$\mathbb{E}\left(\frac{1}{T-q}\sum_{t=q+1}^{T} Y_{t-k}\,Y_{t-l} - \gamma_Y(k-l)\right)^2 = O\left(\frac{1}{T}\right) \qquad \text{for all } \; k,l = 1,\ldots,q.$$

Therefore $\hat{b}_Y$ and thus $\hat{\omega}_n^2$, too, are consistent. The asymptotic normality is also due to standard theory (see e.g. Brockwell and Davis 1991, theorem 8.11.1.). □

If $\{Y_t\}_{t\in\mathbb{Z}}$ is not observable, estimators of its parameters can only be obtained via approximations. By formally replacing Y_t by $\bar{X}_t$ in $\mathcal{L}^{Y}_{n,T}(\theta_b)$, we get $\mathcal{L}^{\bar{X}}_{n,T}(\theta_b)$.

2.4.10 DEFINITION
In analogy to $\mathcal{L}^{Y}_{n,T}(\theta_b)$ of the preceding lemma, define $\mathcal{L}^{\bar{X}}_{n,T}(\theta_b)$ for $\theta_b = (b', \omega_n^2)' \in \Theta_b$ as

$$\mathcal{L}^{\bar{X}}_{n,T}(\theta_b) = \log \omega_n^2 + \log(2\pi) + \frac{1}{\omega_n^2\,(T-q)} \sum_{t=q+1}^{T} \left(b(\mathrm{L})\,\bar{X}_t\right)^2 .$$

We already have seen in lemma 2.3.6, that, if $\operatorname{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right) = O\left(\frac{1}{n}\right)$ for $i \neq j$, then $\mathbb{E}(\bar{X}_t - Y_t)^2 = \mathbb{E}\,\bar{Z}_t^2 = O\left(\frac{1}{n}\right)$. Therefore in this case we can indeed use $\mathcal{L}^{\bar{X}}_{n,T}(\theta)$ as an approximation to $\mathcal{L}^{Y}_{n,T}(\theta)$ for any $\theta \in \Theta_b$.

2.4.11 LEMMA
If we have in the setting of the GICM (assumption 2.3.1) that for $i \neq j$

$$\sigma_n^{ij} = \operatorname{cov}\left(\zeta_t^{(i)}, \zeta_t^{(j)}\right) = O\left(\frac{1}{n}\right) ,$$

then for all $\theta = (b_1, \ldots, b_q, \omega_n^2) \in \Theta_b \subset \mathbb{R}^q \times \mathbb{R}_0^+$

$$\mathbb{E}\left(\mathcal{L}^{Y}_{n,T}(\theta) - \mathcal{L}^{\bar{X}}_{n,T}(\theta)\right)^2 = O\left(\frac{1}{n}\right) .$$

PROOF:
The statement is a corollary of the above mentioned lemma 2.3.6. Since $\bar{X}_t = \bar{Z}_t + Y_t$ for all $t \in \mathbb{Z}$, we can express the mean squared error in terms of $\{\bar{Z}_t\}_{t\in\mathbb{Z}}$:

$$\begin{aligned}
\mathbb{E}\left(\mathcal{L}^{Y}_{n,T}(\theta) - \mathcal{L}^{\bar{X}}_{n,T}(\theta)\right)^2 &= \mathbb{E}\left(\frac{1}{T-q}\sum_{t=q+1}^{T}\left[(b(\mathrm{L})\,Y_t)^2 - (b(\mathrm{L})\,\bar{X}_t)^2\right]\right)^2 \\
&= \mathbb{E}\left(\frac{1}{T-q}\sum_{t=q+1}^{T}\left[(b(\mathrm{L})\,\bar{Z}_t)^2 + 2\,(b(\mathrm{L})\,\bar{Z}_t)\,(b(\mathrm{L})\,Y_t)\right]\right)^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(T-q)^2} \sum_{s,t=q+1}^{T} \Big[\mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_s)^2 \, (b(\mathrm{L})\,\bar{Z}_t)^2 \right) \\
&\qquad + 4\,\mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_s)^2 \, (b(\mathrm{L})\,\bar{Z}_t)\,(b(\mathrm{L})\,Y_t) \right) \\
&\qquad + 4\,\mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_s)\,(b(\mathrm{L})\,\bar{Z}_t)\,(b(\mathrm{L})\,Y_s)\,(b(\mathrm{L})\,Y_t) \right) \Big] \\
&= \frac{1}{(T-q)^2} \sum_{s,t=q+1}^{T} \Big[\mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_s)^2 \right) \mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_t)^2 \right) \\
&\qquad + 2\,\left(\mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_s)\,(b(\mathrm{L})\,\bar{Z}_t) \right) \right)^2 \\
&\qquad + 4\,\delta_{st}\,\omega_n^2\,\mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_s)\,(b(\mathrm{L})\,\bar{Z}_t) \right) \Big] .
\end{aligned}
$$

Here we used the fact that $b(\mathrm{L})\bar{Z}_t$ is Gaussian since the $\zeta_t^{(i)}$ are Gaussian. Therefore, the mixed 4th order moments can be calculated via the 2nd order cumulants as all higher order cumulants are zero. The second term vanishes since $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{\bar{Z}_t\}_{t\in\mathbb{Z}}$ are independent and $\mathbb{E}\,Y_t = 0$. Finally, the last line is due to the fact that $b(\mathrm{L})\,Y_t = \upsilon_t$ for all $t \in \mathbb{Z}$, where υ_s and υ_t are independent for $s \neq t$ and $\mathbb{E}\,\upsilon_t = 0$ by assumption. Furthermore we have (see remark 2.3.5) that

$$
\begin{aligned}
\mathbb{E}\left((b(\mathrm{L})\,\bar{Z}_s)\,(b(\mathrm{L})\,\bar{Z}_t) \right) &= \sum_{k,l=0}^{q} b_k\,b_l \sum_{u,v=0}^{\infty} \varphi_u\,\varphi_v\,\mathbb{E}\left(\bar{\zeta}_{s-k-u}\,\bar{\zeta}_{t-l-v} \right) \\
&= \sum_{k,l=0}^{q} b_k\,b_l \sum_{u=0}^{\infty} \varphi_u\,\varphi_{u+|s-t-k+l|}\,\mathbb{E}\,\bar{\zeta}_t^2 \\
&= \sum_{k,l=0}^{q} b_k\,b_l\,\Phi(s-t-k+l) \left(\frac{1}{n}\,\sigma_n^2 + \frac{n-1}{n}\,\sigma_n^{ij} \right) .
\end{aligned}
$$

Altogether we get if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$ that also

$$
\mathbb{E}\left(\mathcal{L}_{n,T}^{Y}(\theta) - \mathcal{L}_{n,T}^{\bar{X}}(\theta) \right)^2 = O\left(\frac{1}{n^2}\right) + \frac{1}{T-q}\,O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right) . \qquad \square
$$

Consistency and asymptotic normality of $\hat{\theta}_{\bar{X}} = \operatorname{argmin}_{\theta\in\Theta_b} \mathcal{L}_{n,T}^{\bar{X}}(\theta)$ can be obtained if both n and T tend to infinity. We first introduce some notations which then can be used to simplify the proofs of the subsequent theorems showing consistency and asymptotic normality.

2.4.12 NOTATIONS
Let q be the order of the autoregressive process $\{Y_t\}_{t\in\mathbb{Z}}$. For $k,l = 0,\ldots,q$, denote

$$
\hat{\gamma}_{\bar{X}}(k,l) = \frac{1}{T-q} \sum_{t=q+1}^{T} \bar{X}_{t-k}\bar{X}_{t-l} = \frac{1}{T-q} \sum_{t=q+1}^{T} \left(Y_{t-k} + \bar{Z}_{t-k} \right) \left(Y_{t-l} + \bar{Z}_{t-l} \right) ,
$$

$$
\hat{\gamma}_{Y}(k,l) = \frac{1}{T-q} \sum_{t=q+1}^{T} Y_{t-k}\,Y_{t-l} \quad \text{and} \quad \hat{\gamma}_{\bar{Z}}(k,l) = \frac{1}{T-q} \sum_{t=q+1}^{T} \bar{Z}_{t-k}\bar{Z}_{t-l} .
$$

Then $\hat{\Gamma}_{\bar{X}} = \frac{1}{T-q}\sum_{t=q+1}^{T}\bar{\mathbf{x}}_{t-1}\,\bar{\mathbf{x}}'_{t-1}$ can be written as $\hat{\Gamma}_{\bar{X}} = \left(\hat{\gamma}_{\bar{X}}(k,l)\right)_{k,l=1,\ldots,q}$ and $\hat{\gamma}_{\bar{X}} = \frac{1}{T-q}\sum_{t=q+1}^{T}\bar{X}_t\,\bar{\mathbf{x}}_{t-1} = \left(\hat{\gamma}_{\bar{X}}(0,l)\right)'_{l=1,\ldots,q}$, denoting $\bar{\mathbf{x}}_{t-1} = (\bar{X}_{t-1},\ldots,\bar{X}_{t-q})'$. Analogously let

$$\hat{\Gamma}_{\bar{Z}} = \left(\hat{\gamma}_{\bar{Z}}(k,l)\right)_{k,l=1,\ldots,q}\,, \quad \hat{\gamma}_{\bar{Z}} = \left(\hat{\gamma}_{\bar{Z}}(0,l)\right)'_{l=1,\ldots,q}\,,$$
$$\hat{\Gamma}_{Y} = \left(\hat{\gamma}_{Y}(k,l)\right)_{k,l=1,\ldots,q} \quad \text{and} \quad \hat{\gamma}_{Y} = \left(\hat{\gamma}_{Y}(0,l)\right)'_{l=1,\ldots,q}\,.$$

Minimising $\mathcal{L}^{\bar{X}}_{n,T}(\theta)$, we obtain for $\hat{\theta}_{\bar{X}} = \left(\hat{b}'_{\bar{X}}, \hat{\omega}^2_{\bar{X}}\right)' = \operatorname{argmin}_{\theta\in\Theta_b}\mathcal{L}^{\bar{X}}_{n,T}(\theta)$ that

$$\hat{b}_{\bar{X}} = \hat{\Gamma}^{-1}_{\bar{X}}\,\hat{\gamma}_{\bar{X}} \qquad \text{and} \qquad \hat{\omega}^2_{\bar{X}} = \frac{1}{T-q}\sum_{t=q+1}^{T}\left(\hat{b}_{\theta}(\mathrm{L})\,\bar{X}_t\right)^2 .$$

In order to facilitate the proofs of the theorems, we first investigate the behaviour of the bias term.

2.4.13 LEMMA

Using the notations of the preceding remark, we get under the assumptions 2.3.1 of the GICM that the bias term $\hat{\beta}_n = \hat{\Gamma}^{-1}_{\bar{X}}\left(\hat{\gamma}_{\bar{Z}} - \hat{\Gamma}_{\bar{Z}}\,\hat{\gamma}_Y\right)$ *fulfils for* $T\to\infty$

$$\hat{\beta}_n = O_P\left(\frac{1}{n}\right) \qquad \textit{if } \sigma^{ij}_n = O\left(\frac{1}{n}\right) .$$

PROOF:

The autocovariance function $\bar{\gamma}_n(h)$, $h\in\mathbb{Z}$, of the process $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ can be represented as $\bar{\gamma}_n(h) = \gamma_Y(h) + \gamma_{\bar{Z}}(h) = \Phi(h)\,\omega^2_n + \Psi(h)\operatorname{var}\bar{\zeta}_t$ for all $h\in\mathbb{Z}$ (remark 2.3.5), where $\gamma_Y(h)$ and $\gamma_{\bar{Z}}(h)$, $h\in\mathbb{Z}$, are the autocovariance functions of $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{\bar{Z}_t\}_{t\in\mathbb{Z}}$. Since the processes are causal by assumption, the autocovariance functions are absolutely summable (lemma 1.1.4) and therefore $\bar{\gamma}_n(h)$ is square summable. Thus we get as in the proof of lemma 1.2.4 that

$$\begin{aligned}
&\mathbb{E}\left(\hat{\gamma}_n(k,l) - \bar{\gamma}_n(k-l)\right)^2 \\
&= \frac{1}{(T-q)^2}\sum_{h=-(T-q-1)}^{T-q-1}(T-q-|h|)\,\left(\bar{\gamma}_n(h)^2 + \bar{\gamma}_n(h-k+l)\,\bar{\gamma}_n(h-l+k)\right) \\
&= O\left(\frac{1}{T}\right) \qquad \text{for all } k,l = 0,\ldots,q.
\end{aligned}$$

Since $\Gamma_{\bar{X}} = \left(\bar{\gamma}_n(k-l)\right)_{k,l=1,\ldots,q}$ is invertible, this implies that for $T\to\infty$ we have $\hat{\Gamma}^{-1}_{\bar{X}} - \Gamma^{-1}_{\bar{X}} = o_P(1)$ and $\hat{\Gamma}^{-1}_{\bar{X}} = O_P(1)$.

In a similar way it can be easily seen that for all $k,l = 0,\ldots,q$

$$\mathbb{E}\left(\hat{\gamma}_Y(k,l) - \gamma_Y(k-l)\right)^2 = O\left(\frac{\omega^4_n}{T}\right) ;$$

thus we get in particular that $\hat{\gamma}_Y(k,l) = O_P(1)$.

Moreover we know from lemma 2.3.6, where we have proved mean-square convergence of the process $\{\bar{X}_t - Y_t\}_{t\in\mathbb{Z}} = \{\bar{Z}_t\}_{t\in\mathbb{Z}}$, that the autocovariance function $\gamma_{\bar{Z}}$ fulfils $\gamma_{\bar{Z}}(k,l) = \Psi(k-l)\,\mathrm{var}\,\bar{Z}_t = O\left(\frac{1}{n}\right)$ for all $k,l = 0.\ldots,q$ if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$. Therefore we obtain for $\hat{\gamma}_{\bar{Z}}(k,l)$ that, if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$,

$$\begin{aligned}\mathbb{E}\Big(\hat{\gamma}_{\bar{Z}}(k,l)\Big)^2 &= \mathbb{E}\left(\frac{1}{T-q}\sum_{t=q+1}^{T}\bar{Z}_{t-k}\,\bar{Z}_{t-l}\right)^2\\ &= \gamma_{\bar{Z}}(k-l)^2 + \frac{1}{(T-q)^2}\sum_{h=-(T-q-1)}^{T-q-1}(T-q-|h|)\\ &\qquad\times\left(\gamma_{\bar{Z}}(h-k+l)\,\gamma_{\bar{Z}}(h-l+k)+\gamma_{\bar{Z}}(h)^2\right) = O\left(\frac{1}{n^2}\right).\end{aligned}$$

Since $\bar{\Gamma}_{\bar{X}}^{-1} = O_P(1)$ for $T\to\infty$ and $\hat{\gamma}_Y = O_P(1)$, we thus can conclude that the bias term fulfils for $T\to\infty$ that

$$\hat{\beta}_n = \hat{\Gamma}_{\bar{X}}^{-1}\left(\hat{\gamma}_{\bar{Z}} - \hat{\Gamma}_{\bar{Z}}\,\hat{\gamma}_Y\right) = O_P\left(\frac{1}{n}\right)\qquad \text{if } \sigma_n^{ij} = O\left(\frac{1}{n}\right). \qquad \square$$

Now we are in the position to prove the consistency result.

2.4.14 THEOREM

Assume that the assumptions 2.3.1 of the GICM are fulfilled and that $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$. *Then we get with the notations of the preceding remark that* $\hat{\theta}_{\bar{X}} = \mathrm{argmin}_{\theta\in\Theta_b}\mathcal{L}_{n,T}^{\bar{X}}(\theta)$, *where* $\hat{\theta}_{\bar{X}} = \left(\hat{b}'_{\bar{X}},\hat{\omega}^2_{\bar{X}}\right)'$, *is a consistent estimator of the true parameter* $\theta_b = (b',\omega^2)'$:

$$\hat{\theta}_{\bar{X}} - \theta_b = o_P(1) \qquad \textit{for } n,T\to\infty.$$

PROOF:

The processes $\{\bar{Z}_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ are independent. Thus we get for each $\hat{\gamma}_{\bar{X}}(k,l)$, $k,l = 0,\ldots,q$, using the notation $k-l = d$, that

$$\begin{aligned}&\mathbb{E}\Big(\hat{\gamma}_{\bar{X}}(k,l) - \hat{\gamma}_Y(k,l) - \hat{\gamma}_{\bar{Z}}(k,l)\Big)^2\\ &= \frac{1}{(T-q)^2}\,\mathbb{E}\left(\sum_{t=q+1}^{T}\left(Y_{t-k}\,\bar{Z}_{t-l} + Y_{t-l}\,\bar{Z}_{t-k}\right)\right)^2\\ &= \frac{2}{(T-q)^2}\sum_{h=-(T-q-1)}^{T-q-1}(T-q-|h|)\,\left(\gamma_Y(h)\,\gamma_{\bar{Z}}(h) + \gamma_Y(h+d)\,\gamma_{\bar{Z}}(h-d)\right)\\ &= O\left(\frac{1}{n\,T}\right).\end{aligned}$$

As in the proof of the preceding lemma, this is due to the absolute summability of the autocovariance functions and the fact that $\gamma_{\bar{Z}}(h) = O\left(\frac{1}{n}\right)$ for all $h\in\mathbb{Z}$ (lemma 2.3.6). This means that $\hat{\Gamma}_X - \hat{\Gamma}_Y - \hat{\Gamma}_{\bar{Z}} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ and also $\hat{\gamma}_X - \hat{\gamma}_Y - \hat{\gamma}_{\bar{Z}} = O_p\left(\frac{1}{\sqrt{nT}}\right)$.

In the proof of lemma 2.4.9, we have shown that $\hat{b}_Y = \hat{\Gamma}_Y^{-1}\,\hat{\gamma}_Y$. Thus we get

$$\begin{aligned}\hat{b}_{\bar{X}} - \hat{b}_Y &= \hat{\Gamma}_{\bar{X}}^{-1}\,\hat{\gamma}_{\bar{X}} - \hat{\Gamma}_Y^{-1}\,\hat{\gamma}_Y = \hat{\Gamma}_{\bar{X}}^{-1}\,(\hat{\gamma}_{\bar{X}} - \hat{\gamma}_Y) + \left(\hat{\Gamma}_{\bar{X}}^{-1} - \hat{\Gamma}_Y^{-1}\right)\,\hat{\gamma}_Y \\ &= \hat{\Gamma}_{\bar{X}}^{-1}\,(\hat{\gamma}_{\bar{X}} - \hat{\gamma}_Y) + \hat{\Gamma}_{\bar{X}}^{-1}\left(\hat{\Gamma}_Y - \hat{\Gamma}_{\bar{X}}\right)\hat{b}_Y\,.\end{aligned}$$

Now let $\hat{\beta}_n = \hat{\Gamma}_{\bar{X}}^{-1}\left(\hat{\gamma}_{\bar{Z}} - \hat{\Gamma}_{\bar{Z}}\,\hat{\gamma}_Y\right)$ as in the preceding lemma. This gives

$$\begin{aligned}\hat{b}_{\bar{X}} - \hat{b}_Y - \hat{\beta}_n &= \left(\hat{b}_{\bar{X}} - \hat{b}_Y\right) - \hat{\Gamma}_{\bar{X}}^{-1}\left(\hat{\gamma}_Z - \hat{\Gamma}_{\bar{Z}}\,\hat{b}_Y\right) \\ &= \hat{\Gamma}_{\bar{X}}^{-1}\,(\hat{\gamma}_{\bar{X}} - \hat{\gamma}_Y - \hat{\gamma}_{\bar{Z}}) - \hat{\Gamma}_{\bar{X}}^{-1}\left(\hat{\Gamma}_{\bar{X}} - \hat{\Gamma}_Y - \hat{\Gamma}_{\bar{Z}}\right)\hat{b}_Y\,.\end{aligned}$$

As shown in the proof of the preceding lemma, $\hat{\Gamma}_{\bar{X}}^{-1} = O_P(1)$ and $\hat{\gamma}_Y = O_P(1)$. Therefore we get that

$$\hat{b}_{\bar{X}} - \hat{b}_Y - \hat{\beta}_n = O_P\left(\frac{1}{\sqrt{n\,T}}\right)\,.$$

Moreover we have proved in lemma 2.4.9 that $\hat{b}_Y - b = O_P\left(\frac{1}{\sqrt{T}}\right)$. Since $\hat{\beta}_n = O_P\left(\frac{1}{n}\right)$ if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$ (lemma 2.4.13), this leads to

$$\hat{b}_{\bar{X}} - b = \hat{b}_{\bar{X}} - \hat{b}_Y - \hat{\beta}_n + \left(\hat{b}_Y - b\right) + \hat{\beta}_n = o_P(1) \quad \text{for } n, T \to \infty.$$

All that remains to be proven now is the consistency of $\hat{\omega}_{\bar{X}}^2$. As $\hat{\omega}_{\bar{X}}^2$ is obtained (see remark 2.4.12) by $\hat{\omega}_{\bar{X}}^2 = \frac{1}{T-q}\sum_{t=q+1}^{T}\left(\hat{b}_{\bar{X}}(\mathrm{L})\,\bar{X}_t\right)^2$, the consistency of $\hat{b}_{\bar{X}}$ implies that if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$

$$\hat{\omega}_{\bar{X}}^2 - \frac{1}{T-q}\sum_{t=q+1}^{T}\left(b(\mathrm{L})\,\bar{X}_t\right)^2 = o_P(1) \quad \text{for } n, T \to \infty.$$

Furthermore we can derive as above that

$$\mathbb{E}\left(\frac{1}{T-q}\sum_{t=q+1}^{T}\left(b(\mathrm{L})\,\bar{X}_t\right)^2 - \mathrm{var}\left(b(\mathrm{L})\,\bar{X}_t\right)\right)^2 = O\left(\frac{1}{T}\right)\,.$$

Finally, in proposition 2.3.7 we have shown that we have under the assumptions of the GICM and if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$ that $\mathrm{var}\left(b(\mathrm{L})\,\bar{X}_t\right) - \omega_n^2 = O\left(\frac{1}{n}\right)$. Summarising, this implies that if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$

$$\hat{\omega}_{\bar{X}}^2 - \omega_n^2 = o_P(1) \quad \text{for } n, T \to \infty. \qquad \square$$

This enables us to finally obtain asymptotic normality for the estimator $\hat{b}_{\bar{X}}$.

2.4.15 THEOREM

In the setting of the preceding theorem and with $\hat{\beta}_n$ given in lemma 2.4.13 we get

$$\sqrt{T-q}\,\left(\hat{b}_{\bar{X}} - \hat{\beta}_n - b\right) \Rightarrow N \quad \text{for } T \to \infty,$$

where $N \sim \mathrm{N}(0, \Sigma_Y)$ with $\Sigma_Y = (\Phi(k-l))_{k,l=1,\ldots,q}$.

PROOF:
In lemma 2.4.9 we have shown asymptotic normality of $\hat{b}_Y$:

$$\sqrt{T-q}\,\left(\hat{b}_Y - b\right) \Rightarrow N \quad \text{for } T \to \infty.$$

From the proof of the preceding theorem we can see that if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$

$$\hat{b}_{\bar{X}} - \hat{\beta}_n - \hat{b}_Y = O_P\left(\frac{1}{\sqrt{nT}}\right).$$

This implies that

$$\sqrt{T-q}\,\left(\hat{b}_{\bar{X}} - \hat{\beta}_n - b\right) \Rightarrow N \quad \text{for } T \to \infty. \qquad \square$$

2.4.16 REMARK

1. Note that the asymptotic variance of $\hat{a}$ in the preceding theorem is independent of σ^2, the variance of the increments.

2. We have mentioned in remark 2.3.3 that the above result can be viewed as a consequence of the consistency properties of the estimator of the common component in a one-factor model as treated in Forni et al. (2000). However, here the theorem can be proved in a direct way. The same authors determine the rates of convergence in their model in Forni et al. (2001). They show that n tending to infinity guarantees consistency for an arbitrarily slow growth of the time series length T. This is reflected in the convergence behaviour of the parameter estimator obtained above, where consistency depends on the convergence of the bias term $\hat{\beta}_n$.

2.5 Asymptotic Theory for the MLE

The aim of this section is to show asymptotic normality for the parameter estimators in all three cases: $n \to \infty$, T fixed; $T \to \infty$, n fixed, and $n, T \to \infty$.
For $\hat{\theta}_b = \operatorname{argmin}_{\theta \in \Theta_b} \mathcal{L}_{n,T}^{\bar{X}_t}(\theta)$ we already have discussed consistency and asymptotic normality in the last section. Thus it remains to prove asymptotic normality for the ICM parameter estimator and for $\hat{\theta}_a = \operatorname{argmin}_{\theta \in \Theta_a} \mathcal{L}_{n,T}^{\circ}(\theta)$. Asymptotic normality can be derived from uniform convergence conditions on the log-likelihood function. However, as in the conditional log-likelihood function of the ICM all terms depending on the mean process are weighted with $\frac{1}{n}$, they vanish asymptotically for $n \to \infty$. Thus we obtain for the last component of the parameter estimator a different rate of convergence. This implies that we can prove asymptotic normality only by adapting the uniform convergence result mentioned above accordingly. We therefore proceed as follows: first we present the classic theorem. After deriving the pointwise limits of the conditional log-likelihood functions, we turn to our model and prove that conditions corresponding to those of the theorem are fulfilled in the ICM. We conclude by showing asymptotic normality for $\hat{\theta}_a = \operatorname{argmin}_{\theta \in \Theta_a} \mathcal{L}_{n,T}^{\circ}(\theta)$, which can be obtained analogously.

2.5.1 A Classic Theorem on Asymptotic Normality

We now state the theorem our results are based on. Its content is well known but as the proof is short, we include it here for completeness. The proof is split in two steps. The central idea is to apply the mean value theorem in order to obtain a setting where the convergence of the gradient can be used. But first we show that condition (i) already implies consistency of the parameter estimators as this simplifies the proof afterwards.

2.5.1 THEOREM
Let $\Theta \subset \mathbf{R}^d$, Θ *compact, and let* $\mathcal{L}_n : \Theta \to \mathbb{R}$ *be a sequence of functions with pointwise limit* $\mathcal{L}(\theta) = \lim_{n\to\infty} \mathcal{L}_n(\theta)$ *such that* $\theta_0 = \mathrm{argmin}_{\theta\in\Theta}\mathcal{L}(\theta) \in \mathrm{Int}\Theta$ *and* θ_0 *is unique. Moreover we assume that* $\mathcal{L}$ *and* $\nabla^2\mathcal{L}$ *are continuous on* Θ*, and that* $\Gamma = \nabla^2\mathcal{L}(\theta_0)$ *is positive definite, i.e. invertible. Denote* $\mathrm{argmin}\mathcal{L}_n(\theta)$ *by* $\hat{\theta}_n$.
If for $n \to \infty$

(i) $\sup_{\theta\in\Theta} \mid \mathcal{L}_n(\theta) - \mathcal{L}(\theta) \mid = o_P(1)$

(ii) $\sup_{\theta\in\Theta} \mid \nabla^2\mathcal{L}_n(\theta) - \nabla^2\mathcal{L}(\theta) \mid = o_P(1)$ *and*

(iii) $\sqrt{n}\,\nabla\mathcal{L}_n(\theta_0) \Rightarrow N$ *with* $N \sim \mathrm{N}(0,\Sigma)$*,*

then $\hat{\theta}_n$ *is asymptotically normal for* $n \to \infty$*:*

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \Rightarrow N', \quad \textit{where } N' = \Gamma^{-1} N \sim \mathrm{N}(0, \Gamma^{-1}\,\Sigma\,\Gamma^{-1})\,.$$

For consistency, we just need condition (i).

2.5.2 PROPOSITION
In the setting of the preceding theorem the condition of uniform convergence,

$$\sup_{\theta\in\Theta} |\mathcal{L}_n(\theta) - \mathcal{L}(\theta)| = o_P(1) \quad \textit{for } n \to \infty,$$

implies consistency of the parameter estimator: $\hat{\theta}_n - \theta_0 = o_P(1)$ *for* $n \to \infty$.

PROOF:
By definition, $\theta_0 = \mathrm{argmin}_{\theta\in\Theta}\mathcal{L}(\theta)$ and $\hat{\theta}_n = \mathrm{argmin}_{\theta\in\Theta}\mathcal{L}_n(\theta)$. The uniform convergence property gives

$$\mathcal{L}_n(\theta_0) - \mathcal{L}(\theta_0) = o_P(1) \quad \text{and} \quad \mathcal{L}_n(\hat{\theta}_n) - \mathcal{L}(\hat{\theta}_n) = o_P(1) \quad \text{for } n \to \infty.$$

Since $\mathcal{L}_n(\hat{\theta}_n) \leq \mathcal{L}_n(\theta_0)$ and $\mathcal{L}(\theta_0) \leq \mathcal{L}(\hat{\theta}_n)$, we get that also $\mathcal{L}(\hat{\theta}_n) - \mathcal{L}(\theta_0) = o_P(1)$ for $n \to \infty$, with the same rate of convergence. As Θ is compact, the series $\mathcal{L}(\hat{\theta}_n)$ converges to a cluster point $x = \mathcal{L}(\theta_0)$. Since $\mathcal{L}$ is continuous and θ_0 is the unique minimum of $\mathcal{L}$ on Θ, this implies that $\hat{\theta}_n - \theta_0 = o_P(1)$ for $n \to \infty$. □

We can now use this proposition to prove the above theorem.

PROOF OF 2.5.1:
Since $\theta \in \mathbb{R}^d$, the mean value theorem leads to

$$\sqrt{n}\,\nabla\mathcal{L}_n(\hat{\theta}_n) - \sqrt{n}\,\nabla\mathcal{L}_n(\theta_0) = \sqrt{n}\,M_{\mathcal{L}_n}\,(\hat{\theta}_n - \theta_0)\,, \tag{2.3}$$

where $M_{\mathcal{L}_n}$ is given by

$$M_{\mathcal{L}_n} = \begin{pmatrix} \frac{\partial \mathcal{L}_{1,n}(\theta_{1,n})}{\partial x_1} & \cdots & \frac{\partial \mathcal{L}_{1,n}(\theta_{1,n})}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{L}_{d,n}(\theta_{d,n})}{\partial x_1} & \cdots & \frac{\partial \mathcal{L}_{d,n}(\theta_{d,n})}{\partial x_d} \end{pmatrix}$$

with intermediate points $\theta_{i,n} = \theta_0 + \kappa_i\,(\hat{\theta}_n - \hat{\theta}_0)$, $\kappa_i \in [0,1]$, $i = 1,\ldots,d$, and $\mathcal{L}_{i,n}$, $i = 1,\ldots,d$, denoting the ith coordinate function of $\mathcal{L}_n$.
First we look at the entries of $M_{\mathcal{L}_n}$. We have that

$$M_{\mathcal{L}_n} = \left(\frac{\partial \mathcal{L}_{i,n}(\theta_{i,n})}{\partial x_j}\right)_{i,j=1,\ldots,d} - \nabla^2\mathcal{L}(\theta_{i,n}) + \nabla^2\mathcal{L}(\theta_{i,n}) - \nabla^2\mathcal{L}(\theta_0) + \nabla^2\mathcal{L}(\theta_0)\,.$$

Condition (ii) implies that

$$\left(\frac{\partial \mathcal{L}_{i,n}(\theta_{i,n})}{\partial x_j}\right)_{i,j=1,\ldots,d} - \nabla^2\mathcal{L}(\theta_{i,n}) = o_P(1) \quad \text{for } n \to \infty.$$

Since for all $i = 1,\ldots,d$ we have that $||\theta_{i,n} - \theta_0|| \leq ||\hat{\theta}_n - \theta_0||$, we get from the consistency of $\hat{\theta}_n$ (proposition 2.5.2) that for all $i = 1,\ldots,d$ also

$$\theta_{i,n} - \theta_0 = o_P(1) \quad \text{for } n \to \infty.$$

Furthermore $\nabla^2\mathcal{L}$ is continuous on Θ by assumption. This yields for all $i = 1,\ldots,d$

$$\nabla^2\mathcal{L}(\theta_{i,n}) - \nabla^2\mathcal{L}(\theta_0) = o_P(1) \quad \text{for } n \to \infty.$$

Thus we have altogether that

$$M_{\mathcal{L}_n} - \Gamma = o_P(1) \quad \text{for } n \to \infty, \quad \text{where } \Gamma = \nabla^2\mathcal{L}(\theta_0).$$

The second term we treat is $\nabla\mathcal{L}_n(\hat{\theta}_n)$. If $\hat{\theta}_n$ lies in the interior of Θ, then $\nabla\mathcal{L}_n(\hat{\theta}_n) = 0$. If $\hat{\theta}_n$ is on the border of Θ, we get that $||\hat{\theta}_n - \theta_0|| > \delta$ for some $\delta > 0$ as θ_0 lies in the interior of Θ by assumption. The consistency of $\hat{\theta}_n$ then implies that for all $\varepsilon > 0$

$$\mathbb{P}(||\sqrt{n}\,\nabla\mathcal{L}_n(\hat{\theta}_n)|| > \varepsilon) \leq \mathbb{P}(||\hat{\theta}_n - \theta_0|| > \delta) \;\rightarrow\; 0 \quad \text{for } n \to \infty.$$

Since Γ is assumed to be invertible, we therefore can transform equation (2.3) such that

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) + \sqrt{n}\,\Gamma^{-1}\,\nabla\mathcal{L}_n(\theta_0) = o_P(1) \quad \text{for } n \to \infty.$$

Because of condition (iii) we then directly get

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \;\Rightarrow\; N' \quad \text{for } n \to \infty, \quad \text{where } N' \sim \mathrm{N}(0, \Gamma^{-1}\,\Sigma\,\Gamma^{-1})\,. \qquad \square$$

2.5.2 Asymptotic Properties of the Conditional Log-Likelihood-Function

Depending on whether we regard the asymptotic behaviour of $\mathcal{L}_{n,T}(\theta)$ for $n \to \infty$ or for $T \to \infty$, n fixed, the limits of the conditional log-likelihood function differ. If n, the number of time series in the panel, tends to infinity, the terms depending on the mean process vanish asymptotically, which means its information loses weight. Therefore in this case our estimator is asymptotically equivalent to the one derived by Hjellvik and Tjøstheim (1999a), which is the same as our GICM estimator $\hat{\theta}_a$ restricted to the ICM case (see remark 2.4.8). This estimator is only based on the residual processes $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$. In the case of $T \to \infty$, n fixed, however, we can use all of the information and thus are able to improve the estimator. We discuss the implications of this fact more thoroughly at the end of this section in remark 2.5.36. In the present subsection we derive the pointwise limit of the conditional log-likelihood function in each case. This can be achieved easily due to the mean-square consistency of the panel covariance estimator (lemma 1.2.4).
The conditional log-likelihood function for the ICM at $\theta = (\alpha', \sigma^2, \tau^2)' \in \Theta$ was obtained in theorem 2.4.2 as

$$\begin{aligned}\mathcal{L}_{n,T}(\theta) &= -\frac{2}{n\,(T-p)} \log \mathcal{L}(\mathbf{X}_{p+1}, \ldots, \mathbf{X}_T \mid \mathbf{X}_1, \ldots, \mathbf{X}_p) \\ &= \frac{1}{\sigma^2}\, A_{n,T}(\alpha) + \frac{1}{n\,\omega_\theta^2}\, B_{n,T}(\alpha) \\ &\quad + \frac{n-1}{n} \log \sigma^2 + \frac{1}{n}\, \log \omega_\theta^2 + \frac{1}{n}\, \log n + \log\, (2\pi)\end{aligned}$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)'$ and $\omega_\theta^2 = \tau^2 + \frac{\sigma^2}{n}$. By assumption $\sigma^2 \geq c > 0$ and $\tau^2 \geq 0$, thus also $\omega_\theta^2 > 0$ for all $n \in \mathbb{N}$. $A_{n,T}(\alpha)$ and $B_{n,T}(\alpha)$ are given in remark 2.4.3 as

$$A_{n,T}(\alpha) = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \left(\alpha(\mathrm{L})\, \mathring{X}_t^{(i)}\right)^2$$

and

$$B_{n,T}(\alpha) = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(\alpha(\mathrm{L})\, \bar{X}_t\right)^2 \, .$$

Thus it is natural to define the limits of the log-likelihood functions as follows.

2.5.3 DEFINITION
For $\theta = (\alpha', \sigma^2, \tau^2)' \in \Theta$ let c_θ and d_θ be derived from the autocovariance functions given in lemma 2.2.4 such that

$$c_\theta = \sum_{k,l=0}^{p} \alpha_k\, \alpha_l\, c(k-l) \quad \textit{and} \quad d_\theta = \sum_{k,l=0}^{q} \alpha_k\, \alpha_l\, \bar{\gamma}_n(k-l)\, ,$$

denoting $\alpha_0 = -1$. *Using these notations, let*

$$\mathcal{L}_n(\theta) = \mathbb{E}\,\mathcal{L}_{n,T}(\theta) = \frac{1}{\sigma^2}\left(\frac{n-1}{n}\right) c_\theta + \frac{1}{n\,\omega_\theta^2}\,d_\theta + \frac{n-1}{n}\,\log\sigma^2 + \frac{1}{n}\log\omega_\theta^2 + \frac{1}{n}\log n + \log(2\pi)\ ,$$

where $\omega_\theta^2 = \tau^2 + \frac{\sigma^2}{n}$*, and denote*

$$\mathcal{L}(\theta) = \frac{1}{\sigma^2}\,c_\theta + \log\sigma^2 + \log(2\pi)\ .$$

We prove next that the functions defined above are indeed the pointwise limits of $\mathcal{L}_{n,T}(\theta)$ for $T \to \infty$, n fixed, and $n \to \infty$, respectively. These then are used for establishing asymptotic normality for the parameter estimator $\hat{\theta}_{n,T}$. Note that the autocovariance functions $c(h) = \Psi(h)\,\sigma_0^2$ and $\bar{\gamma}_n(h) = \Psi(h)\,\omega_n^2 = \Psi(h)\left(\tau_0^2 + \frac{\sigma_0^2}{n}\right)$, $h \in \mathbb{Z}$, depend only on the true parameter $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$ of the panel.

2.5.4 THEOREM

Under the assumptions of the ICM (assumption 2.2.1) and using the notations introduced in the preceding definition we get for all $\theta = (\alpha', \sigma^2, \tau^2)' \in \Theta$ *such that* $\tau^2 > 0$ *that*

$$\mathbb{E}\left(\mathcal{L}_{n,T}(\theta) - \mathcal{L}_n(\theta)\right)^2 = O\left(\frac{1}{n\,T}\right) \quad \textit{and} \quad \mathbb{E}\left(\mathcal{L}_{n,T}(\theta) - \mathcal{L}(\theta)\right)^2 = O\left(\frac{1}{n}\right)\ .$$

If $\tau^2 = 0$*, we still have that*

$$\mathbb{E}\left(\mathcal{L}_{n,T}(\theta) - \mathcal{L}_n(\theta)\right)^2 = O\left(\frac{\omega_n^2}{T}\right) \quad \textit{and} \quad \mathbb{E}\left(\mathcal{L}_{n,T}(\theta) - \mathcal{L}(\theta)\right)^2 = O\left(\omega_n^2\right)\ ,$$

where $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$ *is derived from the true parameter* $\theta_0 = (a', \sigma_0^2, \tau_0^2)$.

PROOF:

Let $A_{n,T}(\alpha)$ and $B_{n,T}(\alpha)$ be as in remark 2.4.3 and note that the autocovariance function of the residuals fulfils $\mathring{\gamma}_n(h) = \operatorname{cov}\left(\mathring{X}_t^{(i)}, \mathring{X}_{t-h}^{(i)}\right) = \frac{n-1}{n}\,\Psi(h)\,\sigma_0^2 = \frac{n-1}{n}\,c(h)$ for all $h \in \mathbb{Z}$ (lemma 2.2.4). Since in the ICM $u_n = -\frac{1}{n-1} = O\left(\frac{1}{n}\right)$ (remark 2.2.5), we get from lemma 1.2.4 and its corollary 1.2.6, where we have proven mean-square convergence of the autocovariance estimator, that

$$\mathbb{E}\left(A_{n,T}(\alpha) - \sum_{k,l=0}^{p}\alpha_k\,\alpha_l\,\mathring{\gamma}_n(k-l)\right)^2 = O\left(\frac{1}{n\,T}\right) \quad \text{and}$$

$$\mathbb{E}\left(B_{n,T}(\alpha) - d_\theta\right)^2 = O\left(\frac{\omega_n^4}{T}\right) \quad \text{(see remark 2.2.5).}$$

If $\tau^2 > 0$, we have that $\frac{1}{\omega_\theta^2} \le \frac{1}{\tau^2}$ for all $n \in \mathbb{N}$. This proves the first assertion. Furthermore it is obvious that all single terms of $\mathcal{L}_n(\theta) - \mathcal{L}(\theta)$ are mean-square convergent of order $O\left(\frac{1}{\sqrt{n}}\right)$, which implies the second result.

If θ is such that $\tau^2 = 0$, then $\omega_\theta^2 = \frac{\sigma^2}{n}$ and $\frac{1}{n}\log\omega_n^2 = \frac{1}{n}\log\sigma^2 - \frac{1}{n}\log n$. In this case $\frac{1}{n\,\omega_\theta^2} = \frac{1}{\sigma^2}$ does not decay to zero any more for $n \to \infty$, which implies that the convergence properties of the term $B_{n,T}(\alpha)$ depend only on ω_n^2. The above notation covers both of the cases $\tau_0^2 > 0$ and $\tau_0^2 = 0$. □

2.5.5 REMARK
The theorem shows that for $n \to \infty$ all terms of $\mathcal{L}_{n,T}(\theta)$ depending on ω_θ^2 and thus on the mean process $\bar{X}_t$ vanish asymptotically for $n \to \infty$ if $\tau^2 > 0$. If $n \to \infty$ and T is fixed, τ^2 indeed cannot be consistently estimated as we then only have a finite number of observations for the process $\{\eta_t\}_{t\in\mathbb{Z}}$. However we can consistently estimate $\omega_n^2 = \mathrm{var}(\eta_t + \bar{\varepsilon}_t)$ from the mean process if $T \to \infty$. Thus the estimate $\hat{\tau}^2 = \hat{\omega}_n^2 - \frac{\hat{\sigma}^2}{n}$, which we obtain by the minimisation algorithm 2.4.4, is consistent if $T \to \infty$. The asymptotic behaviour of $\hat{\tau}$ is discussed in more detail in remark 2.5.21 after having derived the convergence properties of the parameter estimators explicitly.

2.5.3 Asymptotic Normality of the Parameter Estimators in the ICM

For showing asymptotic normality we follow the lines of the proof of the classic theorem 2.5.1. However we have to adapt the theorem such that we can assess the asymptotic behaviour of the parameter estimator for both $n \to \infty$ and $T \to \infty$, n fixed. If $n \to \infty$, the componentwise limit $\lim_{n\to\infty} \nabla^2\mathcal{L}_n(\theta_0)$ is not positive definite any more. Therefore we regard instead the matrix $\Gamma_n = D_n\,\nabla^2\mathcal{L}_n(\theta_0)\,D_n$, where D_n is given by $D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n}\end{pmatrix}$. We prove in lemma 2.5.10 that $\lim_{n\to\infty}\Gamma_n$ is positive definite and thus invertible. These considerations imply that we cannot use the mean value theorem on $\sqrt{n}\,\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T}) - \sqrt{n}\,\nabla\mathcal{L}_{n,T}(\theta_0)$ but on the same expression multiplied by D_n, i.e.

$$\sqrt{n}\,D_n\,\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T}) - \sqrt{n}\,D_n\,\nabla\mathcal{L}_{n,T}(\theta_0) = \sqrt{n}\,D_n\,M_{\mathcal{L}_{n,T}}(\theta)\,D_n\,D_n^{-1}\left(\hat{\theta}_{n,T} - \theta_0\right),$$

where

$$M_{\mathcal{L}_{n,T}}(\theta) = \begin{pmatrix} \frac{\partial\,\mathcal{L}_{n,T;1}(\theta_{1,n})}{\partial x_1} & \cdots & \frac{\partial\,\mathcal{L}_{n,T;1}(\theta_{1,n})}{\partial x_{p+2}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\,\mathcal{L}_{n,T;p+2}(\theta_{p+2,n})}{\partial x_1} & \cdots & \frac{\partial\,\mathcal{L}_{n,T;p+2}(\theta_{p+2,n})}{\partial x_{p+2}} \end{pmatrix}$$

with intermediate points $\theta_{i,n} = \theta_0 + \kappa_i\,(\hat{\theta}_{n,T} - \hat{\theta}_0)$, $\kappa_i \in [0,1]$, $i = 1,\ldots,p+2$, and $\mathcal{L}_{n,T;i}$, $i = 1,\ldots,p+2$, denoting the ith coordinate function of $\mathcal{L}_{n,T}$. Consistency of the parameter estimator yields that $\sqrt{n}\,D_n\,\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T}) = o_P(1)$. Eventually we obtain that the asymptotic distribution of $\sqrt{n}\,D_n^{-1}\left(\hat{\theta}_{n,T} - \theta_0\right)$ is identical to that of $\sqrt{n}\,\Gamma_n^{-1}\,D_n\,\nabla\mathcal{L}_{n,T}(\theta_0)$.

In order to structure the proof, we prove the conditions of the theorem in separate steps. First of all we have to verify the general premises. We therefore show next that $\theta_0 = \mathrm{argmin}_{\theta\in\Theta}\mathcal{L}_n(\theta) \in \mathbb{R}^{p+2}$ is a unique minimum and that Γ_n and Γ are positive definite. At least this is true if the parameter space is chosen small enough. Subsequently we prove each of the three conditions of theorem 2.5.1 for the ICM. Asymptotic normality of the gradient, which is the last condition, has to be proved differently for $n \to \infty$, T fixed, and $T \to \infty$. In the first case we can employ the standard central limit theorem for independently and identically distributed observations, whereas the serial correlation of the time series implies that we have to use a central limit theorem for martingale arrays in the case of $T \to \infty$.

Preliminaries

In the proof of asymptotic normality we use the fact that the true parameter minimises the pointwise limit functions of theorem 2.5.4. This is stated in the following lemma. Note that we here regard $\mathcal{L}$ as a function on the smaller parameter space $\tilde{\Theta} \subset \mathbb{R}^{p+1}$ as $\frac{\partial}{\partial \tau^2}\mathcal{L}(\theta) = 0$ for all $\theta \in \Theta \subset \mathbb{R}^{p+2}$. Subsequently we change between the two viewpoints according to the actual situation.

2.5.6 LEMMA
In the setting of the ICM (assumption 2.2.1) denote the true parameter by $\theta_0 \in \Theta$*. If* $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$*, let* $\tilde{\theta}_0 = (a', \sigma_0^2)' \in \tilde{\Theta} \subset \mathbb{R}^{p+1}$*. Then the pointwise limits of* $\mathcal{L}_{n,T}(\theta)$*,* $\theta \in \Theta$*, given in definition 2.5.3 fulfil*

$$\theta_0 = \mathrm{argmin}_{\theta \in \Theta} \mathcal{L}_n(\theta) \qquad \textit{and} \qquad \tilde{\theta}_0 = \mathrm{argmin}_{\tilde{\theta} \in \tilde{\Theta}} \mathcal{L}(\tilde{\theta})\,.$$

PROOF:
For $\theta \in \Theta$, let c_θ and d_θ be as in definition 2.5.3. In the ICM the true parameter θ_0 fulfils $c_{\theta_0} = \sigma_0^2$ and $d_{\theta_0} = \omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$ (see remark 1.1.5). Thus it can be easily seen from the derivatives of $\mathcal{L}_n(\theta)$, $\theta \in \Theta$, and $\mathcal{L}(\tilde{\theta})$, $\tilde{\theta} \in \tilde{\Theta}$, which are given in the Appendix C.1 (lemmas C.1.2 and C.1.3) that these functions are minimised by θ_0 and $\tilde{\theta}_0$, respectively. □

The parameter spaces used for the minimisation can be restricted such that the true parameter becomes a unique minimum of the log-likelihood function.

2.5.7 LEMMA
Denote the true parameter in the ICM (assumption 2.2.1) by $\theta_0 = (a', \sigma_0^2, \tau_0^2)' \in \Theta$ *and let* $\tilde{\theta}_0 = (a', \sigma_0^2)' \in \tilde{\Theta} \subset \mathbb{R}^{p+1}$*. Then one can choose compact subspaces* $\Theta' \subseteq \Theta$ *and* $\tilde{\Theta}' \subseteq \tilde{\Theta}$ *such that* θ_0 *and* $\tilde{\theta}_0$ *are unique minima of* $\mathcal{L}_n(\theta)$*,* $\theta \in \Theta'$*, and* $\mathcal{L}(\tilde{\theta})$*,* $\tilde{\theta} \in \tilde{\Theta}'$*, where* $\mathcal{L}_n$ *and* $\mathcal{L}$ *are given in definition 2.5.3.*

PROOF:
We have shown in the preceding lemma that if $\theta_0 \in \Theta$, then $\theta_0 = \mathrm{argmin}_{\theta \in \Theta} \mathcal{L}_n(\theta)$ and $\tilde{\theta}_0 = \mathrm{argmin}_{\tilde{\theta} \in \tilde{\Theta}} \mathcal{L}_n(\tilde{\theta})$. Using the Yule-Walker relations of remark 1.1.5 we can prove that we have on a neighbourhood $\tilde{\Theta}'$ of $\tilde{\theta}_0$ that $\frac{\partial \mathcal{L}}{\partial x_{\tilde{\theta}}}(\tilde{\theta}) < 0$ for all $\tilde{\theta} \in \tilde{\Theta}'$, where $x_{\tilde{\theta}} = -(\tilde{\theta} - \tilde{\theta}_0)$. Thus $\tilde{\theta}_0$ is a unique minimum of $\mathcal{L}$ on $\tilde{\Theta}'$. Analogously we obtain that θ_0 is a unique minimum of $\mathcal{L}_n$ on Θ'. As Θ is compact by assumption, the subspaces Θ' and $\tilde{\Theta}'$ can be chosen such that they are compact as well. Since the calculations are elementary but lengthy, we refer to the Appendix C.1 for details. □

This implies that without loss of generality the compact parameter spaces Θ and $\tilde{\Theta}$ can be chosen such that θ_0 and $\tilde{\theta}_0$ are unique minima. For ease of notation we therefore use from now on the conventions of the subsequent assumption.

2.5.8 ASSUMPTION
The true parameter in the ICM is $\theta_0 = (a', \sigma_0^2, \tau_0^2)' \in \Theta$. The parameter spaces Θ and $\tilde{\Theta}$ are compact such that $\theta_0 \in \Theta$ and $\tilde{\theta}_0 = (a', \sigma_0^2)' \in \tilde{\Theta}$ are unique minima of the limit functions $\mathcal{L}_n$ and $\mathcal{L}$ given in definition 2.5.3. Furthermore there exists a $c_2 > 0$ such that we have for all $\theta = (\alpha', \sigma^2, \tau^2)' \in \Theta$ that $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n} \geq c_2$ for all $n \in \mathbb{N}$.

2.5.9 REMARK
Note that we have already $\sigma^2 \geq c > 0$ due to the assumptions of the ICM (assumption 2.2.1). As $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n}$, the last condition ensures that $\frac{1}{\omega_n^2}$ is uniformly bounded in the case of $n \to \infty$. This is for example needed for proving the uniform convergence properties in the next subsection. If we regard the case $n \to \infty$, the condition implies the restriction of $\tau^2 \geq c_2 > 0$ on the parameter space. If $T \to \infty$, n fixed, this can be relaxed to $\tau^2 \geq 0$.

We can conclude from the proof of lemma 2.5.7 that the second derivatives $\nabla^2 \mathcal{L}(\tilde{\theta}_0)$ and $\nabla^2 \mathcal{L}_n(\theta_0)$ are positive definite. However this can also be seen directly.

2.5.10 LEMMA
In the setting of the preceding lemma, the matrices $\nabla^2 \mathcal{L}(\tilde{\theta}_0)$ and $\nabla^2 \mathcal{L}_n(\theta_0)$ are positive definite. If we let $D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n} \end{pmatrix}$, the transformed matrix $\Gamma_n = D_n \, \nabla^2 \mathcal{L}_n(\theta_0) \, D_n$ is positive definite as well. Under assumption 2.5.8 we have that $\Gamma = \lim_{n\to\infty} \Gamma_n$ exists. It is positive definite and fulfils $\nabla^2 \mathcal{L}(\tilde{\theta}_0) = (\,\Gamma\,)_{k,l=1,\ldots,p+1}$. All of the above matrices are continuous on Θ.

PROOF:
Straightforward calculations give (see the Appendix C.1, corollary C.1.4) that

$$\nabla^2 \mathcal{L}(\tilde{\theta}_0) = \begin{pmatrix} 2\,\left(\Psi(k-l)\right)_{k,l=1,\ldots,p} & 0 \\ 0 & \frac{1}{\sigma_0^4} \end{pmatrix}$$

and that

$$\nabla^2 \mathcal{L}_n(\theta_0) = \begin{pmatrix} 2\,\left(\Psi(k-l)\right)_{k,l=1,\ldots,p} & 0 & 0 \\ 0 & \frac{n-1}{n\,\sigma_0^4} + \frac{1}{n^3\,\omega_n^4} & \frac{1}{n^2\,\omega_n^4} \\ 0 & \frac{1}{n^2\,\omega_n^4} & \frac{1}{n\,\omega_n^4} \end{pmatrix},$$

where $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$. As $\Psi(h)$, $h \in \mathbb{Z}$, is derived from the autocovariance functions (see lemma 2.2.4), we obtain directly that $\nabla^2 \mathcal{L}(\tilde{\theta}_0)$ and $\nabla^2 \mathcal{L}_n(\theta_0)$ are positive definite and continuous. For the transformed matrix we get that

$$\Gamma_n = D_n \, \nabla^2 \mathcal{L}_n(\theta_0) \, D_n = \begin{pmatrix} 2\,\left(\Psi(k-l)\right)_{k,l=1,\ldots,p} & 0 & 0 \\ 0 & \frac{n-1}{n\,\sigma_0^4} + \frac{1}{n^3\,\omega_n^4} & \frac{1}{n\,\sqrt{n}\,\omega_n^4} \\ 0 & \frac{1}{n\,\sqrt{n}\,\omega_n^4} & \frac{1}{\omega_n^4} \end{pmatrix}.$$

It can easily be seen that Γ_n and also the componentwise limit $\Gamma = \lim_{n\to\infty} \Gamma_n$ are positive definite and that moreover $\nabla^2 \mathcal{L}(\tilde{\theta}_0) = (\,\Gamma\,)_{k,l=1,\ldots,p+1}$. □

After having verified the general premises, we now prove the adapted versions of conditions (i) to (iii) of theorem 2.5.1.

Condition (i)

The intercorrelation structure in the ICM implies that the conditional maximum-likelihood estimator $\hat{\theta}_{n,T} = \operatorname{argmin}_{\theta\in\Theta} \mathcal{L}_{n,T}(\theta)$ cannot be directly estimated from the data but has to be obtained through a recursive algorithm (see section 2.4.2). Thus we also cannot prove consistency in a direct way. We need the concept of equicontinuity in probability, which e.g. is defined in Dahlhaus (2000b):

2.5.11 DEFINITION
We call a sequence of random variables $Z_n(\theta)$, $\theta \in \Theta$, equicontinuous in probability, if for each $\eta > 0$ and $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{||\theta_1-\theta_2||\le\delta} |Z_n(\theta_1) - Z_n(\theta_2)| > \eta \right) < \varepsilon\,.$$

On a compact space equicontinuity in probability and pointwise convergence in probability imply uniform convergence. Thus, in order to prove condition (i) we first show that $\mathcal{L}_{n,T}(\theta)$, $\theta \in \Theta$, is equicontinuous in probability. This is achieved if we restrict the parameter space such that σ^2 and ω_n^2 are bounded away from zero.

2.5.12 LEMMA
Under assumption 2.5.8, i.e. if we have that $\sigma^2 \ge c > 0$ and $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n} \ge c_2 > 0$ for all $\theta = (a', \sigma^2, \tau^2)' \in \Theta$, the conditional log-likelihood function $\mathcal{L}_{n,T}(\theta)$ is equicontinuous in probability for $n\,T \to \infty$.

PROOF:
The derivatives of $\mathcal{L}_{n,T}$ are given in remark 2.4.3. For facilitating the notation, denote $a_\theta = (-1, a_1, \ldots, a_p)'$, $\bar{\mathbf{x}}_t = \left(\bar{X}_t, \ldots, \bar{X}_{t-p}\right)'$ and $\mathring{\mathbf{x}}_t^{(i)} = \left(\mathring{X}_t^{(i)}, \ldots, \mathring{X}_{t-p}^{(i)}\right)'$ for $t \in \mathbb{Z}$, $i = 1, \ldots, n$. Then all terms $a_\theta'\,\bar{\mathbf{x}}_t$ and $a_\theta'\,\mathring{\mathbf{x}}_t^{(i)}$, $t \in \mathbb{Z}$, $i = 1, \ldots, n$, which appear in $\nabla\,\mathcal{L}_{n,T}(\theta)$ are bounded by $||a_\theta||\,||\bar{\mathbf{x}}_t||$ and $||a_\theta||\,||\mathring{\mathbf{x}}_t^{(i)}||$, $i = 1, \ldots, n$. Due to the assumption that Θ is compact, $||a_\theta|| \le \sup_{\theta\in\Theta} ||a_\theta|| \le M_a$ for some $M_a < \infty$. Furthermore the restriction of the parameter space to $\sigma^2 \ge c$ and $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n} \ge c_2$ implies that there exists a $M_c < \infty$ such that $\frac{1}{\sigma^2} \le M_c$ and $\frac{1}{\omega_n^2} \le M_c$. Because of the norm inequalities, $\sup_{\theta\in\Theta} ||\nabla \mathcal{L}_{n,T}(\theta)||$ thus is bounded by a function $f_{M_a,M_c}\left(\bar{X}_t, \mathring{X}_t^{(i)}\right)$ of the processes $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ and $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, depending on M_a and M_c. As the processes $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ and $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, are Gaussian by assumption, all higher moments exist. Using the Hölder inequality we therefore obtain that

$$\mathbb{E}\,||\sup_{\theta\in\Theta} f_{M_a,M_c}\left(\bar{X}_t, \mathring{X}_t^{(i)}\right)||^2 \le M < \infty\,,$$

where M is a bound independent of n, T, and θ.
By the mean value theorem and the Chebyshev inequality we thus get that

$$\mathbb{P}\left(\sup_{||\theta_1-\theta_2||\leq\delta}|\mathcal{L}_{n,T}(\theta_1)-\mathcal{L}_{n,T}(\theta_2)|>\eta\right)$$
$$\leq \mathbb{P}(\delta\sup_{\theta\in\Theta}||f_{M_a,M_c}(\bar{X}_t,\mathring{X}_t^{(i)})||>\eta)\leq\frac{\delta^2}{\eta^2}\,\mathbb{E}\,||\sup_{\theta\in\Theta}f_{M_a,M_c}(\bar{X}_t,\mathring{X}_t^{(i)})||^2\leq\frac{\delta^2}{\eta^2}\,M\,.$$

Since M does not depend on n or T, this is also true for the limit. Therefore

$$\limsup_{nT\to\infty}\mathbb{P}\left(\sup_{||\theta_1-\theta_2||\leq\delta}|\mathcal{L}_{n,T}(\theta_1)-\mathcal{L}_{n,T}(\theta_2)|>\eta\right)\leq\frac{\delta^2}{\eta^2}\,M\,. \qquad \square$$

This enables us to establish uniform convergence for $\mathcal{L}_{n,T}(\theta)$.

2.5.13 THEOREM
If Θ is chosen as in the preceding lemma, we get that

$$\sup_{\theta\in\Theta}|\mathcal{L}_{n,T}(\theta)-\mathcal{L}_n(\theta)|=o_P(1)\quad \textit{for } T\to\infty,\ n \textit{ fixed},$$

and

$$\sup_{\theta\in\Theta}|\mathcal{L}_{n,T}(\theta)-\mathcal{L}(\theta)|=o_P(1)\quad \textit{for } n\to\infty,\ T \textit{ fixed, and } n,T\to\infty\,.$$

PROOF:
The proof is based on the fact that equicontinuity in probability and mean-square convergence imply uniform convergence.
We first treat the case $T\to\infty$, n fixed. Due to the equicontinuity in probability we can choose for each $\varepsilon>0$ and $\eta>0$ a $\delta>0$ such that

$$\limsup_{nT\to\infty}\mathbb{P}\left(\sup_{||\theta_1-\theta_2||\leq\delta}|\mathcal{L}_{n,T}(\theta_1)-\mathcal{L}_{n,T}(\theta_2)|>\eta\right)<\varepsilon\,.$$

Since Θ is compact, we can cover Θ by a finite number of open balls of radius $r\leq\delta$, i.e. there exist $\vartheta_1,\dots,\vartheta_k\in\Theta$ such that $\Theta\subset\bigcup_{i=1}^k B_r(\vartheta_i)$.
Therefore

$$\begin{aligned}\mathbb{P}\Big(\sup_{\theta\in\Theta}|\mathcal{L}_{n,T}(\theta)-\mathcal{L}_n(\theta)|>3\eta\Big)\\ &=\mathbb{P}\left(\max_{i\in\{1,\dots,k\}}\left(\sup_{\theta\in B_r(\vartheta_i)\cap\Theta}|\mathcal{L}_{n,T}(\theta)-\mathcal{L}_n(\theta)|\right)>3\eta\right)\\ &\leq\mathbb{P}\left(\max_{i\in\{1,\dots,k\}}\left(\sup_{\theta\in B_r(\vartheta_i)\cap\Theta}|\mathcal{L}_{n,T}(\theta)-\mathcal{L}_{n,T}(\vartheta_i)|\right)>\eta\right)\\ &\quad+\mathbb{P}\left(\max_{i\in\{1,\dots,k\}}|\mathcal{L}_{n,T}(\vartheta_i)-\mathcal{L}_n(\vartheta_i)|>\eta\right)\end{aligned}$$

$$+ \mathbb{P}\left(\max_{i\in\{1,\ldots,k\}} \left(\sup_{\theta \in B_r(\vartheta_i)\cap\Theta} |\mathcal{L}_n(\vartheta_i) - \mathcal{L}_n(\theta)| \right) > \eta \right) .$$

This leads to

$$\begin{aligned}
\limsup_{T\to\infty} \mathbb{P}\left(\sup_{\theta\in\Theta} |\mathcal{L}_{n,T}(\theta) - \mathcal{L}_n(\theta)| > 3\eta \right) \\
&\leq \limsup_{T\to\infty} \mathbb{P}\left(\sup_{||\theta_1-\theta_2||<r} |\mathcal{L}_{n,T}(\theta_1) - \mathcal{L}_{n,T}(\theta_2)| > \eta \right) \\
&\quad + \sum_{i=1}^{k} \limsup_{T\to\infty} \mathbb{P}\left(|\mathcal{L}_{n,T}(\vartheta_i) - \mathcal{L}_n(\vartheta_i)| > \eta \right) \\
&\quad + \mathbb{P}\left(\sup_{||\theta_1-\theta_2||<r} |\mathcal{L}_n(\theta_1) - \mathcal{L}_n(\theta_2)| > \eta \right)
\end{aligned}$$

The first of these terms is bounded by ε due to the equicontinuity of $\mathcal{L}_{n,T}(\theta)$, $\theta \in \Theta$, which has been proved in the preceding lemma.
Due to the mean-square convergence of $\mathcal{L}_{n,T}(\theta)$ for $T \to \infty$, n fixed (theorem 2.5.4), there exists a T_0 such that

$$\mathbb{P}\left(|\mathcal{L}_{n,T}(\vartheta_i) - \mathcal{L}_n(\vartheta_i)| > \eta \right) \leq \frac{\varepsilon}{k} \quad \text{for all } T \geq T_0,\ i = 1, \ldots, k.$$

The limiting function $\mathcal{L}_n$ (see definition 2.5.3) is deterministic and continuous on Θ. If $r \leq \delta$ is small enough, we therefore have

$$\sup_{||\theta_1-\theta_2||<r} |\mathcal{L}_n(\theta_1) - \mathcal{L}_n(\theta_2)| < \eta$$

and thus

$$\mathbb{P}\left(\sup_{||\theta_1-\theta_2||<r} |\mathcal{L}_n(\theta_1) - \mathcal{L}_n(\theta_2)| > \eta \right) = 0\,.$$

Altogether we obtain the result if the radius of the covering balls is $r \leq \delta$ such that the last term vanishes.

For the cases of $n \to \infty$, T fixed, and $n, T \to \infty$, the proof is analogous. The limiting function for $n \to \infty$, $\mathcal{L}$ (see definition 2.5.3), is deterministic and continuous on Θ. Furthermore we have that for all $\theta \in \Theta$

$$\mathbb{P}\left(|\mathcal{L}_{n,T}(\theta) - \mathcal{L}(\theta)| > \eta \right) \to 0 \quad \text{for } n \to \infty$$

due to the pointwise mean-square convergence of $\mathcal{L}_{n,T}(\theta)$ (theorem 2.5.4). The result follows as above. □

From the uniform convergence we can conclude consistency of the parameter estimator $\hat{\theta}_{n,T}$. This enables us to prove convergence in probability for $\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T})$ and later for $D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n$ in the expansion used for establishing asymptotic normality.

2.5.14 LEMMA

Denote the ICM parameter estimator by $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)' = \operatorname{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta)$ *and the true parameter by* $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$*. Then we get under assumption 2.5.8 that*

$$\begin{aligned} (\hat{a}', \hat{\sigma}^2)' - (a', \sigma_0^2)' &= o_P(1) \qquad \text{for } n\,T \to \infty \\ \text{and} \qquad \hat{\tau}^2 - \tau_0^2 &= o_P(1) \qquad \text{for } T \to \infty. \end{aligned}$$

PROOF:

We have shown uniform convergence of $\mathcal{L}_{n,T}(\theta)$ in the above theorem 2.5.13. By assumption the two limiting functions $\mathcal{L}_n(\theta)$ and $\mathcal{L}(\tilde{\theta})$ are continuous on the compact parameter spaces Θ and $\tilde{\Theta}$, with unique minima at $\theta = \theta_0$ and $\tilde{\theta} = \tilde{\theta}_0$, respectively. As in the proof of proposition 2.5.2, this yields directly the consistency of $\hat{\theta}_{n,T}$ in the case of $T \to \infty$, n fixed, and of $(\hat{a}', \hat{\sigma}^2)'$ if $n \to \infty$.

It now remains to prove the statement for $\hat{\tau}^2 = \hat{\omega}_n^2 - \frac{\hat{\sigma}^2}{n}$ in the case of $n, T \to \infty$. Here $\hat{\omega}_n^2$ is estimated as $\hat{\omega}_n^2 = \max\left(B(\hat{a}), c_2 + \frac{\hat{\sigma}^2}{n}\right)$ with $B(\hat{a}) = \frac{1}{T-p}\sum_{t=p+1}^{T}\left(\hat{a}(\mathrm{L})\,\bar{X}_t\right)^2$ (see section 2.4.2). Furthermore

$$\mathbb{E}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}\bar{X}_{t-k}\,\bar{X}_{t-l} - \bar{\gamma}_n(k-l)\right)^2 = O\left(\frac{1}{T}\right)$$

due to lemma 1.2.4. As $\bar{\gamma}(k-l) = \Psi(k-l)\left(\tau^2 + \frac{\sigma^2}{n}\right)$ (lemma 2.2.4), we thus have due to the consistency of $\hat{a}$ and $\hat{\sigma}^2$ that $\hat{\tau}^2 - \tau^2 = o_P(1)$ even in the case of $n, T \to \infty$.
□

We however do not get any asymptotic results for $\hat{\tau}$ in the case of $n \to \infty$, T fixed, as we then have only a finite number of observations for $\{\eta_t\}_{t\in\mathbb{Z}}$. The consistency properties thus imply convergence of the gradient $\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T})$ in the following sense.

2.5.15 COROLLARY

In the setting of the ICM (assumption 2.2.1), let $\hat{\theta}_{n,T} = \operatorname{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta)$*. If the true parameter* $\theta_0 = (a', \sigma_0^2, \tau_0^2)' \in \operatorname{Int}\Theta$*, we obtain under assumption 2.5.8 that the gradient* $\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T})$ *fulfils*

$$\begin{aligned} \sqrt{n\,(T-p)}\,D_n\,\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T}) &= o_P(1) \qquad \text{for } T \to \infty \\ \text{and} \qquad \sqrt{n\,(T-p)}\,\nabla\tilde{\mathcal{L}}_{n,T}(\hat{\theta}_{n,T}) &= o_P(1) \qquad \text{for } n \to \infty,\ T \text{ fixed.} \end{aligned}$$

Here $\nabla\tilde{\mathcal{L}}_{n,T}(\hat{\theta}_{n,T}) = \left(\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T})\right)_{k=1,\ldots,p+1}$ *is the vector consisting of the first* $p+1$ *components of* $\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T})$*.*

PROOF:

The statement follows from the consistency properties of $\hat{\theta}_{n,T}$ as in the proof of the classic theorem 2.5.1: if $\hat{\theta}_{n,T} \in \operatorname{Int}\Theta$, then $\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T}) = 0$. Otherwise we have for $\varepsilon > 0$ due to the consistency that

$$\mathbb{P}\left(||\sqrt{n\,(T-p)}\,D_n\,\nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T})|| > \varepsilon\right) \leq \mathbb{P}\left(||\hat{\theta}_{n,T} - \theta_0|| > \delta\right) \qquad \text{for some } \delta > 0$$

as by assumption $\theta_0 \in \mathrm{Int}\Theta$. Due to the consistency properties proved in the preceding lemma, $\mathbb{P}\left(||\hat{\theta}_{n,T} - \theta_0|| > \delta\right) \to 0$ if $T \to \infty$, which proves the first statement. In the case of $n \to \infty$, T fixed, we have to restrict ourselves to the first components of $\hat{\theta}_{n,T}$ and θ_0. Altogether this yields the result as stated. □

Condition (ii)

For assessing the asymptotic behaviour of $D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n$, we can proceed as in the preceding subsection and prove equicontinuity for each component of the matrix. Applying the mean value theorem, we get from the mean-square convergence together with equicontinuity in probability that $D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n$ is uniformly convergent. This then can be used for showing convergence in probability to the matrix Γ_n. However we can also prove convergence in probability of $D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n$ directly, as it is possible to compute an explicit representation of $\nabla^2 \mathcal{L}_{n,T}(\theta)$ and as we have already shown consistency of the parameter estimators in the last subsection.

2.5.16 THEOREM
Denote the true parameter in the ICM by $\theta_0 = (a', \sigma_0^2, \tau_0^2) \in \Theta$ *and the corresponding estimator by* $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)' = \mathrm{argmin}_{\theta\in\Theta} \mathcal{L}_{n,T}(\theta)$. *Furthermore let*

$$D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n} \end{pmatrix} \quad \textit{and} \quad M_{\mathcal{L}_{n,T}}(\theta) = \begin{pmatrix} \frac{\partial \mathcal{L}_{n,T;1}(\theta_{1,n})}{\partial x_1} & \cdots & \frac{\partial \mathcal{L}_{n,T;1}(\theta_{1,n})}{\partial x_{p+2}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{L}_{n,T;p+2}(\theta_{p+2,n})}{\partial x_1} & \cdots & \frac{\partial \mathcal{L}_{n,T;p+2}(\theta_{p+2,n})}{\partial x_{p+2}} \end{pmatrix}$$

with intermediate points $\theta_{i,n} = \theta_0 + \kappa_i\,(\hat{\theta}_{n,T} - \hat{\theta}_0)$, $\kappa_i \in [0,1]$, $i = 1,\ldots,p+2$, *and* $\mathcal{L}_{n,T;i}$, $i = 1,\ldots,p+2$, *denoting the* i*th coordinate function of* $\mathcal{L}_{n,T}$.
Under assumption 2.5.8 we have that

$$\begin{aligned} D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n - \Gamma_n &= o_P(1) \quad \textit{for } T \to \infty,\ n \textit{ fixed}, \\ D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n - \Gamma &= o_P(1) \quad \textit{for } n, T \to \infty \\ \textit{and} \qquad D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n - \tilde{\Gamma} &= o_P(1) \quad \textit{for } n \to \infty,\ T \textit{ fixed}, \end{aligned}$$

where $\Gamma_n = D_n\, \nabla^2 \mathcal{L}_n(\theta_0)\, D_n$ *and* $\Gamma = \lim_{n\to\infty} \Gamma_n$ *are given in lemma 2.5.10 and*

$$\tilde{\Gamma} = \begin{pmatrix} 2\ (\psi(k-l))_{k,l=1,\ldots,p} & 0 & 0 \\ 0 & \frac{1}{\sigma_0^4} & 0 \\ 0 & 0 & Y \end{pmatrix} = \begin{pmatrix} \nabla^2 \mathcal{L}(\tilde{\theta}_0) & 0 \\ 0 & Y \end{pmatrix},$$

Y *being independent of* n.

PROOF:
An explicit representation of the second derivatives of $\mathcal{L}_{n,T}(\theta)$ can be found in the Appendix C.1 in lemma C.1.1. As $\theta_{i,n} = \theta_0 + \kappa_i\,(\hat{\theta}_{n,T} - \hat{\theta}_0)$ with $\kappa_i \in [0,1]$, the

consistency of the parameter estimator shown in lemma 2.5.14 implies consistency of all intermediate points $\theta_{i,n}$: for all $i = 1, \dots, p+2$ we have that

$$\theta_{i,n} - \theta_0 = o_P(1) \quad \text{for } T \to \infty$$
$$\text{and} \qquad (\theta_{i,n})_{k=1,\dots,p+1} - (a', \sigma_0^2)' = o_P(1) \quad \text{for } n, T \to \infty.$$

Due to the well-known mean-square convergence of the panel autocovariance estimator (lemma 1.2.4) we thus get for each entry of $D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n$ the above stated convergence. This is straightforward except for the last component, $n \frac{\partial^2}{(\partial \tau^2)^2} \mathcal{L}_{n,T}(\theta_{p+2,n})$, in the case of $n \to \infty$, T fixed. Denoting $\theta_{p+2,n} = (\alpha_{p+2,1}, \dots, \alpha_{p+2,p}, \sigma^2_{p+2}, \tau^2_{p+2})'$, $\alpha_{p+1,0} = -1$ and $\omega_{p+2} = \tau^2_{p+2} + \frac{\sigma^2_{p+2}}{n}$, we get here that

$$n \frac{\partial^2}{(\partial \tau^2)^2} \mathcal{L}_{n,T}(\theta_{p+2,n}) = -\frac{1}{\omega^4_{p+2}} + \frac{2}{(T-p)\,\omega^6_{p+2}} \sum_{t=p+1}^{T} \sum_{k,l=0}^{p} \alpha_{p+2,k}\, \alpha_{p+2,l}\, \bar{X}_{t-k}\, \bar{X}_{t-l}\,.$$

Now let $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{\bar{Z}_t\}_{t\in\mathbb{Z}}$ be the processes defined by $Y_t = \sum_{u=0}^{\infty} \psi_u\, \eta_{t-u}$ and $\bar{Z}_t = \sum_{u=0}^{\infty} \psi_u\, \bar{\varepsilon}_{t-u}$ for all $t \in \mathbb{Z}$. Then we have $\{\bar{X}_t\}_{t\in\mathbb{Z}} = \{Y_t + \bar{Z}_t\}_{t\in\mathbb{Z}}$. As $\gamma_Y(h) = \operatorname{cov}(Y_t, Y_{t+h}) = \Psi(h)\, \tau_0^2$ and $\gamma_{\bar{Z}}(h) = \operatorname{cov}\left(\bar{Z}_t, \bar{Z}_{t+h}\right) = \Psi(h)\, \frac{\sigma_0^2}{n}$ for all $h \in \mathbb{Z}$, this yields

$$\begin{aligned}\mathbb{E}\Big(\bar{X}_{t-k}\, \bar{X}_{t-l} - Y_{t-k}\, Y_{t-l}\Big)^2 &= \mathbb{E}\left(\bar{Z}_{t-k}\, Y_{t-l} + Y_{t-k}\, \bar{Z}_{t-l} + \bar{Z}_{t-k}\, \bar{Z}_{t-l}\right)^2 \\ &= 2\,\gamma_{\bar{Z}}(0)\,\gamma_Y(0) + 2\,\gamma_{\bar{Z}}(k-l)\,\gamma_Y(k-l) + \gamma_{\bar{Z}}(0)^2 + 2\,\gamma_{\bar{Z}}(k-l)^2 = O\left(\frac{1}{n}\right)\,.\end{aligned}$$

Denoting $a_0 = -1$, let $\tilde{Y}_T = \frac{1}{T-p} \sum_{t=p+1}^{T} \sum_{k,l=0}^{p} a_k\, a_l\, Y_{t-k}\, Y_{t-l}$, where $(a_1, \dots, a_p)'$ is the true autoregressive parameter in the ICM. Then we obtain from the consistency of $\hat{a}$ that

$$n \frac{\partial^2}{(\partial \tau^2)^2} \mathcal{L}_{n,T}(\theta_{p+2,n}) - Y = o_P(1) \quad \text{for } n \to \infty\,,$$

where $Y = -\frac{1}{\omega^4_{p+2}} + \frac{2}{\omega^6_{p+2}}\, \tilde{Y}_T$. This completes the proof. □

2.5.17 REMARK
Moreover the random variable $\tilde{Y}_T$ defined in the proof of the preceding theorem fulfils that $\tilde{Y}_T = O_P(1)$ for $T \to \infty$:
$\mathbb{E}\left(\tilde{Y}_T^2 - \tau_0^2\right)^2 = O\left(\frac{1}{T}\right)$ due to the mean-square convergence of the autocovariance estimator (lemma 1.2.4) and the fact that $\sum_{k,l=0}^{p} a_k\, a_l\, \gamma_Y(k-l) = \tau_0^2$ (remark 1.1.5). Thus in particular also $Y = O_P(1)$ and $\frac{1}{Y} = O_P(1)$ for $T \to \infty$.

Condition (iii)

Here we have to study the asymptotic properties of the gradient $D_n\, \nabla \mathcal{L}_{n,T}$ at the true parameter θ_0. In the case of $n \to \infty$, T fixed, asymptotic normality of the first $p+1$

components of $\nabla\mathcal{L}_{n,T}(\theta_0)$ follows from the standard central limit theorem for independently and identically distributed data. In this case we do not obtain a result for the last component. If $T \to \infty$, we have to use a central limit theorem for martingale arrays due to the double (serial and cross-sectional) correlation structure. Altogether we obtain asymptotic normality as follows. We start with the non-degenerated case where $\tau_0^2 > 0$. The special case of independent time series ($\tau_0^2 = 0$) is treated afterwards in theorem 2.5.22.

2.5.18 THEOREM

The gradient of the conditional log-likelihood function $\mathcal{L}_{n,T}$, which has been derived in proposition 2.4.2, fulfils at the true parameter $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$ that for $T \to \infty$, n fixed,

$$\sqrt{(T-p)}\,\sqrt{n}\,D_n \nabla\mathcal{L}_{n,T}(\theta_0) \Rightarrow Z_n, \qquad \textit{where } Z_n \sim \mathrm{N}(0, \Sigma_n)\,;$$

and, if $\tau_0^2 > 0$, for $n, T \to \infty$,

$$\sqrt{T-p}\,\sqrt{n}\,D_n \nabla\mathcal{L}_{n,T}(\theta_0) \Rightarrow Z, \qquad \textit{where } Z \sim \mathrm{N}(0, \Sigma).$$

In the case of $n \to \infty$, T fixed, let $\nabla\tilde{\mathcal{L}}_{n,T}(\tilde{\theta}_0) = \Big(\nabla\tilde{\mathcal{L}}_{n,T}(\theta_0)\Big)_{k=1,\dots,p+1}$. We get that

$$\sqrt{T-p}\,\sqrt{n}\,\nabla\tilde{\mathcal{L}}_{n,T}(\tilde{\theta}_0) \Rightarrow Z_T, \qquad \textit{where } Z_T \sim \mathrm{N}(0, \Sigma_T).$$

Here we have, denoting $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$, that

$$\Sigma_n = 2\begin{pmatrix} (2\,\Psi(k-l))_{k,l=1,\dots,p} & 0 & 0 \\ 0 & \frac{n-1}{n\,\sigma_0^4} + \frac{1}{n^3\,\omega_n^4} & \frac{1}{n\,\sqrt{n}\,\omega_n^4} \\ 0 & \frac{1}{n\,\sqrt{n}\,\omega_n^4} & \frac{1}{\omega_n^4} \end{pmatrix}$$

$$\Sigma = \lim_{n\to\infty} \Sigma_n = 2\begin{pmatrix} (2\,\Psi(k-l))_{k,l=1,\dots,p} & 0 & 0 \\ 0 & \frac{1}{\sigma_0^4} & 0 \\ 0 & 0 & \frac{1}{\omega_n^4} \end{pmatrix}$$

$$\textit{and} \qquad \Sigma_T = 2\begin{pmatrix} (2\,\Psi(k-l))_{k,l=1,\dots,p} & 0 \\ 0 & \frac{1}{\sigma_0^4} \end{pmatrix}.$$

PROOF:

Due to the double nature of correlation in the ICM we have to distinguish between the cases $n \to \infty$, T fixed, and $T \to \infty$. In the first case we can employ the standard central limit theorem for independently and identically distributed observations for the proof, whereas we have to use a martingale limit theorem in the second case. In order to enhance readability, we have moved the proof to the separate section 2.5.4, which can be found after the main result. □

2.5.19 REMARK

It can be easily be seen from the explicit representations of $\Gamma_n = D_n\,\nabla^2\mathcal{L}_n(\theta_0)\,D_n$, $\Gamma = \lim_{n\to\infty}\Gamma_n$ and $\nabla^2\mathcal{L}(\tilde{\theta}_0)$ in lemma 2.5.10 that the above matrices fulfil $\Sigma_n = 2\,\Gamma_n$, $\Sigma = 2\,\Gamma$ and $\Sigma_T = 2\,\nabla^2\mathcal{L}(\tilde{\theta}_0)$.

Conclusion

We now have verified all conditions of the classic theorem 2.5.1 stated at the beginning of the chapter. This implies asymptotic normality of the parameter estimator in the ICM. Summarising, we obtain our main result:

2.5.20 THEOREM (ASYMPTOTIC NORMALITY OF THE MLE)
Let $\mathcal{L}_{n,T}(\theta)$ be the conditional log-likelihood function in the ICM derived in proposition 2.4.2 and let assumption 2.5.8 on the parameter spaces $\Theta \subset \mathbb{R}^{p+2}$ and $\tilde{\Theta} \subset \mathbb{R}^{p+1}$ be fulfilled. Let $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)' = \operatorname{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta)$ and denote the reduced vector by $\tilde{\hat{\theta}}_{n,T} = (\hat{a}', \hat{\sigma}^2)'$. Analogously, let $\theta_0 = (a', \sigma_0^2, \tau_0^2)' \in \mathrm{Int}\Theta$ be the true parameter in the ICM and denote $\tilde{\theta}_0 = (a', \sigma_0^2)' \in \mathrm{Int}\tilde{\Theta}$.

Furthermore let D_n be the transformation matrix $D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n} \end{pmatrix}$.

Then we have for $T \to \infty$, n fixed,

$$\sqrt{n\,(T-p)}\, D_n^{-1} \left(\hat{\theta}_{n,T} - \theta_0\right) \Rightarrow N_n\,, \qquad \text{where } N_n \sim \mathrm{N}\left(0, 2\,\Gamma_n^{-1}\right);$$

for $n, T \to \infty$

$$\sqrt{n\,(T-p)}\, D_n^{-1} \left(\hat{\theta}_{n,T} - \theta_0\right) \Rightarrow N\,, \qquad \text{where } N \sim \mathrm{N}\left(0, 2\,\Gamma^{-1}\right),$$

and for $n \to \infty$, T fixed,

$$\sqrt{n\,(T-p)}\, \left(\tilde{\theta}_{n,T} - \tilde{\theta}_0\right) \Rightarrow N_T\,, \qquad \text{where } N_T \sim \mathrm{N}\left(0, 2\,\Gamma^{\circ^{-1}}\right).$$

Here $\Gamma_n = D_n\,\nabla^2\mathcal{L}_n(\theta_0)\,D_n$, $\Gamma = \lim_{n\to\infty}\Gamma_n$ and $\Gamma^\circ = \nabla^2\mathcal{L}(\tilde{\theta}_0)$, where $\mathcal{L}_n$ is the pointwise limit of the log-likelihood function $\mathcal{L}_{n,T}$ in the case of $T \to \infty$, n fixed, and $\mathcal{L}$ the limit for $n \to \infty$.

PROOF:
The mean value theorem leads to

$$\begin{aligned} &\sqrt{n\,(T-p)}\, D_n\, \nabla\mathcal{L}_{n,T}(\hat{\theta}_{n,T}) - \sqrt{n\,(T-p)}\, D_n\, \nabla\mathcal{L}_{n,T}(\theta_0) \\ &\qquad = \sqrt{n\,(T-p)}\, D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n\, D_n^{-1} \left(\hat{\theta}_{n,T} - \theta_0\right), \end{aligned}$$

where

$$M_{\mathcal{L}_{n,T}}(\theta) = \begin{pmatrix} \frac{\partial\,\mathcal{L}_{n,T;1}(\theta_{1,n})}{\partial x_1} & \cdots & \frac{\partial\,\mathcal{L}_{n,T;1}(\theta_{1,n})}{\partial x_{p+2}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\,\mathcal{L}_{n,T;p+2}(\theta_{p+2,n})}{\partial x_1} & \cdots & \frac{\partial\,\mathcal{L}_{n,T;p+2}(\theta_{p+2,n})}{\partial x_{p+2}} \end{pmatrix}$$

with intermediate points $\theta_{i,n} = \theta_0 + \kappa_i\,(\hat{\theta}_{n,T} - \hat{\theta}_0)$, $\kappa_i \in [0,1]$, $i = 1, \ldots, p+2$, and $\mathcal{L}_{n,T;i}$, $i = 1, \ldots, p+2$, denoting the ith coordinate function of $\mathcal{L}_{n,T}$. The result is obtained by reasoning analogously to the proof of the classic theorem 2.5.1:

we have seen in lemma 2.5.6 that the true parameter fulfils $\theta_0 = \text{argmin}_{\theta\in\Theta}\mathcal{L}_n(\theta)$ and $(a', \sigma_0^2)' = \text{argmin}_{\tilde\theta\in\tilde\Theta}\mathcal{L}(\tilde\theta)$. From the subsequent results we know due to the consistency of $\hat\theta_{n,T}$ (condition (i), lemma 2.5.14) that

$$\sqrt{n\,(T-p)}\,D_n\,\nabla\mathcal{L}_{n,T}(\hat\theta_{n,T}) = o_P(1) \quad \text{for} \;\; T\to\infty \;\; \text{(corollary 2.5.15).}$$

In the case of $n\to\infty$, T fixed, we do not get a statement on $\hat\tau^2$ with the above methods. However $(\hat a', \hat\sigma^2)'$ is a consistent estimator of $(a', \sigma_0^2)'$; and we thus have that

$$\sqrt{n\,(T-p)}\,\nabla\tilde{\mathcal{L}}_{n,T}(\hat\theta_{n,T}) = o_P(1) \quad \text{for } n\to\infty,\, T \text{ fixed.}$$

where $\nabla\tilde{\mathcal{L}}_{n,T}(\theta)$ denotes the reduced vector $(\nabla\mathcal{L}_{n,T}(\theta))_{k=1,\ldots,p+1}$.
In theorem 2.5.16, which corresponds to condition (ii), we have shown that consistency together with mean-square convergence lead to

$$D_n\,M_{\mathcal{L}_{n,T}}(\theta)\,D_n - \Gamma_{n,T} = o_P(1) \quad \text{for } n\,T\to\infty,$$

where $\Gamma_{n,T}$ stands for Γ_n, Γ and $\tilde\Gamma = \begin{pmatrix}\Gamma^\circ & 0\\ 0 & Y\end{pmatrix}$ according to the case $T\to\infty$, n fixed; $n, T \to \infty$ and $n\to\infty$, T fixed, respectively. Γ_n, Γ and Γ° are given in theorem 2.5.16; they are invertible (lemma 2.5.10).
We therefore obtain from the above equation in the case of $T\to\infty$ that

$$\sqrt{n\,(T-p)}\,D_n^{-1}\left(\hat\theta_{n,T}-\theta_0\right) + \sqrt{n\,(T-p)}\,\Gamma_{n,T}^{-1}\,D_n\,\nabla\mathcal{L}_{n,T}(\theta_0) = o_P(1)\,,$$

and in the case of $n\to\infty$, T fixed, where only $\sqrt{n\,(T-p)}\,\nabla\tilde{\mathcal{L}}_{n,T}(\hat\theta_{n,T}) = o_P(1)$, we have due to the structure of $\tilde\Gamma$ given in theorem 2.5.16 that

$$\sqrt{n\,(T-p)}\,\left(\tilde\theta_{n,T}-\tilde\theta_0\right) + \sqrt{n\,(T-p)}\,\Gamma^{\circ\,-1}\,\nabla\tilde{\mathcal{L}}_{n,T}(\theta_0) = o_P(1)\,.$$

Finally the preceding theorem yields for $T\to\infty$ that $\sqrt{n\,(T-p)}\,D_n\,\nabla\mathcal{L}_{n,T}(\theta_0)$ is $\text{AN}(0,\Sigma_n)$ (condition (iii); notation as in Brockwell and Davis (1991)). As $\Sigma_n = 2\,\Gamma_n$ (remark 2.5.19), this implies that

$$\sqrt{n\,(T-p)}\,D_n^{-1}\left(\hat\theta_{n,T}-\theta_0\right) \;\text{ is }\; \text{AN}(0, 2\,\Gamma_n^{-1})\,.$$

In the case of $n\to\infty$, T fixed, we get analogously from the preceding theorem that

$$\sqrt{n\,(T-p)}\,\left((\hat a', \hat\sigma^2)' - (a', \sigma_0^2)'\right) \;\text{ is }\; \text{AN}(0, 2\,\Gamma^{\circ\,-1})\,.$$

This completes the proof. □

These properties can be observed nicely in the simulations presented in the Appendix A.

2.5.21 Remark

1. The theorem shows that $\hat\tau^2$ converges with a different rate than the other estimators. If $T\to\infty$, $\hat\tau^2$ is only $\sqrt{T}$-consistent, whereas the other parameter

estimators are $\sqrt{n\,T}$-consistent. As $\hat{\tau}^2$ is the estimator of the variance of the single process $\{\eta_t\}_{t\in\mathbb{Z}}$, this is a natural result. In the case of $n \to \infty$, T fixed, we cannot infer about the asymptotic behaviour of $\hat{\tau}^2$, as we then only have a finite number of observations for the common influence $\{\eta_t\}_{t\in\mathbb{Z}}$; $\hat{\tau}^2$ is not a consistent estimator of τ_0^2 in this case.

2. We can investigate the asymptotic behaviour of the parameter estimators more closely. Since $\sqrt{n\,(T-p)}\,D_n^{-1}\left(\hat{\theta}_{n,T}-\theta_0\right)$ is $\mathrm{AN}(0, 2\,\Gamma_n^{-1})$, this means that

$$\sqrt{n\,(T-p)}\left(\hat{\theta}_{n,T}-\theta_0\right) \text{ is } \mathrm{AN}(0, D_n\,2\,\Gamma_n^{-1}\,D_n)\,,$$

where $D_n\,\Gamma_n^{-1}\,D_n = \nabla^2\mathcal{L}_n(\theta_0)^{-1}$.

It is easily seen that for $n\,T \to \infty$, the variance of the estimators decreases. Denote $B = (\Psi(k-l))_{k,l=1,\ldots,p}$. As (compare lemma 2.5.10)

$$2\,\nabla^2\mathcal{L}_n(\theta_0)^{-1} = \begin{pmatrix} B^{-1} & 0 & 0 \\ 0 & \frac{2\,n\,\sigma_0^4}{n-1} & -\frac{2\,\sigma_0^4}{n-1} \\ 0 & -\frac{2\,\sigma_0^4}{n-1} & \frac{2\,\sigma_0^4}{n\,(n-1)} + 2\,n\,\omega_n^4 \end{pmatrix}$$

$$\text{and} \quad 2\,\nabla^2\mathcal{L}(\tilde{\theta}_0)^{-1} = \begin{pmatrix} B^{-1} & 0 \\ 0 & 2\,\sigma_0^4 \end{pmatrix},$$

we obtain that the asymptotic variance of $\hat{a}$ is not influenced by the strength of intercorrelation $\rho = \mathrm{cov}\left(X_t^{(i)}, X_t^{(j)}\right)/\mathrm{var}\left(X_t^{(i)}\right) = \frac{\tau_0^2}{\sigma_0^2+\tau_0^2}$ $(i \neq j)$. However the variances of $\hat{\sigma}^2$ and $\hat{\tau}^2$ vary with the size of σ_0^2 and τ_0^2. The variance of $\hat{\omega}_n^2$ can be obtained from $2\,\nabla^2\mathcal{L}_n(\theta_0)^{-1}$ as

$$\mathrm{var}\,\hat{\omega}_n^2 = \mathrm{var}\,\hat{\tau}^2 + \mathrm{var}\,\frac{\hat{\sigma}^2}{n} + 2\,\mathrm{cov}\left(\hat{\tau}^2, \frac{\hat{\sigma}^2}{n}\right) = 2\,n\,\omega_n^4\,.$$

We see in the Appendix A that the simulation results are close to the theoretical values.

3. The factor 2 in the asymptotic variance is due to the standardisation used in theorem 2.4.2. There we have multiplied the conditional log-likelihood function with the factor $-\frac{2}{n\,(T-p)}$ for convenience. This standardisation has simplified the notation in our proofs. Thus the asymptotic variances $2\,\Gamma_n^{-1}$ and $2\,\Gamma^{-1}$ are the same as those we would have obtained without using the factor 2 in the standardisation. In particular the asymptotic variance of $\hat{a}$ is $\mathrm{var}\,\hat{a} = \frac{1}{n\,(T-p)}\,B^{-1}$, which corresponds to standard theory (see e.g. Brockwell and Davis 1991, theorem 8.1.1).

The above theorem proves asymptotic normality of the parameter estimator $\hat{\theta}_{n,T}$ under assumption 2.5.8. In particular it requires $\frac{1}{\hat{\omega}_n^2}$ to be uniformly bounded in probability for $n\,T \to \infty$, implying the condition $\tau^2 \geq c_2 > 0$ for all $\theta = (\alpha', \sigma^2, \tau^2)' \in \Theta$ if $n \to \infty$ (remark 2.5.9). This is necessary for proving equicontinuity in probability for $\mathcal{L}_{n,T}$, which in turn is needed for obtaining consistency of the parameter estimators.

The latter property cannot be assessed directly due to the recursive structure of the parameter estimator (remark 2.4.8). However we can omit the condition on τ^2 if the intercorrelation in the panel is zero, i.e. if the time series in the panel are independent. This case is addressed in the following theorem. The notation follows Brockwell and Davis (1991).

2.5.22 THEOREM
Let $\theta_0 = (a', \sigma_0^2, \tau_0^2)' \in \Theta$ be the true parameter in the ICM (assumption 2.2.1). Assume that $\Theta \subseteq \tilde{\Theta} \times \mathbb{R}_0^+$, where $\tilde{\Theta}$ is such that $\sigma^2 \geq c > 0$ for all $\tilde{\theta} = (\alpha', \sigma^2)' \in \tilde{\Theta}$. Furthermore $(a', \sigma_0^2)' \in \mathrm{Int}\tilde{\Theta}$ and $\tau_0^2 = 0$. In this special case of no intercorrelation, we get for the ICM parameter estimator $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)' = \mathrm{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta)$ that

$$\sqrt{n\,(T-p)}\,\left((\hat{a}', \hat{\sigma}^2)' - (a', \sigma_0^2)'\right) \Rightarrow N, \quad \textit{where } N \sim \mathrm{N}(0, 2\,\Gamma^{\circ\,-1}),$$

with $\Gamma^\circ = \nabla^2\mathcal{L}(\tilde{\theta}_0) = \begin{pmatrix} 2\,B & 0 \\ 0 & \frac{1}{\sigma_0^4} \end{pmatrix}$ *and* $B = (\Psi(k-l))_{k,l=1,\dots,p}$.
Furthermore $\hat{\tau}^2 = O_P\left(\frac{1}{n\sqrt{T}}\right)$.

PROOF:
We first treat the case $T \to \infty$, n fixed, using the notations of the preceding theorem and adapting its proof to the present setting. As n is fixed, we do not need to use D_n for deriving the asymptotic distribution. Furthermore the restriction of the parameter space to $\hat{\sigma}^2 \geq c > 0$ implies that $\frac{1}{\hat{\sigma}^2}$ and $\frac{1}{\hat{\omega}_n^2}$ are uniformly bounded if n is fixed, such that all steps of the preceding proofs can be performed analogously. Thus we have that $\sup_{\theta\in\Theta} ||\nabla\mathcal{L}_{n,T}(\theta)|| < \infty$, which yields that the function $\mathcal{L}_{n,T}(\theta)$ is equicontinuous in probability for $T \to \infty$, n fixed. Using the pointwise mean-square convergence of $\mathcal{L}_{n,T}(\theta) - \mathcal{L}_n(\theta)$ (theorem 2.5.4) we therefore get that $\hat{\theta}_{n,T} - \theta_0 = o_P(1)$. As $(a', \sigma_0^2)' \in \mathrm{Int}\tilde{\Theta}$ by assumption, we then can conclude as in the proof of theorem 2.5.1 that $\sqrt{n\,(T-p)}\,\nabla\tilde{\mathcal{L}}_{n,T}(\hat{\theta}_{n,T}) = o_P(1)$. Since we allow for $\hat{\tau}^2 = 0$, we have that $\frac{\partial}{\partial\tau^2}\mathcal{L}_{n,T}(\hat{\theta}_{n,T}) = 0$ by construction (remark 2.4.3). Furthermore the matrix $M_{\mathcal{L}_{n,T}}(\theta)$ in the mean value theorem fulfils that $M_{\mathcal{L}_{n,T}}(\theta) - \nabla^2\mathcal{L}_n(\theta_0) = o_P(1)$ due to the consistency of $\hat{\theta}_{n,T}$ and the mean-square convergence of the panel autocovariance estimator (compare the proof of the preceding theorem). In section 2.5.4 we prove asymptotic normality of $\nabla\mathcal{L}_{n,T}(\theta_0)$, the gradient at the true parameter, for any $\tau_0^2 \geq 0$. We get from corollary 2.5.33 that in the case of $T \to \infty$, n fixed,

$$\sqrt{n\,(T-p)}\,\nabla\mathcal{L}_{n,T}(\theta_0) \Rightarrow Z_n \quad \text{where } Z_n \sim \mathrm{N}(0, 2\,\nabla^2\mathcal{L}_n(\theta_0))\,.$$

The matrix $\nabla^2\mathcal{L}_n(\theta_0)$ is positive definite (lemma 2.5.10). These considerations allow us to conclude from the mean value theorem as in the proof of the preceding theorem that

$$\sqrt{n\,(T-p)}\,(\hat{\theta}_{n,T} - \theta_0) \Rightarrow N_n \quad \text{for } T \to \infty,\ n \text{ fixed},$$

where $N_n \sim \mathrm{N}(0, 2\,\nabla^2\mathcal{L}_n(\theta_0)^{-1})$. As $\tau_0^2 = 0$, we have that $\omega_n^2 = \frac{\sigma_0^2}{n}$ and therefore $\left(\nabla^2\mathcal{L}(\tilde{\theta}_0)^{-1}\right)_{k,l=1,\dots,p+1} = \Gamma^{\circ\,-1}$.
Now we have to consider the case $n \to \infty$. Here $\frac{1}{\hat{\omega}_n^2}$ is not uniformly bounded any more as the parameter space is not restricted to $\tau^2 > 0$. Thus we do not obtain consistency

in the same way as before, because we cannot prove equicontinuity in probability of $\mathcal{L}_{n,T}(\theta)$ on the given parameter space. We however can employ a more direct procedure in this case, based on the fact that $\mathbb{E}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}\bar{X}_{t-k}\,\bar{X}_{t-l}\right)^2 = O\left(\omega_n^4\right)$ (compare remark 2.2.5). For ease of notation let

$$\hat{A}_1 = \frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{X}_t^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)}\,, \qquad \hat{B}_1 = \frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{\mathbf{x}}_{t-1}^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)\prime}\,,$$

$$\hat{A}_2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\bar{X}_t\,\bar{\mathbf{x}}_{t-1} \qquad \text{and} \qquad \hat{B}_2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\bar{\mathbf{x}}_{t-1}\,\bar{\mathbf{x}}_{t-1}'\,,$$

where $\mathring{\mathbf{x}}_{t-1}^{(i)} = \left(\mathring{X}_{t-1}^{(i)},\ldots,\mathring{X}_{t-p}^{(i)}\right)'$ for $i = 1,\ldots,n$, and $\bar{\mathbf{x}}_{t-1} = \left(\bar{X}_{t-1},\ldots,\bar{X}_{t-p}\right)'$. Then $\hat{a}$ fulfils the equation $\hat{B}\,\hat{a} = \hat{A}$ (remark 2.4.3), where $\hat{B} = \frac{1}{\hat{\sigma}^2}\,\hat{B}_1 + \frac{1}{n\,\hat{\omega}_n^2}\,\hat{B}_2$ and $\hat{A} = \frac{1}{\hat{\sigma}^2}\,\hat{A}_1 + \frac{1}{n\,\hat{\omega}_n^2}\,\hat{A}_2$. If we let $\hat{a}_0 = -1$, the estimator $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \frac{n}{n-1}\,A_{n,T}(\hat{a}) = \frac{1}{(n-1)\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\sum_{k,l=0}^{p}\hat{a}_k\,\hat{a}_l\,\mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(i)}\,.$$

Denote the estimator based on the equation $\hat{B}_1\,\hat{a}_{HT} = \hat{A}_1$ by $\hat{a}_{HT}$; the corresponding variance estimator is $\hat{\sigma}_{HT}^2 = \frac{n}{n-1}A_{n,T}(\hat{a}_{HT})$. Then $\hat{\theta}_a = (\hat{a}_{HT}', \hat{\sigma}_{HT}^2)'$ is the estimator of the parameter (a', σ_0^2) of the individual effects in the GICM (proposition 2.4.7) and is the same as the one of Hjellvik and Tjøstheim (1999a) (see remark 2.4.8).
As Θ is assumed to be compact such that $\sigma^2 \geq c > 0$ for all $\theta = (\alpha', \sigma^2, \tau^2)' \in \Theta$, we know that $\hat{\alpha}$, $\hat{\sigma}^2$ and, as $\hat{\omega}_n^2 = \hat{\tau}^2 + \frac{\hat{\sigma}^2}{n} \geq \frac{c}{n}$, that $\frac{1}{n\,\hat{\omega}_n^2} \leq \frac{1}{c}$ are bounded in probability. Moreover we have here that $\frac{1}{T-p}\sum_{t=p+1}^{T}\bar{X}_{t-k}\,\bar{X}_{t-l} = O_P\left(\frac{1}{n}\right)$ since $\omega_n^2 = \frac{\sigma_0^2}{n}$ in the special case of $\tau_0^2 = 0$. These considerations yield that

$$\hat{\sigma}^2\,\hat{A} - \hat{A}_1 = \frac{\hat{\sigma}^2}{n\,\hat{\omega}_n^2}\,\hat{A}_2 = O_P\left(\frac{1}{n}\right) \quad \text{and} \quad \hat{\sigma}^2\,\hat{B} - \hat{B}_1 = O_P\left(\frac{1}{n}\right)\,.$$

As $\operatorname{cov}\left(\mathring{X}_t^{(i)}, \mathring{X}_t^{(j)}\right) = -\frac{1}{n}\,\sigma_0^2$ if $i \neq j$, we furthermore get from the mean-square convergence of the panel autocovariance estimator that $\hat{B}_1 - \frac{n-1}{n}\,B\,\sigma_0^2 = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$, where the matrix $\lim_{n\to\infty}\frac{n-1}{n}\,B\,\sigma_0^2$ is invertible (lemma 1.2.4).
Therefore $\hat{B}_1^{-1} = O_P(1)$. Since

$$\hat{B}_1\,(\hat{a}-\hat{a}_{HT}) = \hat{B}_1\,\hat{a} - \hat{A}_1 + \hat{\sigma}^2\,\hat{A} - \hat{\sigma}^2\,\hat{B}\,\hat{a} = (\hat{B}_1 - \hat{\sigma}^2\,\hat{B})\,\hat{a} + (\hat{\sigma}^2\,\hat{A} - \hat{A}_1) = O_P\left(\frac{1}{n}\right)\,,$$

we have that $\hat{a} - \hat{a}_{HT} = O_P\left(\frac{1}{n}\right)$. This yields for the estimators of the variances that also $\hat{\sigma}^2 - \hat{\sigma}_{HT}^2 = O_P\left(\frac{1}{n}\right)$. Thus $(\hat{a}', \hat{\sigma}^2)'$ has the same asymptotic distribution as $(\hat{a}_{HT}', \hat{\sigma}_{HT}^2)$ if $n \to \infty$. Hjellvik and Tjøstheim (1999a) have shown that their estimator fulfils $\sqrt{n\,(T-p)}\;((\hat{a}_{HT}, \sigma_{HT}^2)' - (a, \sigma_0^2)')$ is $\mathrm{AN}(0, 2\,\Gamma_n^{\circ\,-1})$ and we conclude the same result as a special case of theorem 2.5.34 in the next section, where we prove asymptotic normality of the estimator $\hat{\theta}_a = (\hat{a}'_{HT}, \hat{\sigma}^2_{HT})'$ in the GICM, see corollary 2.5.35. Therefore $\hat{\theta}_{n,T}$ fulfils that for $n \to \infty$

$$\sqrt{n\,(T-p)}\,((\hat{a}', \hat{\sigma}^2)' - (a', \sigma_0^2)') \Rightarrow N, \quad \text{where } N \sim \mathrm{N}\left(0, 2\,\Gamma^{\circ\,-1}\right) .$$

Finally we consider the behaviour of the estimators $\hat{\omega}_n^2 = \hat{\tau}^2 + \frac{\hat{\sigma}^2}{n}$ and $\hat{\tau}^2$. The minimisation algorithm 2.4.4 yields that $\hat{\omega}_n^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\sum_{k,l=0}^{p} \hat{a}_k\,\hat{a}_l\,\bar{X}_{t-k}\,\bar{X}_{t-l}$, where we denote $\hat{a}_0 = -1$. Moreover we can write $\omega_n^2 = \sum_{k,l=0}^{p} a_k\,a_l\,\Psi(k-l)\,\omega_n^2$ with $a_0 = -1$ (remark 1.1.5). From the consistency of $\hat{a}$ we know that $\hat{a}_k\,\hat{a}_l - a_k\,a_l = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$. As $\mathbb{E}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}\bar{X}_{t-k}\,\bar{X}_{t-l} - \bar{\gamma}_n(k-l)\right)^2 = O\left(\frac{\omega_n^4}{T}\right)$, where $\bar{\gamma}_n(h) = \Psi(h)\,\omega_n^2$ for all $h \in \mathbb{Z}$ (remark 2.2.5), this leads to

$$\begin{aligned}\hat{\omega}_n^2 - \omega_n^2 = \sum_{k,l=0}^{p} \hat{a}_k\,\hat{a}_l\,\frac{1}{T-p}\sum_{t=p+1}^{T}\left(\bar{X}_{t-k}\,\bar{X}_{t-l} - \Psi(k-l)\,\omega_n^2\right) \\ + \sum_{k,l=0}^{p}\left(\hat{a}_k\,\hat{a}_l - a_k\,a_l\right)\,\Psi(k-l)\,\omega_n^2 = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) .\end{aligned}$$

Therefore we have due to the consistency of $\hat{\sigma}^2$ that also

$$\hat{\tau}^2 - \tau_0^2 = \hat{\omega}_n^2 - \omega_n^2 - \left(\frac{\hat{\sigma}^2}{n} - \frac{\sigma_0^2}{n}\right) = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) ,$$

i.e. for $\tau_0^2 = 0$ we get $\hat{\tau}^2 = O_P\left(\frac{1}{n\,\sqrt{T}}\right)$ since $\omega_n^2 = \tau_0^2 + \frac{\hat{\sigma}^2}{n}$. □

2.5.23 REMARK

1. In contrast to the preceding theorem, here the terms depending on the mean process vanish asymptotically if $n \to \infty$. This implies that the information contained in the mean process $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ is asymptotically not used in the estimation anymore. The proof shows that in this case $(\hat{a}, \hat{\sigma}^2)'$ indeed has the same asymptotic distribution as the parameter estimator $(\hat{a}_{HT}, \hat{\sigma}^2_{HT})'$. The latter is derived in the present work in corollary 2.5.35 after proving asymptotic normality in the GICM.

2. The last part of the proof of the preceding theorem shows that the asymptotic behaviour of $\hat{\tau}^2$ and $\hat{\omega}_n^2$ is only based on the consistency of $\hat{a}$ and $\hat{\sigma}^2$. Thus we get the same results also under the assumption that $\tau_0^2 > 0$, i.e. we have for any $\tau_0^2 \geq 0$ that

$$\hat{\tau}^2 - \tau_0^2 = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) \quad \text{and} \quad \hat{\omega}_n^2 - \omega_n^2 = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) ,$$

where $\hat{\omega}_n^2 = \hat{\tau}^2 + \frac{\hat{\sigma}^2}{n}$ and $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$.

2.5.4 Proof of Theorem 2.5.18

As the proof requires distinguishing the cases $n \to \infty$, T fixed, and $T \to \infty$, we have decided to present it in this separate section. First recall the basic properties of the true parameter θ_0 in the ICM.

2.5.24 REMARK
We know that in the ICM (assumption 2.2.1) the true parameter $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$ fulfils

$$a(\mathrm{L})\, \mathring{X}_t^{(i)} = \mathring{\varepsilon}_t^{(i)} \quad \text{and} \quad a(\mathrm{L})\, \bar{X}_t = \eta_t + \bar{\varepsilon}_t \quad \text{(see remark 2.2.3)},$$

where $a_\theta(\mathrm{L})$ is the backward shift operator defined in section 1.1. In order to simplify the notation let $\xi_t = \eta_t + \bar{\varepsilon}_t$ and $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n} = \operatorname{var} \xi_t$.
The gradient of the conditional log-likelihood function $\mathcal{L}_{n,T}$, given in proposition 2.4.2, at θ_0 is therefore

$$\nabla \mathcal{L}_{n,T}(\theta_0) = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \begin{pmatrix} \left(-\frac{2}{\sigma_0^2} \left(\mathring{\varepsilon}_t^{(i)}\, \mathring{X}_{t-k}^{(i)}\right) - \frac{2}{n\,\omega_n^2}\, \xi_t\, \bar{X}_{t-k}\right)_{k=1,\dots,p} \\ -\frac{1}{\sigma_0^4}\, \mathring{\varepsilon}_t^{(i)\,2} + \frac{n-1}{n\,\sigma_0^2} - \frac{1}{n^2\,\omega_n^4}\, \xi_t^2 + \frac{1}{n^2\,\omega_n^2} \\ -\frac{1}{n\,\omega_n^4}\, \xi_t^2 + \frac{1}{n\,\omega_n^2} \end{pmatrix}.$$

For showing convergence of vectors, we use the *Cramér-Wold-Device*:

2.5.25 PROPOSITION (CRAMÉR-WOLD-DEVICE)
A sequence of d-dimensional random vectors $\{X_n\}_{n\geq 0}$ converges weakly to a d-dimensional random vector X if and only if for all $\lambda \in \mathbb{R}^d$

$$\lambda' X_n \Rightarrow \lambda' X \qquad \textit{for } n \to \infty\,.$$

PROOF:
See for example Brockwell and Davis (1991), proposition 6.3.1. □

We split the proof of the theorem in two parts. The first case is $n \to \infty$, T fixed. Here we can use the central limit theorem for independently and identically distributed data. In the second part, $T \to \infty$, we employ a central limit theorem for martingale arrays taken from Hall and Heyde (1980).

Case $n \to \infty$, T fixed

If $n \to \infty$, T fixed, we do only have a finite number of observations for the process $\{\eta_t\}_{t\in\mathbb{Z}}$. Therefore it is not possible to obtain the asymptotic distribution of $\hat{\tau}^2$ in this case. The variance σ^2 however can consistently be estimated, even if T is fixed. Thus we here regard $\nabla \tilde{\mathcal{L}}_{n,T}(\theta_0) = \left(\nabla \mathcal{L}_{n,T}(\theta_0)\right)_{k=1,\dots,p+1}$, i.e. we omit $\frac{\partial}{\partial \tau^2}\, \mathcal{L}_{n,T}(\theta_0)$ from the analysis.
In the following we construct a sequence of independently and identically distributed random vectors $\mathbf{S}_T^{(i)}$, $i \geq 1$, to which we apply the standard central limit theorem. We then show that $\sqrt{n\,(T-p)}\, \nabla \tilde{\mathcal{L}}_{n,T}(\theta_0)$ has the same asymptotic distribution as

$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{S}_T^{(i)}$. In order to motivate the choice of $\mathbf{S}_T^{(i)}$, $i \geq 1$, we preliminarily consider some asymptotic properties of the entries of $\sqrt{n\,(T-p)}\,\nabla\tilde{\mathcal{L}}_{n,T}(\theta_0)$. In particular the terms of $\sqrt{n\,(T-p)}\,\nabla\tilde{\mathcal{L}}_{n,T}(\theta_0)$ depending on $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ vanish asymptotically if $n \to \infty$.

2.5.26 LEMMA

The entries of $\sqrt{n\,(T-p)}\,\nabla\mathcal{L}_{n,T}(\theta_0)$, where $\nabla\mathcal{L}_{n,T}(\theta_0)$ is given in remark 2.5.24, fulfil for all $k = 1, \ldots, p$ that

$$\mathbb{E}\left(\frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \frac{2}{\omega_n^2}\,\xi_t\,\bar{X}_{t-k}\right)^2 = O\left(\frac{1}{n}\right).$$

Furthermore

$$\mathbb{E}\left(\frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \frac{1}{n\,\omega_n^2}\left(-\frac{1}{\omega_n^2}\,\xi_t^2 + 1\right)\right)^2 = O\left(\frac{1}{n^3\,\omega_n^4}\right).$$

We moreover have, if we let $Z_t^{(i)} = \sum_{u=0}^{\infty} \psi_u\,\varepsilon_{t-u}^{(i)}$ as in lemma 2.2.4, that

$$\mathbb{E}\left(\frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \sum_{i=1}^{n}\left(-\frac{2}{\sigma_0^2}\left(\mathring{\varepsilon}_t^{(i)}\,\mathring{X}_{t-k}^{(i)} - \varepsilon_t^{(i)}\,Z_{t-k}^{(i)}\right)\right)\right)^2 = O\left(\frac{1}{n}\right)$$

and

$$\mathbb{E}\left(\frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \sum_{i=1}^{n}\left(-\frac{1}{\sigma_0^4}\left(\mathring{\varepsilon}_t^{(i)\,2} - \varepsilon_t^{(i)\,2}\right) - \frac{1}{n}\,\frac{1}{\sigma_0^2}\right)\right)^2 = O\left(\frac{1}{n}\right).$$

PROOF:

The proof of the above statements is straightforward. In order to enhance readability, we have moved it to the Appendix C.2.1. □

Now we can prove asymptotic normality in the case $n \to \infty$, T fixed. We apply the central limit theorem to the independently and identically distributed sequence $\{\mathbf{S}_n\}_{n\in\mathbb{N}}$ defined below.

2.5.27 THEOREM

In the setting of the ICM described in assumption 2.2.1 denote $Z_t^{(i)} = \sum_{u=0}^{\infty} \psi_u\,\varepsilon_{t-u}^{(i)}$ (see lemma 2.2.4). For $i = 1, \ldots, n$, let $\mathbf{S}_T^{(i)} = \frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \mathbf{Z}_t^{(i)}$, where

$$\mathbf{Z}_t^{(i)} = \begin{pmatrix} \left(-\frac{2}{\sigma_0^2}\,\varepsilon_t^{(i)}\,Z_{t-k}^{(i)}\right)_{k=1,\ldots,p} \\ -\frac{1}{\sigma_0^4}\,\varepsilon_t^{(i)2} + \frac{1}{\sigma_0^2} \end{pmatrix},$$

and denote $\mathbf{S}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{S}_T^{(i)}$.
Then

$$\sqrt{n}\,\mathbf{S}_n \Rightarrow N_T\,, \quad \text{with } N_T \sim \mathrm{N}\left(0, \Sigma_T\right), \quad \text{for } n \to \infty,\ T \text{ fixed},$$

where $\Sigma_T = 2\begin{pmatrix} \left(2\,\Psi(k-l)\right)_{k,l=1,\ldots,p} & 0 \\ 0 & \frac{1}{\sigma_0^4} \end{pmatrix}$.

PROOF:
Let $\lambda \in \mathbb{R}^{p+1}$. Then $\lambda' \mathbf{S}_T^{(i)}$ and $\lambda' \mathbf{S}_T^{(j)}$ are identically distributed and independent for $i \neq j$ with $\mathbb{E}\left(\lambda' \mathbf{S}_T^{(i)}\right) = 0$ and $\operatorname{var}\left(\lambda' \mathbf{S}_T^{(i)}\right) = 4 \sum_{k,l=1}^{p} \lambda_k \lambda_l \Psi(k-l) + \lambda_{p+1}^2 \frac{2}{\sigma_0^4}$ for all $i = 1, \ldots, n$. Therefore the standard central limit theorem for independently and identically distributed observations directly gives that

$$\sqrt{n}\, \lambda' \mathbf{S}_n = \frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \lambda' \mathbf{Z}_t^{(i)} \Rightarrow N_T, \quad \text{where } N_T \sim \mathrm{N}(0, \lambda' \Sigma_T \lambda)$$

with Σ_T as stated above. Using the Cramér-Wold device (proposition 2.5.25) we obtain the result. □

This allows us to show asymptotic normality for $\sqrt{n(T-p)}\, \nabla \tilde{\mathcal{L}}_{n,T}(\theta_0)$.

2.5.28 COROLLARY (THEOREM 2.5.18, CASE $n \to \infty$, T FIXED)
Let $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$ be the true parameter in the ICM. Then the reduced gradient $\nabla \tilde{\mathcal{L}}_{n,T}(\theta_0) = \left(\nabla \mathcal{L}_{n,T}(\theta_0)\right)_{k=1,\ldots,p+1}$, where $\nabla \mathcal{L}_{n,T}(\theta_0)$ is given in remark 2.5.24, is asymptotically normal:

$$\sqrt{n\,(T-p)}\, \nabla \tilde{\mathcal{L}}_{n,T}(\theta_0) \Rightarrow N_T\,, \quad \textit{with } N_T \sim \mathrm{N}(0, \Sigma_T), \quad \textit{for } n \to \infty,\ T \textit{ fixed},$$

where Σ_T is as stated in the above theorem.

PROOF:
We use the notations of the preceding theorem. Then

$$\begin{aligned}
\Delta_n &= \sqrt{n}\, \left(\sqrt{(T-p)}\, \nabla \tilde{\mathcal{L}}_{n,T}(\theta_0) - \mathbf{S}_n\right) \\
&= \frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \begin{pmatrix} \left(-\frac{2}{\sigma_0^2}\left(\mathring{\varepsilon}_t^{(i)} \mathring{X}_{t-k}^{(i)} - \varepsilon_t^{(i)} Z_{t-k}^{(i)}\right)\right)_{k=1,\ldots,p} \\ -\frac{1}{\sigma_0^4}\left(\mathring{\varepsilon}_t^{(i)\,2} - \varepsilon_t^{(i)\,2}\right) - \frac{1}{n}\frac{1}{\sigma_0^2} \end{pmatrix} \\
&\quad + \frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \begin{pmatrix} \left(-\frac{2}{\omega_n^2} \xi_t \bar{X}_{t-k}\right)_{k=1,\ldots,p} \\ \frac{1}{n}\frac{1}{\omega_n^2}\left(-\frac{1}{\omega_n^2}\, \xi_t^2 + 1\right) \end{pmatrix}.
\end{aligned}$$

Lemma 2.5.26 shows that we have $\mathbb{E}\left(\lambda' \Delta_n\right)^2 = O\left(\frac{1}{n}\right)$ for all $\lambda \in \mathbb{R}^{p+1}$. Thus $\sqrt{n\,(T-p)}\, \nabla \tilde{\mathcal{L}}_{n,T}(\theta_0)$ has the same asymptotic distribution as $\sqrt{n}\, \mathbf{S}_n$ in the case of $n \to \infty$, T fixed □

Case $T \to \infty$

In the ICM, not only $\mathring{X}_t^{(i)}$ and $\mathring{X}_t^{(j)}$ are correlated for $i, j = 1, \ldots, n$. Due to their autoregressive structure, also $\mathring{X}_s^{(i)}$ and $\mathring{X}_t^{(i)}$ are dependent for $s \neq t$. Therefore, in the case of $T \to \infty$ we cannot reduce the proof of the theorem to the central limit theorem for independently and identically distributed data as before. However, the conditions of the following central limit theorem for martingale arrays are fulfilled.

2.5.29 THEOREM (HALL AND HEYDE)
Let $\{S_{T,t}, \mathcal{F}_{T,t}, T \geq 1, 1 \leq t \leq T\}$ *be a martingale array with* $\mathcal{F}_{T,t} \subseteq \mathcal{F}_{T+1,t}$*, such that* $\mathbb{E}(S_{T,t}) = 0$ *and* $\mathbb{E}(S_{T,t}^2) < \infty$ *for all* $T \geq 1$, $1 \leq t \leq T$. *Denote the differences by* $D_{T,t}$. *If the differences fulfil*

(i) the conditional Lindeberg condition that for all $\varepsilon > 0$

$$\sum_{t=1}^{T} \mathbb{E}\left(D_{T,t}^2 \, I(|D_{T,t}| > \varepsilon) \mid \mathcal{F}_{T,t-1}\right) = o_P(1) \quad \text{for } T \to \infty$$

(ii) and an analogous condition for the conditional variance,

$$\sum_{t=1}^{T} \mathbb{E}\left(D_{T,t}^2 \mid \mathcal{F}_{T,t-1}\right) - \eta^2 = o_P(1) \quad \text{for } T \to \infty,$$

we get that for $T \to \infty$

$$S_T = \sum_{t=1}^{T} D_{T,t} \;\Rightarrow\; N, \quad \text{where } N \sim \mathrm{N}(0, \eta^2)\,.$$

PROOF:
Hall and Heyde (1980), theorem 3.2 and corollary 3.1. □

For proving asymptotic normality we now construct a martingale array fulfilling the conditions of the preceding theorem. It is defined in the next lemma; the convergence properties of the martingale differences are derived in the subsequent proposition. By regarding the transformed gradient $D_n \nabla\mathcal{L}_{n,T}(\theta_0)$ we obtain the asymptotic normality result for both of the cases $T \to \infty$, n fixed, and $n, T \to \infty$ in one step.

2.5.30 LEMMA
Assume that the assumptions of the ICM (assumption 2.2.1) are fulfilled and let the sequence $(T, n_T)_{T \geq p+1}$ *be such that* $n_{T+1} \geq n_T \geq 1$.
Define the σ*-field* $\mathcal{F}_{T,\tau}$ *by*

$$\mathcal{F}_{T,\tau} = \sigma\{\varepsilon_t^{(i)}, \eta_t;\ -\infty < t \leq \tau, i = 1, \ldots, n_T\}$$

and let

$$S_{T,\tau} = \frac{1}{\sqrt{n_T\,(T-p)}} \sum_{t=p+1}^{\tau} \sum_{i=1}^{n_T} \lambda' \, \mathbf{Z}_t^{(i)}\,,$$

where $\lambda \in \mathbb{R}^{p+2}$ *and the variables* $\mathbf{Z}_t^{(i)}$, $p+1 \leq t \leq T$, $i = 1, \ldots, n_T$, *are such that*

$$D_n \, \nabla\mathcal{L}_{n_T,T}(\theta_0) = \frac{1}{n_T\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n_T} \mathbf{Z}_t^{(i)}\,,$$

where $\nabla\mathcal{L}_{n,T}(\theta_0)$ *is given in remark 2.5.24 and* $D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n} \end{pmatrix}$.
Then $\{S_{T,\tau}, \mathcal{F}_{T,\tau}, T \geq p+1, \tau = p+1, \ldots, T\}$ *is a martingale array with* $\mathbb{E}\, S_{T,\tau} = 0$ *and* $\operatorname{var} S_{T,\tau} < \infty$. *Furthermore* $\mathcal{F}_{T,\tau} \subseteq \mathcal{F}_{T+1,\tau}$. *The martingale differences, depending on the given* λ, *are* $D_{T,t,\lambda} = \frac{1}{\sqrt{n_T\,(T-p)}} \sum_{i=1}^{n_T} \lambda' \, \mathbf{Z}_t^{(i)}$, $t = p+1, \ldots, T$.

PROOF:
The assertions are a direct consequence of the choice of $\mathbf{Z}_t^{(i)}$ and $\mathcal{F}_{T,\tau}$. □

The martingale differences fulfil the conditions of the above theorem of Hall and Heyde (1980).

2.5.31 PROPOSITION
In the setting of the preceding lemma we get for the martingale differences that

$$\sum_{t=p+1}^{T} \mathbb{E}\Big(D_{T,t,\lambda}^2 \mid \mathcal{F}_{T,t-1}\Big) - \lambda' \Sigma_n \lambda = o_P(1) \quad \textit{for } T \to \infty,$$

where Σ_n is as stated in theorem 2.5.18. If the true parameter $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$ fulfils that $\tau_0^2 > 0$, we furthermore have for all $\varepsilon > 0$ that

$$\sum_{t=p+1}^{T} \mathbb{E}\Big(D_{T,t,\lambda}^2 \, I(|D_{T,t,\lambda}| > \varepsilon) \mid \mathcal{F}_{T,t-1}\Big) = o_P(1) \quad \textit{for } T \to \infty.$$

In the case of $T \to \infty$, n fixed, it is sufficient to require that $\tau_0^2 \geq 0$.

PROOF:
Straightforward calculations give that

$$\begin{aligned}
&\sum_{t=p+1}^{T} \mathbb{E}\Big(D_{T,t,\lambda}^2 \mid \mathcal{F}_{T,t-1}\Big) \\
&\qquad = \frac{1}{n_T\,(T-p)} \sum_{t=p+1}^{T} \Big[\sum_{k,l=1}^{p} \lambda_k \, \lambda_l \\
&\qquad\qquad \times \left(\frac{4}{\sigma_0^2} \sum_{i,j=1}^{n_T} \mathring{X}_{t-k}^{(i)} \, \mathring{X}_{t-l}^{(j)} \left(\delta_{ij} - \frac{1}{n_T} \right) + \frac{4}{\omega_n^2} \, \bar{X}_{t-k} \, \bar{X}_{t-l} \right) \Big] \\
&\qquad + \lambda_{p+1}^2 \left(\frac{2\,(n_T-1)}{n_T\,\sigma_0^4} + \frac{2}{n_T^3\,\omega_n^4} \right) + \lambda_{p+2}^2 \, \frac{2}{\omega_n^4} \\
&\qquad + 2\,\lambda_{p+1}\,\lambda_{p+2} \, \frac{2}{n_T\,\sqrt{n_T}\,\omega_n^4} \,.
\end{aligned}$$

Due to the mean-square convergence of the first terms, we thus get that

$$\sum_{t=p+1}^{T} \mathbb{E}(D_{T,t,\lambda}^2 \mid \mathcal{F}_{T,t-1}) \, - \lambda' \, \Sigma_n \, \lambda \, = o_P(1)$$

with Σ_n as stated in theorem 2.5.18. The details can be found in the Appendix C.2.2 (proposition C.2.2).
It remains to prove the second assertion. Applying the Hölder and Chebyshev inequalities we obtain

$$\mathbb{E}\Big(\sum_{t=p+1}^{T} \mathbb{E}\big(D_{T,t,\lambda}^2 \, I(|D_{T,t,\lambda}| > \varepsilon) \mid \mathcal{F}_{T,t-1}\big) \Big) = \mathbb{E}\,\big(D_{T,t,\lambda}^2 \, I(|D_{T,t,\lambda}| > \varepsilon)\big)$$

$$\leq \sum_{t=p+1}^{T} \sqrt{\mathbb{E}\left(D_{T,t,\lambda}^4\right)} \sqrt{\frac{1}{\varepsilon^2}\,\mathbb{E}\left(D_{T,t,\lambda}^2\right)}\,.$$

Note that because of

$$(T-p)\,\mathbb{E}\left(D_{T,t,\lambda}^2\right) = \mathbb{E}\left(\sum_{t=p+1}^{T} \mathbb{E}\left(D_{T,t,\lambda}^2 \mid \mathcal{F}_{T,t-1}\right)\right)\,,$$

we have $(T-p)\,\mathbb{E}\left(D_{T,t,\lambda}^2\right) = \lambda'\,\Sigma_n\,\lambda$. If $\frac{1}{\omega_n^2} = O(1)$ for $n\,T \to \infty$, i.e. if $\tau_0^2 > 0$ or if $T \to \infty,\ n$ fixed, we therefore get that

$$\mathbb{E}\left(D_{T,t,\lambda}^2\right) = O\left(\frac{1}{T}\right)\,.$$

As the calculations needed to show that $\mathbb{E}\left(D_{T,t,\lambda}^4\right)$ remains bounded if $n_T \to \infty$ are too lengthy to be included here, the proof also has been moved to the Appendix C.2.2 (proposition C.2.3). Indeed we have that

$$\mathbb{E}\left(D_{T,t,\lambda}^4\right) = O\left(\frac{1}{T^2}\right)\,.$$

This yields, in both of the cases $n_T \to \infty$ and n fixed, that

$$\mathbb{E}\Big(\sum_{t=p+1}^{T} \mathbb{E}\left(D_{T,t,\lambda}^2\, I(|D_{T,t,\lambda}| > \varepsilon) \mid \mathcal{F}_{T,t-1}\right)\Big) = O\left(\frac{1}{\sqrt{T}}\right) \qquad \text{for all } \varepsilon > 0.$$

As $\mathbb{E}\left(D_{T,t,\lambda}^2\, I(|D_{T,t,\lambda}| > \varepsilon) \mid \mathcal{F}_{T,t-1}\right) \geq 0$, we just have shown that the term converges to zero in the L^1 norm. This completes the proof. □

Theorem 2.5.18 is now a direct conclusion from Hall and Heyde's theorem (notation as in Brockwell and Davis (1991)).

2.5.32 THEOREM

In the setting of the ICM (assumption 2.2.1), the gradient of the conditional log-likelihood function at the true parameter $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$*, given in remark 2.5.24, is asymptotically normal for* $T \to \infty$ *if* $\tau_0^2 > 0$*:*

$$\sqrt{n\,(T-p)}\,D_n\,\nabla\mathcal{L}_{n,T}(\theta_0) \;\textit{is}\; \mathrm{AN}(0, \Sigma_n) \qquad \textit{for } T \to \infty,$$

where Σ_n *is as stated in theorem 2.5.18 and* $D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n} \end{pmatrix}$*.*

In the case of $T \to \infty,\ n$ *fixed, the condition on* τ_0^2 *can be relaxed to* $\tau_0^2 \geq 0$*.*

PROOF:

First note that we have for $\Sigma = \lim_{n\to\infty} \Sigma_n$ that

$$\lambda'\,\Sigma_n\,\lambda - \lambda'\,\Sigma\,\lambda = O_P\left(\frac{1}{n}\right) \qquad \text{(see lemma 2.5.10);}$$

which implies that

$$\sum_{t=p+1}^{T} \mathbb{E}\Big(D^2_{T,t,\lambda} \mid \mathcal{F}_{T,t-1}\Big) - \lambda' \Sigma \lambda = o_P(1) \quad \text{for } n, T \to \infty$$

(see also the proof of the preceding proposition). Depending on the λ chosen in the definition of $D_{T,t,\lambda} = \lambda' \left(\frac{1}{\sqrt{(T-p)\,n}} \sum_{i=1}^{n} \mathbf{Z}_t^{(i)} \right)$, we thus obtain from theorem 2.5.29 that

$$S_T = \sum_{t=p+1}^{T} D_{T,t,\lambda} \text{ is } \mathrm{AN}(0, \lambda' \Sigma_n \lambda) \quad \text{for } T \to \infty.$$

This means that $S_T = \sum_{t=p+1}^{T} D_{T,t,\lambda} \Rightarrow N_\lambda$, where the random variable N_λ is distributed as $N_\lambda \sim \mathrm{N}(0, \lambda' \Sigma \lambda)$ in the case of $n, T \to \infty$ and as $N_\lambda \sim \mathrm{N}(0, \lambda' \Sigma_n \lambda)$ if $T \to \infty$, n fixed. Using the Cramér-Wold device (proposition 2.5.25), this leads to asymptotic normality of $\frac{1}{\sqrt{n\,(T-p)}} \sum_{i=1}^{n} \mathbf{Z}_t^{(i)}$. Since $\mathbf{Z}_t^{(i)}$ has been chosen such that $\nabla \mathcal{L}_{n,T}(\theta_0) = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathbf{Z}_t^{(i)}$, result follows directly. □

Distinguishing the two cases $n \to \infty$ and n fixed, we get the following explicit version of the theorem.

2.5.33 COROLLARY (THEOREM 2.5.18, CASE $T \to \infty$)
In the above setting, we obtain for the gradient of the conditional log-likelihood function at the true parameter $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$ *that for* $T \to \infty$, n *fixed,*

$$\sqrt{n\,(T-p)}\, D_n\, \nabla \mathcal{L}_{n,T}(\theta_0) \;\Rightarrow\; Z_n \quad \textit{where } Z_n \sim \mathrm{N}(0, \Sigma_n),$$

and for $n, T \to \infty$, *if* $\tau_0^2 > 0$,

$$\sqrt{n\,(T-p)}\, D_n\, \nabla \tilde{\mathcal{L}}_{n,T}(\theta_0) \;\Rightarrow\; Z \quad \textit{where } Z \sim \mathrm{N}(0, \Sigma).$$

Here Σ_n *and* $\Sigma = \lim_{n\to\infty} \Sigma_n$ *are as stated in theorem 2.5.18 and* $D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n} \end{pmatrix}$.

PROOF:
The above statements are the longer expressions for the "AN"-notation. □

2.5.5 Asymptotic Normality in the GICM

In the GICM, the parameter $\theta_a = (a_1, \ldots, a_p, \tilde{\sigma}_n^2)'$ of the individual effects and the parameter $\theta_b = (b_1, \ldots, b_q, \omega_n^2)'$ of the background process are estimated separately. Thus we can also investigate their asymptotic properties separately.
We have already discussed asymptotic normality for the estimator θ_b obtained by minimising $\mathcal{L}_{n,T}^{\bar{X}_t}$ in theorem 2.4.15. Asymptotic normality of $\hat{\theta}_a$ derived from $\mathcal{L}_{n,T}^{\circ}$ follows in a similar way as in the case of the ICM:

2.5.34 THEOREM
Under the assumptions of the GICM (assumption 2.3.1), we get if $\sigma_n^{ij} = O\left(\frac{1}{n}\right)$ *that*

$$\sqrt{n\,(T-p)}\,\left(\hat{\theta}_a - \theta_a\right) \text{ is } \mathrm{AN}(0, 2\,\Gamma_n^{\circ\,-1})$$

(notation as in Brockwell and Davis (1991)), i.e. we have for $T \to \infty$, n *fixed, that*

$$\sqrt{n\,(T-p)}\,\left(\hat{\theta}_a - \theta_a\right) \Rightarrow N_n\,,$$

and for $n \to \infty$

$$\sqrt{n\,(T-p)}\,\left(\hat{\theta}_a - \theta_a\right) \Rightarrow N\,,$$

where $N_n \sim \mathrm{N}(0, 2\,\Gamma_n^{\circ\,-1})$ *and* $N \sim \mathrm{N}(0, 2\,\Gamma^{\circ\,-1})$ *with*

$$\Gamma_n^\circ = \mathbb{E}\left(\nabla^2 \mathcal{L}_{n,T}^\circ(\theta_a)\right) = \frac{n-1}{n}\begin{pmatrix} 2\,B & 0 \\ 0 & \frac{1}{\tilde{\sigma}_n^4}\end{pmatrix} \quad \text{and} \quad \Gamma^\circ = \lim_{n\to\infty} \Gamma_n^\circ = \begin{pmatrix} 2\,B & 0 \\ 0 & \frac{1}{\sigma_0^4}\end{pmatrix},$$

where $B = (\Psi(k-l))_{k,l=1,\dots,p}$ *is derived from the autocovariance function of the processes (see remark 2.3.5).*

PROOF:
In the case of the GICM, the estimation of the parameter is based solely on the individual time series $\mathring{X}_t^{(i)} = \mathring{Z}_t^{(i)}$, $t \in \mathbb{Z}$, $i = 1, \dots, n$. The estimator of the autoregressive parameter is obtained as (see proposition 2.4.7)

$$\hat{a} = \left(\sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{\mathbf{x}}_{t-1}^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)\,\prime}\right)^{-1} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{X}_t^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)}\,,$$

where we denote $\mathring{\mathbf{x}}_{t-1}^{(i)} = (\mathring{X}_{t-1}^{(i)}, \dots, \mathring{X}_{t-p}^{(i)})'$, $i = 1, \dots, n$. The estimator of the innovations' variance is given by

$$\hat{\sigma}_n^2 = \frac{1}{(n-1)(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \left(\hat{a}(\mathrm{L})\,\mathring{X}_t^{(i)}\right)^2.$$

We have already proved in proposition 2.4.7 that $(\hat{a}', \hat{\sigma}_n^2)' - (a', \tilde{\sigma}_n^2)' = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$. By assumption $\sigma_n^{ij} = \mathrm{cov}\left(\mathring{\zeta}_t^{(i)}, \mathring{\zeta}_t^{(j)}\right) = O\left(\frac{1}{n}\right)$ for $i \neq j$. Due to the mean-square consistency of the panel covariance estimator (lemma 1.2.4) it is easy to see that

$$\nabla^2 \mathcal{L}_{n,T}^\circ(\theta_a) - \Gamma_n^\circ = O_P\left(\frac{1}{\sqrt{n\,T}}\right);$$

the second derivatives of $\mathcal{L}_{n,T}^\circ(\theta)$ are given in the Appendix C.1, lemma C.1.5.
As $\nabla \mathcal{L}_{n,T}^\circ(\hat{\theta}_a) = 0$ by construction, we obtain from the mean value theorem that

$$\sqrt{n\,(T-p)}\,\nabla \mathcal{L}_{n,T}^\circ\big(\theta_a\big) = \sqrt{n\,(T-p)}\,M_{\mathcal{L}^\circ}(\theta)\,(\hat{\theta}_a - \theta_a)\,,$$

$$\text{where} \qquad M_{\mathcal{L}^\circ}(\theta) = \begin{pmatrix} \frac{\partial \mathcal{L}^\circ_{n,T;1}(\theta_{1,n})}{\partial x_1} & \cdots & \frac{\partial \mathcal{L}^\circ_{n,T;1}(\theta_{1,n})}{\partial x_{p+1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{L}^\circ_{n,T;p+1}(\theta_{p+1,n})}{\partial x_1} & \cdots & \frac{\partial \mathcal{L}^\circ_{n,T;p+1}(\theta_{p+1,n})}{\partial x_{p+1}} \end{pmatrix}$$

with intermediate points $\theta_{i,n} = \theta_a + \kappa_i\,(\hat{\theta}_a - \hat{\theta}_a)$, $\kappa_i \in [0,1]$, $i = 1,\ldots,p+1$, and $\mathcal{L}^\circ_{n,T;i}$, $i = 1,\ldots,p+1$, denoting the ith coordinate function of $\mathcal{L}^\circ_{n,T}$ (compare proof of theorem 2.5.20, where we have proved asymptotic normality of the estimators in the ICM). Due to the consistency of the parameter estimator and the $\sqrt{n\,T}$-consistency of $\nabla^2\mathcal{L}^\circ_{n,T}(\theta_a)$ we can reason as in the proof of condition (ii) of the preceding theorem (theorem 2.5.16) that also $M_{\mathcal{L}^\circ}(\theta) - \Gamma^\circ_n = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$. It thus only remains to prove asymptotic normality of the gradient $\nabla\mathcal{L}^\circ_{n,T}\big(\theta_a\big)$, which then yields the result. Here we have to proceed as in the case of the ICM (theorem 2.5.18). The gradient at the true parameter is

$$\nabla\mathcal{L}^\circ_{n,T}(\theta_a) = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \begin{pmatrix} \left(-\frac{2}{\tilde{\sigma}_n^2}\,\mathring{\zeta}^{(i)}_t\,\mathring{Z}^{(i)}_{t-k}\right)_{k=1,\ldots,p} \\ -\frac{1}{\tilde{\sigma}_n^4}\,\mathring{\zeta}^{(i)\,2}_t + \frac{n-1}{n\,\tilde{\sigma}_n^2} \end{pmatrix},$$

which can formally be obtained from $\nabla\mathcal{L}_{n,T}(\theta_0)$ by replacing $\mathring{\varepsilon}^{(i)}_t$ by $\mathring{\zeta}^{(i)}_t$, σ_0^2 by $\tilde{\sigma}_n^2$ and by omitting all terms dependent on ω_n^2. Recall that $\mathring{Z}^{(i)}_t = \mathring{X}^{(i)}_t$ is given for all $t \in \mathbb{Z}$ by $\mathring{Z}^{(i)}_t = \sum_{u=0}^{\infty} \psi_u\,\mathring{\zeta}^{(i)}_{t-u}$ (assumption 2.3.1) and that $\operatorname{cov}\left(\mathring{\zeta}^{(i)}_t, \mathring{\zeta}^{(j)}_t\right) = \left(\delta_{ij} - \frac{1}{n}\right)\tilde{\sigma}_n^2$ (remark 2.3.5). Proceeding as in the proof of theorem 2.5.18, we get in the case of $T \to \infty$ that

$$\sqrt{n\,(T-p)}\,\nabla\mathcal{L}^\circ_{n,T}(\theta_a) \ \text{ is } \ \mathrm{AN}(0, 2\,\Gamma^{\circ\,-1}_n)\,.$$

The factor $\frac{n-1}{n}$ in Γ°_n is induced by omitting the terms depending on ω_n^2. In the remaining case of $n \to \infty$, T fixed, we have been able to employ the independence of $\varepsilon^{(i)}_t$ and $\varepsilon^{(j)}_t$ for $i \neq j$ in the ICM case (see theorem 2.5.27). This however cannot be mimicked for the GICM as we allow $\zeta^{(i)}_t$ and $\zeta^{(j)}_t$ to be correlated for $i \neq j$. Therefore we must again use the above central limit theorem for martingale arrays. For $n \in \mathbb{N}$ and $\nu = 1,\ldots,n$ let $\mathcal{F}_{n,\nu} = \sigma\{\zeta^{(i)}_t, i = 1,\ldots,\nu\}$, which implies $\mathcal{F}_{n,\nu} \subseteq \mathcal{F}_{n+1,\nu}$. Instead of $\zeta^{(i)}_t$ we now regard $\tilde{\zeta}^{(i)}_t = \zeta^{(i)}_t - \mathbb{E}\left(\zeta^{(i)}_t \mid \mathcal{F}_{n,i-1}\right)$. Then we have for $i \neq j$ that $\tilde{\zeta}^{(i)}_t$ and $\tilde{\zeta}^{(j)}_t$ are independent. Additionally replace $Z^{(i)}_t$ by $\tilde{Z}^{(i)}_t = \sum_{u=0}^{\infty} \psi_u\,\tilde{\zeta}^{(i)}_{t-u}$. As $Z^{(j)}_s$ and $\zeta^{(i)}_t$ are independent for all $s < t$, $i,j = 1,\ldots,n$, also $\tilde{Z}^{(j)}_s$ and $\tilde{\zeta}^{(i)}_t$ are independent for $s < t$. Using these notations, we form

$$\tilde{\mathbf{Z}}^{(i)}_t = \begin{pmatrix} \left(-\frac{2}{\tilde{\sigma}_n^2}\,\tilde{\zeta}^{(i)}_t\,\tilde{Z}^{(i)}_{t-k}\right)_{k=1,\ldots,p} \\ -\frac{1}{\tilde{\sigma}_n^4}\,\tilde{\zeta}^{(i)\,2}_t + \frac{\sigma_n^2}{\tilde{\sigma}_n^4} \end{pmatrix}$$

and let $\tilde{S}_{n,\nu} = \frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \sum_{i=1}^{\nu} \lambda'\,\tilde{\mathbf{Z}}^{(i)}_t$, where $\lambda \in \mathbb{R}^{p+1}$.

Then $\{S_{n,\nu}, \mathcal{F}_{n,\nu}, n \in \mathbb{N}, \nu = 1,\ldots,n\}$ is a martingale array with $\mathbb{E}\,S_{n,\nu} = 0$ and $\operatorname{var} S_{n,\nu} < \infty$. Furthermore, the differences $D_{n,i} = \frac{1}{\sqrt{n\,(T-p)}} \sum_{t=p+1}^{T} \lambda'\,\tilde{\mathbf{Z}}^{(i)}_t$ fulfil the other two conditions of Hall and Heyde's theorem 2.5.29. Thus $S_n = \sum_{i=1}^{n} D_{n,i}$ is

asymptotically normal with asymptotic variance $\lambda' \, 2\, \Gamma^\circ \, \lambda$. As in the ICM, case $n \to \infty$, it can be shown that $\mathbb{E} \left(\sqrt{n\,(T-p)}\, \nabla \mathcal{L}^\circ_{n,T}(\theta_a) - S_n \right)^2 = O\left(\frac{1}{n}\right)$ (lemma 2.5.26). This proves asymptotic normality for the gradient in the case of $n \to \infty$, T fixed.
Therefore we can conclude from the mean value theorem that

$$\sqrt{n\,(T-p)}\, \left(\hat{\theta}_a - \theta_a\right) \text{ is } \mathrm{AN}(0, 2\,\Gamma^{\circ\,-1}_n)\,,$$

covering all three cases $T \to \infty$, n fixed; $n, T \to \infty$, and $n \to \infty$, T fixed. □

We now can deduce the asymptotic distribution of the estimator of the parameter of the individual effects in the ICM, which allows us comparing the relative efficiencies of the ICM parameter estimator to the estimator of Hjellvik and Tjøstheim (1999a).

2.5.35 COROLLARY
In the special case of the ICM, the above theorem yields that

$$\sqrt{n\,(T-p)}\, \left(\hat{\theta}_a - (a', \sigma_0^2)'\right) \textit{ is } \mathrm{AN}\left(0, 2\,\frac{n}{n-1}\,\Gamma^{\circ\,-1}\right)\,,$$

where $\sigma_0^2 = \operatorname{var} \varepsilon_t^{(i)}$ *and* $\Gamma^\circ = \begin{pmatrix} 2\,B & 0 \\ 0 & \frac{1}{\sigma_0^4} \end{pmatrix} = \nabla^2 \mathcal{L}(\tilde{\theta}_0)$ *with* $B = (\Psi(k-l))_{k,l=1,\ldots,p}$ *(compare lemma 2.5.10).*

PROOF:
The random variables $\zeta_t^{(i)}$ in the GICM correspond to $\mathring{\varepsilon}_t^{(i)}$ in the ICM for all $t \in \mathbb{Z}$, $i = 1, \ldots, n$. Thus we get in particular that $\mathring{\zeta}_t^{(i)} = \mathring{\varepsilon}_t^{(i)} - \frac{1}{n}\sum_{i=1}^n \mathring{\varepsilon}_t^{(i)} = \mathring{\varepsilon}_t^{(i)} = \zeta_t^{(i)}$, where $\operatorname{cov}\left(\mathring{\varepsilon}_t^{(i)}, \mathring{\varepsilon}_t^{(j)}\right) = \left(\delta_{ij} - \frac{1}{n}\right) \sigma_0^2$, i.e. here $\tilde{\sigma}_n^2 = \sigma_0^2 = \lim_{n\to\infty} \tilde{\sigma}_n^2$ (remark 2.3.5). Thus the above theorem yields directly that

$$\sqrt{n\,(T-p)}\, \left(\hat{\theta}_a - (a', \sigma_0^2)'\right) \text{ is } \mathrm{AN}\left(0, 2\,\left(\frac{n-1}{n}\,\Gamma^\circ\right)^{-1}\right)\,. \qquad \square$$

2.5.36 REMARK
We have discussed in remark 2.4.8 that, if the model is restricted to the ICM case, the estimator of Hjellvik and Tjøstheim (1999a) equals the GICM estimator $\hat{\theta}_a$ of θ_a, which is obtained from the individual effects $\{\mathring{Z}_t^{(i)}\}_{t\in\mathbb{Z}} = \{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$. We can derive the asymptotic relative efficiencies of the autoregressive parameter estimators from the asymptotic variances obtained in theorem 2.5.20 and in the above theorem. For this, denote $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)'$ and $\hat{\theta}_a = (\hat{a}'_{HT}, \hat{\sigma}^2_{HT})'$. For $n \to \infty$ the ICM parameter estimator $(\hat{a}', \hat{\sigma}^2)'$ and the estimator $\hat{\theta}_a$ are asymptotically equivalent. Thus we could have obtained asymptotic normality of $\hat{\theta}_a$ in the special case of the ICM also in a direct way from $\mathcal{L}^\circ_{n,T}(\theta_a)$, as it has been done in Hjellvik and Tjøstheim (1999a), see also remark 2.4.8. However, if $T \to \infty$, n fixed, the asymptotic results in the ICM are based on $\mathcal{L}_n(\theta)$, which includes the information contained in $\{\bar{X}_t\}_{t\in\mathbb{Z}}$. In this case we get that the asymptotic variance of $\sqrt{n\,(T-p)}\,(\hat{a} - a)$ is B^{-1}, where $B = (\Psi(k-l))_{k,l=1,\ldots,p}$ is obtained from the upper left block of Γ_n (remark 2.5.23).

If we denote the upper left block of $\frac{n-1}{n}\,\Gamma^\circ$ by B°, we thus have that $\frac{n-1}{n}\,B = B^\circ$, i.e. $B^{\circ\,-1} = \frac{n}{n-1}\,B^{-1}$, Therefore the relative asymptotic efficiency of $\hat{a}$ to $\hat{a}_{HT}$ is for $T \to \infty$, n fixed,

$$\text{eff}_{rel}\,(\hat{a}, \hat{a}_{HT}) = \frac{n-1}{n}\,.$$

This effect is also illustrated in the simulations in the Appendix A.

2.6 Properties of the Parameter Estimators

In the last section we have proved consistency of the parameter estimators and have proved asymptotic normality in both the ICM (theorem 2.5.20) and the GICM (theorem 2.5.34). These results yield a $\sqrt{n\,T}$-rate of convergence. Now we investigate the asymptotic behaviour more closely. It is generally known (compare also remark 2.6.13 at the end of this section) that determining the mean squared error of the parameter estimators is a difficult task. We thus restrict ourselves to a stochastic expansion and examine the mean squared error of its dominating term. The calculation is straightforward for the parameter estimators in the GICM; in theorem 2.6.9 we derive the term responsible for the asymptotic behaviour of $\hat{a}_{HT} - a$, where $\hat{a}_{HT}$ is the estimator of Hjellvik and Tjøstheim (1999a) and a denotes the true autoregressive parameter. We however begin with the case of the ICM parameter estimator $\hat{a}$, where the structure of the estimators is more complex. Here we give an explicit expression of the dominating term in theorem 2.6.5.

Comparing the mean squared errors of the dominating terms we obtain the main result of this section: the ICM estimator $\hat{a}$ has not only a higher relative efficiency compared to $\hat{a}_{HT}$ (see remark 2.4.8), but also the mean squared error of the dominating term is smaller. In order to enhance readability, some proofs have been moved to the Appendix C.3.1. The section concludes with a short discussion of the bias in the ICM and GICM. In particular we prove the mean-square rate of convergence of the bias term.

Rates of Convergence

For reference we first recall some results obtained previously.

2.6.1 REMARK

1. We see from remark 2.4.3 and the algorithm 2.4.4 that, if the ICM parameter estimator $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)' = \text{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta) \in \text{Int}\Theta$, it fulfils $\hat{B}\,\hat{a} = \hat{A}$ with

$$\hat{B} = \frac{1}{\hat{\sigma}^2}\frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{\mathbf{x}}_{t-1}^{(i)}\,\mathring{\mathbf{x}}_{t-1}^{(i)\prime} + \frac{1}{\hat{\omega}_n^2}\frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\bar{\mathbf{x}}_{t-1}\,\bar{\mathbf{x}}_{t-1}'$$

and

$$\hat{A} = \frac{1}{\hat{\sigma}^2}\frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{\mathbf{x}}_{t-1}^{(i)}\,\mathring{X}_t^{(i)} + \frac{1}{\hat{\omega}_n^2}\frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\bar{\mathbf{x}}_{t-1}\,\bar{X}_t\,,$$

where $\mathring{\mathbf{x}}_{t-1}^{(i)} = \left(\mathring{X}_{t-1}^{(i)}, \ldots, \mathring{X}_{t-p}^{(i)}\right)'$, $i = 1, \ldots, n$, $\bar{\mathbf{x}}_{t-1} = \left(\bar{X}_{t-1}, \ldots, \bar{X}_{t-p}\right)'$.

$\hat{\tau}^2$ is obtained from $\hat{\sigma}^2$ and $\hat{\omega}_n^2$ by $\hat{\tau}^2 = \hat{\omega}_n^2 - \frac{\hat{\sigma}^2}{n}$.

In the ICM (assumption 2.2.1) we assume that the parameter space Θ is compact and that there exists a $c > 0$ such that for all $\theta = (a', \sigma^2, \tau^2) \in \Theta$ we have that $\sigma^2 \geq c$.

2. In the ICM, $\{\mathring{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, and $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ are autoregressive processes with the same autoregressive parameter $a = (a_1, \ldots, a_p)'$. For $h \in \mathbb{Z}$ their autocovariances are $\mathring{\gamma}_n(h) = \frac{n-1}{n}\,\sigma_0^2\,\Psi(h)$ and $\bar{\gamma}_n(h) = \omega_n^2\,\Psi(h) = \left(\tau_0^2 + \frac{\sigma_0^2}{n}\right)\,\Psi(h)$, where σ_0^2 and τ_0^2 denote the true variances of $\{\varepsilon_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, and $\{\eta_t\}_{t\in\mathbb{Z}}$ (see lemma 2.2.4). Therefore we know from standard theory (e.g. Brockwell and Davis 1991, p. 239) that a fulfils the Yule-Walker equation $B\,a = A$, where

$$B = (\Psi(k-l))_{k,l=1,\ldots,p} \qquad \text{and} \qquad A = (\Psi(l))_{l=1,\ldots,p}\ .$$

As B is derived from the autocovariance function, it is obvious that B is positive definite and thus invertible.

3. Since in the ICM $u_n = -\frac{1}{n-1} = O\left(\frac{1}{n}\right)$ (remark 2.2.5), the mean-square convergence property shown in lemma 1.2.4 gives for all $k, l = 0, \ldots, p$, that

$$\mathbb{E}\left(\frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(i)} - \mathring{\gamma}_n(k-l)\right)^2 = O\left(\frac{1}{n\,T}\right)\ .$$

Furthermore we have seen in remark 2.2.5 that

$$\mathbb{E}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}\bar{X}_{t-k}\,\bar{X}_{t-l} - \bar{\gamma}_n(k-l)\right)^2 = O\left(\frac{\omega_n^4}{T}\right)\ ,$$

where again $\mathring{\gamma}_n(h) = \frac{n-1}{n}\,\sigma_0^2\,\Psi(h)$ and $\bar{\gamma}_n(h) = \omega_n^2\,\Psi(h)$, $h \in \mathbb{Z}$.

Consistency of the parameter estimators can be obtained as a conclusion from the central result of the preceding section, where we have established asymptotic normality of the parameter estimators.

2.6.2 LEMMA

Let the true parameter in the ICM be $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$, and let $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$. Then the components of the ICM estimator $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)' = \mathrm{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta)$ admit the following rates of convergence:

$$(\hat{a}', \hat{\sigma}^2)' - (a', \sigma_0^2)' = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$$

and for the last entry we get that

$$\hat{\tau}^2 - \tau_0^2 = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right)\ .$$

Furthermore we have for $\hat{\omega}_n^2 = \hat{\tau}^2 + \frac{\hat{\sigma}^2}{n}$ *that*

$$\hat{\omega}_n^2 - \omega_n^2 = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) .$$

PROOF:
We have proved asymptotic normality of $\sqrt{n\,(T-p)}\,D_n^{-1}\left(\hat{\theta}_{n,T} - \theta_0\right)$ in the case of $T \to \infty$ in theorem 2.5.20 under assumption 2.5.8. There we furthermore have got in the case of $n \to \infty$, that still $\sqrt{n\,(T-p)}\,((\hat{a}', \hat{\sigma}^2)' - (a', \sigma_0^2)')$ is asymptotically normal. Assumption 2.5.8 implies a restriction of the parameter space as it requires $\tau_0^2 \in \mathrm{Int}\Theta$ and $\omega_n^2 \geq c_2 > 0$. For each θ_0 with $\tau_0^2 > 0$ we can choose a subspace $\Theta' \in \Theta$ such that this assumption is fulfilled (lemma 2.5.7). The case $\tau_0^2 = 0$ has been treated in theorem 2.5.22. There we get asymptotic normality of $\sqrt{n\,(T-p)}\,((\hat{a}', \hat{\sigma}^2)' - (a', \sigma_0^2)')$. Furthermore we can conclude from the proof of the theorem that

$$\hat{\tau}^2 - \tau_0^2 = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) \quad \text{and} \quad \hat{\omega}_n^2 - \omega_n^2 = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) ,$$

where $\hat{\omega}_n^2 = \hat{\tau}^2 + \frac{\hat{\sigma}^2}{n}$ and $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$. This expression is derived from the consistency of $\hat{a}$ and $\hat{\sigma}^2$ and is also valid if $\tau_0^2 > 0$ (see remark 2.5.23). The notation includes the dependence on τ_0^2 via ω_n^2 and covers the already proved case of $T \to \infty$, too. □

In particular the following conclusions are of practical interest.

2.6.3 COROLLARY
In the setting of the above lemma we have $\frac{1}{\hat{\sigma}^2} = O_P(1)$, $\frac{1}{n\,\hat{\omega}_n^2} = O_P\left(\frac{1}{n\,\omega_n^2}\right)$ *and moreover*

$$\frac{1}{\hat{\sigma}^2} - \frac{1}{\sigma_0^2} = O_P\left(\frac{1}{\sqrt{n\,T}}\right) \qquad \text{and} \qquad \frac{1}{n\,\hat{\omega}_n^2} - \frac{1}{n\,\omega_n^2} = O_P\left(\frac{1}{n\,\omega_n^2\,\sqrt{T}}\right) .$$

The estimator $\hat{B}$ *of the covariance matrix, given in remark 2.6.1, fulfils that*

$$\hat{B} - B = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$$

where B *is the matrix* $B = (\Psi(k-l))_{k,l=1,\dots,p}$.

PROOF:
For the proof of this lemma is straightforward. In order to enhance readability we have moved it to the Appendix C.3.1. □

2.6.4 REMARK
The asymptotic behaviour of $\hat{\omega}_n^2$ and $\hat{\tau}^2$ in the case of $n \to \infty$, T fixed is determined by the actual realisation of the process $\{\eta_t\}_{t\in\mathbb{Z}}$. Let $Y_t = \sum_{u=0}^{\infty} \psi_u\,\eta_{t-u}$ for all $t \in \mathbb{Z}$ and denote $\tilde{Y}_T = \frac{1}{T-p}\sum_{t=p+1}^{T}\sum_{k,l=0}^{p} a_k\,a_l\,Y_{t-k}\,Y_{t-l}$, where $a_0 = -1$ and $(a_1, \dots, a_p)'$ is the true autoregressive parameter in the ICM. In the proof of theorem 2.5.16 we

have seen that $\mathbb{E}\left(\bar{X}_{t-k}\,\bar{X}_{t-l} - Y_{t-k}\,Y_{t-l}\right)^2 = O\left(\frac{1}{n}\right)$. Since $\hat{\omega}_n^2$ is estimated through $\hat{\omega}_n^2 = B_{n,T}(\hat{a}) = \frac{1}{T-p}\sum_{t=p+1}^{T}\sum_{k,l=0}^{p}\hat{a}_k\,\hat{a}_l\,\bar{X}_{t-k}\,\bar{X}_{t-l}$ (remark 2.4.3), the consistency of $\hat{a}$ implies that $\hat{\omega}_n^2 - \tilde{Y}_T = O_P\left(\frac{1}{\sqrt{n}}\right)$ (see the proof of lemma 2.6.2). Therefore also $\hat{\tau}^2 - \tilde{Y}_T = \hat{\omega}_n^2 - \frac{\hat{\sigma}^2}{n} - \tilde{Y}_T = O_P\left(\frac{1}{\sqrt{n}}\right)$. Note that $\tilde{Y}_T - \tau_0^2 = O_P\left(\frac{1}{\sqrt{T}}\right)$ (remark 2.5.17).

We now identify the main term responsible for the convergence behaviour of $\hat{a}$ by proving that there exists a $\tilde{C}_{n,T}$ such that $\hat{a} - a$ can be written as $\hat{a} - a = B^{-1}\tilde{C}_{n,T} + O_P\left(\frac{1}{n\,T}\right)$ with dominating term $B^{-1}\tilde{C}_{n,T}$. Since the bias is of lower order, the asymptotic behaviour mainly depends on the dominating term in the stochastic expansion. Thus the explicit expression for the dominating term enables us to compare the large sample properties of the ICM parameter estimator $\hat{\theta}_{n,T}$ to those of the GICM estimator $\hat{\theta}_a$. The proof of the theorem is based on a recursive representation of $\hat{a} - a$ similar to the one in Dahlhaus and Giraitis (1998), where the rates of the single terms can be obtained using the above consistency results.

2.6.5 THEOREM

In the setting of the ICM (assumption 2.2.1) let $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\tau}^2)' = \operatorname{argmin}_{\theta\in\Theta}\mathcal{L}_{n,T}(\theta)$ *be obtained as described in remark 2.4.3. Denote the true parameter by* $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$ *and let* $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$. *Furthermore let*

$$\hat{C}_1 = \frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{\mathbf{x}}_{t-1}^{(i)}\,\mathring{\varepsilon}_t^{(i)} \quad \text{and} \quad \hat{C}_2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\bar{\mathbf{x}}_{t-1}\,\left(\eta_t + \bar{\varepsilon}_t\right)\,,$$

where $\mathring{\mathbf{x}}_{t-1}^{(i)} = \left(\mathring{X}_{t-1}^{(i)},\ldots,\mathring{X}_{t-p}^{(i)}\right)'$ *and* $\bar{\mathbf{x}}_{t-1} = \left(\bar{X}_{t-1},\ldots,\bar{X}_{t-p}\right)'$ *for all* $t \geq p+1$, $i = 1,\ldots,n$*; and define* $\tilde{C}_{n,T}$ *as*

$$\tilde{C}_{n,T} = \frac{1}{\sigma_0^2}\,\hat{C}_1 + \frac{1}{n\,\omega_n^2}\,\hat{C}_2\,.$$

Then

$$\hat{a} - a = B^{-1}\,\tilde{C}_{n,T} + O_P\left(\frac{1}{n\,T}\right)\,.$$

with dominating term $B^{-1}\,\tilde{C}_{n,T} = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$.

PROOF:

Throughout this proof we use the notations of remark 2.6.1. There we have seen that $\hat{B}\,\hat{a} = \hat{A}$, $B\,a = A$ and furthermore that B is invertible. In the ICM the true parameter $a = (a_1,\ldots,a_p)'$ fulfils

$$\mathring{X}_t^{(i)} = \mathring{\mathbf{x}}_{t-1}^{(i)\prime}\,a + \mathring{\varepsilon}_t^{(i)} \quad \text{and} \quad \bar{X}_t = \bar{\mathbf{x}}_{t-1}'\,a + (\eta_t + \bar{\varepsilon}_t) \quad \text{for all } t\in\mathbb{Z}, i = 1,\ldots,n.$$

Thus it is straightforward that a also fulfils the equation $\hat{B}\,a = \hat{A} - \hat{C}_{n,T}$, where $\hat{C}_{n,T}$ is given by $\hat{C}_{n,T} = \frac{1}{\hat{\sigma}^2}\,\hat{C}_1 + \frac{1}{n\,\hat{\omega}_n^2}\,\hat{C}_2$. This means that $\hat{C}_{n,T} = \hat{A} - \hat{B}\,a = \hat{B}\,(\hat{a} - a)$, which leads to

$$B\,(\hat{a} - a) = B\,(\hat{a} - a) - \hat{B}\,(\hat{a} - a) + \hat{B}\,(\hat{a} - a) = (B - \hat{B})\,(\hat{a} - a) + \hat{C}_{n,T}\,.$$

As we know from corollary 2.6.3 that $B - \hat{B} = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$, we get from the rate of convergence of $\hat{a}$ that $(B-\hat{B})\,(\hat{a}-a) = O_P\left(\frac{1}{n\,T}\right)$. The term $\hat{C}_{n,T}$ can be split further. It is straightforward to show that $\frac{1}{\hat{\sigma}^2}\,\hat{C}_1 = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$ and $\frac{1}{n\,\hat{\omega}_n^2}\,\hat{C}_2 = O_P\left(\frac{1}{n\,\sqrt{T}}\right)$, which also means that $\hat{C}_1 = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$ and $\hat{C}_2 = O_P\left(\frac{\omega_n}{n\,\sqrt{T}}\right)$. The details of the proof can be found in the Appendix C.3.1, lemma C.3.1. Under the assumptions on the parameter space it thus is easily seen from lemma 2.6.2 and its corollary 2.6.3 that $\hat{C}_{n,T}$ fulfils

$$\begin{aligned}\hat{C}_{n,T} &= \frac{1}{\sigma_0^2}\,\hat{C}_1 + \frac{1}{n\,\omega_n^2}\,\hat{C}_2 + \frac{1}{\hat{\sigma}^2\,\sigma_0^2}\,\left(\sigma_0^2 - \hat{\sigma}^2\right)\,\hat{C}_1 + \frac{1}{n\,\hat{\omega}_n^2\,\omega_n^2}\,\left(\omega_n^2 - \hat{\omega}_n^2\right)\,\hat{C}_2 \\ &= \tilde{C}_{n,T} + O_P\left(\frac{1}{n\,T}\right)\,,\end{aligned}$$

and that $\tilde{C}_{n,T} = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$. This yields the result. □

2.6.6 REMARK

It is not surprising that $\hat{C}_{n,T} = \frac{1}{\hat{\sigma}^2}\,\hat{C}_1 + \frac{1}{n\,\hat{\omega}_n^2}\,\hat{C}_2$ is closely related to $\nabla\mathcal{L}_{n,T}(\theta_0)$. Indeed we have that $-2\,\hat{C}_{n,T} = \frac{\partial}{\partial a}\mathcal{L}_{n,T}(\theta)$ for $\theta = (a', \hat{\sigma}^2, \hat{\tau}^2)'$, where a denotes the true autoregressive parameter in the ICM, but $\hat{\sigma}^2$ and $\hat{\tau}^2$ are the parameter estimators obtained from the recursive algorithm. For $\tilde{C}_{n,T}$ we even have that $-2\,\tilde{C}_{n,T} = \frac{\partial}{\partial a}\mathcal{L}_{n,T}(\theta_0)$. Note that the factor 2, which was only introduced into the likelihood function for computational convenience, cancels out in the representation $\hat{a} - a = B^{-1}\,\hat{C}_{n,T} + O_P\left(\frac{1}{n\,T}\right)$ because $B = (\Psi(k-l))_{k,l=1,\ldots,p} = \frac{1}{2}\,\frac{\partial^2}{(\partial a)^2}\mathcal{L}(\theta_0)$ (see remark 2.5.21). As $\hat{B}$ is the estimator of $B = \frac{\partial^2}{(\partial a)^2}\mathcal{L}(\theta_0)$, the equation $\hat{C}_{n,T} = \hat{B}\,(\hat{a} - a)$ is an empirical counterpart of the often used representation based on the mean value theorem (see e.g. the proof of theorem 2.5.20). However it cannot directly be employed for the parameter estimation due to the intercorrelation present in the ICM. For example, it is not possible to obtain consistency of $\hat{a}$ in a direct way because of the recursive estimation procedure needed. This implies in turn that we e.g. cannot easily identify with a direct method the lower order terms in the above representation of $\hat{C}_{n,T} = \tilde{C}_{n,T} + O_P\left(\frac{1}{n\,T}\right)$.

We can moreover give an explicit formula for computing the mean squared error of the dominating term.

2.6.7 PROPOSITION

Using the notations of the preceding theorem, we get for the mean squared error of $\hat{m} = B^{-1}\,\tilde{C}_{n,T}$ *that*

$$\mathbb{E}\,||\hat{m}||^2 = \frac{1}{n\,(T-p)}\,\mathrm{tr}\left(B^{-1}\right)\,.$$

PROOF:

$\tilde{C}_{n,T}$ was defined in the preceding theorem as $\tilde{C}_{n,T} = \frac{1}{\sigma_0^2}\,\hat{C}_1 + \frac{1}{n\,\omega_n^2}\hat{C}_2$. In the ICM, $\hat{C}_1$ and $\hat{C}_2$ are independent because the processes $\{\mathring{\varepsilon}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1,\ldots,n$, are independent of $\{\eta_t + \bar{\varepsilon}_t\}_{t\in\mathbb{Z}}$ (see remark 2.2.3). Denoting the entries of B^{-1} by $b_{k,l}$, $k,l = 1,\ldots,p$ we therefore get

$$\mathbb{E}\,||\hat{m}||^2 = \mathbb{E}\,||B^{-1}\,\hat{C}_1||^2 + \mathbb{E}\,||B^{-1}\,\hat{C}_2||^2$$

$$= \sum_{g=1}^{p} \sum_{k,l=1}^{p} b_{g,k}\, b_{g,l} \left(\frac{1}{\sigma_0^4} \frac{n-1}{n^2\,(T-p)} \Psi(k-l)\,\sigma_0^4 + \frac{1}{n^2\,\omega_0^4} \Psi(k-l)\,\omega_0^4 \right)$$

$$= \frac{1}{n\,(T-p)} \sum_{g=1}^{p} \sum_{k,l=1}^{p} b_{g,k}\, b_{g,l}\, \Psi(k-l) \;=\; \frac{1}{n\,(T-p)} \operatorname{tr}\left(B^{-1}\right) ,$$

as $B = (\Psi(k-l))_{k,l=1,\dots,p}$ and thus $\sum_{k,l=1}^{p} b_{g,k}\,\Psi(k-l)\,b_{g,l} = \sum_{l=1}^{p} \delta_{gk}\, b_{g,l} = b_{g,g}$, where δ_{gk} denotes the Kronecker delta. □

In particular we now have an explicit formula in the AR(1) case.

2.6.8 COROLLARY
As a special case of the last proposition we get for a panel of AR(1) processes with autoregressive parameter a, $|a| < 1$, that

$$\mathbb{E}\,||\hat{m}||^2 = \frac{1-a^2}{n\,(T-p)} .$$

PROOF:
In the AR(1) case $\Psi(0) = \sum_{u=0}^{\infty} \psi_u^2 = \sum_{u=0}^{\infty} a^{2u} = (1-a^2)^{-1}$. This yields

$$\mathbb{E}\,||\hat{m}||^2 = \frac{1}{\Psi(0)\,n\,(T-p)} = \frac{1-a^2}{n\,(T-p)} . \qquad \square$$

We can use the same method for investigating the convergence properties of Hjellvik and Tjøstheim's estimator $\hat{a}_{HT}$.

2.6.9 PROPOSITION
Let $B_1 = \frac{n-1}{n}\,\sigma_0^2\, B$. Then the estimator $\hat{\theta}_a = (\hat{a}_{HT}', \hat{\sigma}_{HT}^2)' = \operatorname{argmin}_{\theta\in\Theta_a} \mathcal{L}_{n,T}^{\circ}(\theta)$ obtained in proposition 2.4.7 fulfils

$$\hat{a}_{HT} - a = B_1^{-1}\,\hat{C}_1 + O_P\left(\frac{1}{n\,T}\right) ,$$

where the mean squared error of the dominating term is

$$\mathbb{E}\,||B_1^{-1}\,\hat{C}_1||^2 = \frac{1}{(n-1)\,(T-p)} \operatorname{tr}\left(B^{-1}\right) .$$

PROOF:
We know from proposition 2.4.7 that $\hat{a}_{HT}$ fulfils $\hat{B}_1\,\hat{a}_{HT} = \hat{A}_1$, where

$$\hat{B}_1 = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{\mathbf{x}}_{t-1}^{(i)}\, \mathring{\mathbf{x}}_{t-1}^{(i)\prime}$$

$$\text{and} \qquad \hat{A}_1 = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{\mathbf{x}}_{t-1}^{(i)}\, \mathring{X}_t^{(i)}$$

with $\mathring{\mathbf{x}}_{t-1}^{(i)} = \left(\mathring{X}_{t-1}^{(i)}, \ldots, \mathring{X}_{t-p}^{(i)}\right)'$. It is easily seen that $\mathbb{E}\,\hat{B}_1 = B_1$. Thus we get similar to the proof of theorem 2.6.5 that

$$\hat{a}_{HT} - a = B_1^{-1}\,\hat{C}_1 + B_1^{-1}\,\left(B_1 - \hat{B}_1\right)\,(\hat{a}_{HT} - a)\ ,$$

with $\hat{C}_1$ as in theorem 2.6.5. Due to the mean-square convergence of the panel autocovariance estimator (see remark 2.6.1), we have that $B_1 - \hat{B}_1 = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$. Since $\hat{a}_{HT} - a = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$ (lemma 2.5.34), we get as in the proof of theorem 2.6.5 that

$$\hat{a}_{HT} - a = B_1^{-1}\,\hat{C}_1 + O_P\left(\frac{1}{n\,T}\right)\ .$$

As we can see from the proof of lemma C.3.1 in the Appendix C.3.1, $\hat{C}_1$ fulfils that

$$\mathbb{E}\,||\hat{C}_1||^2 = \frac{p\,(n-1)}{n^2\,(T-p)}\,\sigma_0^4\,\Psi(0)\,.$$

Analogously to the proof of proposition 2.6.7 we furthermore obtain that

$$\begin{aligned}\mathbb{E}\,||B_1^{-1}\,\hat{C}_1||^2 &= \frac{n^2}{(n-1)^2\,\sigma_0^4}\sum_{g=1}^{p}\sum_{k,l=1}^{p} b_{g,k}\,b_{g,l}\,\frac{n-1}{n^2\,(T-p)}\,\Psi(k-l)\,\sigma_0^4\\ &= \frac{1}{(n-1)\,(T-p)}\,\mathrm{tr}\left(B^{-1}\right)\ ,\end{aligned}$$

if we again denote the entries of B^{-1} by $b_{k,l}$, $k, l = 1, \ldots, p$. □

These differences in the asymptotic behaviour of the ICM parameter estimator $\hat{a}$ and the estimator $\hat{a}_{HT}$ of Hjellvik and Tjøstheim (1999a) are the main result of this section. They show that the dominating term in the stochastic expansion has a smaller mean squared errof than the corresponding term based on $\hat{a}_{HT}$.

2.6.10 REMARK
The above results illustrate again the differences between the ICM estimator $\hat{\theta}_{n,T}$ and $\hat{\theta}_a$, which is the estimator of Hjellvik and Tjøstheim (1999a) (see remark 2.4.8). We have already seen in theorem 2.5.20, where we have summarised the results on the asymptotic normality of $\hat{\theta}_{n,T}$ in the ICM, that in the case of $T \to \infty$, n fixed, the asymptotic variance of the ICM estimator $\hat{a}$ is B^{-1}, whereas in the GICM it is $\frac{n}{n-1}\,B^{-1}$ (see theorem 2.5.34). Thus the relative asymptotic efficiency of $\hat{a}$ compared to Hjellvik and Tjøstheim's estimator $\hat{a}_{HT}$, which is the same estimator as the estimator of a in the GICM, is in this case $\mathrm{eff}_{rel}(\hat{a}, \hat{a}_{HT}) = \frac{n-1}{n}$. This already has been discussed in remark 2.5.36. The above considerations further show that already the ratio of the mean squared errors of the respective dominating terms equals $\frac{n-1}{n}$. These properties also are illustrated by the simulations (Appendix A).

Bias in the ICM and GICM

We conclude with some considerations on the bias of $\hat{a}$.

2.6.11 REMARK
In the proof of theorem 2.6.5 we have seen that $\hat{C}_{n,T} = \hat{A} - \hat{B}\,a$. We thus get

$$\begin{aligned}\hat{a} - a &= \hat{B}^{-1}\left(\hat{A} - \hat{B}\,a\right) = B^{-1}\,\hat{C}_{n,T} + \left(\hat{B}^{-1} - B^{-1}\right)\hat{C}_{n,T} \\ &= B^{-1}\,\hat{C}_{n,T} + \hat{B}^{-1}\left(B - \hat{B}\right)B^{-1}\,\hat{C}_{n,T} \\ &= B^{-1}\,\hat{C}_{n,T} + B^{-1}\left(B - \hat{B}\right)B^{-1}\,\hat{C}_{n,T} \\ &\qquad + \hat{B}^{-1}\left(B - \hat{B}\right)B^{-1}\left(B - \hat{B}\right)B^{-1}\,\hat{C}_{n,T}\,.\end{aligned}$$

The first term fulfils, as we have seen in the proof of theorem 2.6.5, that

$$B^{-1}\,\hat{C}_{n,T} = B^{-1}\,\tilde{C}_{n,T} + \frac{1}{\hat{\sigma}^2\,\sigma_0^2}\left(\sigma_0^2 - \hat{\sigma}^2\right)\hat{C}_1 + \frac{1}{n\,\hat{\omega}_n^2\,\omega_n^2}\left(\omega_n^2 - \hat{\omega}_n^2\right)\hat{C}_2$$

where $B^{-1}\,\tilde{C}_{n,T} = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$ and the other terms are of order $O_P\left(\frac{1}{n\,T}\right)$. As also $B - \hat{B} = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$ (corollary 2.6.3), the third term in the expansion is of lower order:

$$\hat{B}^{-1}\left(B - \hat{B}\right)B^{-1}\left(B - \hat{B}\right)B^{-1}\,\hat{C}_{n,T} = O_P\left(\frac{1}{(n\,T)^{3/2}}\right).$$

The orthogonality properties of the innovations imply that $\mathbb{E}\left(B^{-1}\,\tilde{C}_{n,T}\right) = 0$. Thus the main bias term is

$$\begin{aligned}\hat{\beta}_{n,T} &= B^{-1}\left(B - \hat{B}\right)B^{-1}\,\hat{C}_{n,T} \\ &\qquad + \frac{1}{\hat{\sigma}^2\,\sigma_0^2}\left(\sigma_0^2 - \hat{\sigma}^2\right)\hat{C}_1 + \frac{1}{n\,\hat{\omega}_n^2\,\omega_n^2}\left(\omega_n^2 - \hat{\omega}_n^2\right)\hat{C}_2\end{aligned}$$

and fulfils $\hat{\beta}_{n,T} = O_P\left(\frac{1}{n\,T}\right)$ but $\mathbb{E}\,\hat{\beta}_{n,T} \neq 0$.
In the GICM, we do not have weighted averages in the estimator. Here simply

$$\begin{aligned}\hat{a} - a &= B_1^{-1}\,\hat{C}_1 + B_1^{-1}\left(B_1 - \hat{B}_1\right)B_1^{-1}\,\hat{C}_1 + o_P\left(\frac{1}{n\,T}\right) \\ \text{and}\quad \hat{b} - b &= B_2^{-1}\,\hat{C}_2 + B_2^{-1}\left(B_2 - \hat{B}_2\right)B_2^{-1}\,\hat{C}_2 + o_P\left(\frac{1}{T}\right).\end{aligned}$$

Thus we get analogously to the above cosiderations that the bias terms $\hat{\beta}_1$ and $\hat{\beta}_2$ are

$$\hat{\beta}_1 = B_1^{-1}\left(B_1 - \hat{B}_1\right)B_1^{-1}\,\hat{C}_1 \quad\text{and}\quad \hat{\beta}_2 = B_2^{-1}\left(B_2 - \hat{B}_2\right)B_2^{-1}\,\hat{C}_2\,,$$

where $\hat{C}_1$ and $\hat{C}_2$ are given in lemma C.3.1; $\hat{B}_1 = \frac{1}{n\,(T-p)}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\mathring{\mathbf{x}}_{t-1}^{(i)}\mathring{\mathbf{x}}_{t-1}^{(i)\prime}$, $\hat{B}_2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\bar{\mathbf{x}}_{t-1}\bar{\mathbf{x}}_{t-1}'$, and furthermore $B_1 = \mathbb{E}\,\hat{B}_1$ and $B_2 = \mathbb{E}\,\hat{B}_2$.

Due to $\hat{B}$ and $\hat{C}_{n,T}$ being weighted averages with random weights, it is difficult to obtain $\mathbb{E}\,\hat{\beta}_{n,T}$ or $\mathbb{E}\,||\hat{\beta}_{n,T}||^2$ in the ICM. Therefore we restrict ourselves in the following to the case of the GICM, where the variances cancel out in the computation of the estimators. Here we obtain the mean-square rates of convergence.

2.6.12 LEMMA
Using the notations of the preceding remark, we get for the bias terms $\hat{\beta}_1$ and $\hat{\beta}_2$ in the GICM that

$$\mathbb{E}\,||\hat{\beta}_1||^2 = O\left(\frac{1}{n^2\,T^2}\right) \quad \textit{and} \quad \mathbb{E}\,||\hat{\beta}_2||^2 = O\left(\frac{1}{T^2}\right).$$

PROOF:
The proof requires some results on cumulants and some considerations on the relations between the autoregressive parameters a and the autocovariance function $\gamma(h)$, $h \in \mathbb{Z}$, of an autoregressive process $\{X_t\}_{t\in\mathbb{Z}}$. These are developed in the Appendix C.3.2. As the proof of the lemma involves some lengthy calculations, we have moved it to the end of the Appendix C.3.2. □

2.6.13 REMARK

1. The rates derived in this section correspond to the standard theory for Yule-Walker and least squares estimation, where the bias is of order n^{-1} in the number of observations, while the asymptotic standard deviation is of order $n^{-1/2}$. Tjøstheim and Paulsen (1983) have shown that for Yule-Walker estimators the coefficient of the bias term may become large as it depends on the roots of the characteristic polynomial of the autoregressive process. Thus here the bias term can become appreciably larger than the $n^{-1/2}$-term of the standard deviation for a wide range of n. However this is not the case for least squares estimators, which we get in the conditional maximum likelihood estimation under the assumption of Gaussianity. Therefore we do not need to derive the coefficient of the bias term as in our setting it is sufficient to regard the behaviour of the $n^{-1/2}$-term.

2. In the case of the ICM it would already have been very difficult to derive the mean-square rate of convergence of the main bias term in the stochastic expansion, as the method used in the preceding lemma for calculating the mean squared error of $\hat{\beta}_1$ and $\hat{\beta}_2$ cannot be directly applied to the case of the ICM. We have seen in remark 2.6.11 that the main bias term in the ICM, $\hat{\beta}_{n,T}$, has a more complicated structure. Since the different terms are weighted with $\hat{\sigma}^2$ and $\hat{\omega}_n^2$, one would have to find a representation of $\hat{\beta}_{n,T}$ which makes a Taylor expansion possible. This is not easily obtained. Furthermore it is not clear under which conditions the Taylor expansion of e.g. $\frac{1}{\hat{\sigma}^2}$ around $\frac{1}{\sigma_0^2}$ can be obtained as $\hat{\sigma}^2$ in turn depends on the estimator of the autoregressive parameter, $\hat{a}$.

3. Directly calculating the large sample bias for the maximum likelihood estimator in an autoregressive process is a difficult task in general. This is shown by the existing literature. Shaman and Stine (1988) give formulae for the bias of Yule-Walker and least squares estimators of the autoregressive parameter a. Their proof is based on the assumption that $\mathbb{E}\,||\hat{B}^{-1} - B^{-1}||^8_{spec}$ is bounded, where

$||A||_{spec} = \max\{|\lambda|; \lambda$ eigenvalue of $A\}$ denotes the spectral norm of a matrix. This cannot easily be verified. Tjøstheim and Paulsen (1983) derive an explicit formula of the bias based on the zeros of the autoregressive polynomial. However they focus on the AR(2) case, and their method cannot be generalised to higher autoregressive orders in a straightforward way.

2.7 Discussion

The ICM (assumption 2.2.1) and its generalisation, the GICM (assumption 2.3.1), are two models which include a common effect influencing all time series in a panel simultaneously. In particular the GICM is very flexible as it allows the common effect to have e.g. an autoregressive structure. Possible applications include the investigation of panels of time series in medical studies, in population dynamics or even the modelling of business cycles (if it can be assumed that a single cycle is responsible for the common structure). The latter model is a special case of the dynamic factor model developed by Forni et al. (2000). However the focus here is on parameter estimation, whereas Forni et al. are predominantly concerned with finding the number of common factors. For a more detailed comparison see remark 2.3.3.
The ICM has been investigate before by Hjellvik and Tjøstheim (1999a), who treat the common influence η_t as a nuisance parameter which they eliminate by subtracting the mean process $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ from each time series. The factorisation of the conditional likelihood obtained in theorem 2.4.2 now makes it possible to obtain a log-likelihood function which uses the information contained in $\bar{X}_t$ as well. The corresponding estimator is based on weighted averages of two sample covariance matrices. Thus we have to employ a recursive algorithm for the estimation. Due to this structure, we furthermore are not able to obtain an asymptotic normality result directly but have to use a proof based on a uniform convergence condition on $\mathcal{L}_{n,T}$. Additionally we need in the case of $T \to \infty$ a martingale limit theorem for the last step of the proof. Hjellvik and Tjøstheim's estimator $\hat{a}_{HT}$ can be computed by minimising $\mathcal{L}^{\circ}_{\underline{n},T}(\theta)$ (see remark 2.4.8). If the number n of time series in the panel tends to infinity, $\bar{X}_t \to 0$, which means that the information contained in the mean process loses weight. In this case the ICM estimator $\hat{a}$ is asymptotically equivalent to $\hat{a}_{HT}$; but in the case of small n, $\hat{a}$ is more efficient than $\hat{a}_{HT}$. Indeed we have seen that the relative asymptotic efficiency is $\text{eff}_{rel}(\hat{a}, \hat{a}_{HT}) = \frac{n-1}{n}$ (remark 2.5.36). This is also the ratio of the mean squared errors of the dominating terms in the stochastic expansion (remark 2.6.10). Simulations show that in practice the ICM parameter estimator performs as well as the estimator of Hjellvik and Tjøstheim in spite of the iterative algorithm employed for its computation (see the Appendix A and the discussion in section A.4). Moreover they illustrate the theoretical properties of the estimators. Thus we conclude that if the data can be modelled using the ICM and if n is not very large, the ICM algorithm should be used. If n is large, or if the order of the autoregressive process is high, it is computationally more convenient to calculate $\hat{a}_{HT}$, which is sufficiently accurate. We must always use the GICM if there is no theoretical reason why the underlying background process should have the same dynamical structure as the residual processes, i.e. if we cannot assume that the common error $\{\eta_t\}_{t\in\mathbb{Z}}$ is a white noise process. If necessary, one could test whether $\hat{a}$ and $\hat{b}$ coincide, which can be achieved using a bootstrap procedure.

We only have briefly discussed the calculation of the bias in section 2.6 as it is very difficult to derive. In the ICM the computation is complicated by the complex structure of the estimator which makes it necessary to use a recursive procedure for estimation. However it is in general difficult to calculate the bias for a conditional maximum likelihood estimator. This is due to the fact that in contrast to the Yule-Walker procedure all components of the estimated covariance matrix and the corresponding autocovariance vector are distinct. Thus $||\hat{a} - a||$ cannot be bounded as it is done in Dahlhaus and Giraitis (1998) (see also Whittle (1963) cited therein). Moreover the influence of the bias term on the asymptotic behaviour of the parameter estimator is small for least squares estimators (see remark 2.6.13). Therefore it is not necessary to calculate the bias explicitely in our case. These effects are discussed in the literature on large sample bias estimation (Tjøstheim and Paulsen 1983, Shaman and Stine 1988). There formulae for the asymptotic bias are given. However the complexity of the problem is reflected by the restrictions used. Tjøstheim and Paulsen (1983) focus on the AR(2) case and Shaman and Stine (1988) rely on the assumption that $\mathbb{E}|| \left(\hat{B}^{-1} - B^{-1}\right)^8 ||_{spec} < \infty$, which means that the 8th moment of the eigenvalues is bounded. This cannot easily be verified. Kiviet (1995) derives an approximate small sample bias in a model also containing exogenous variables, but the dynamic part is restricted to first-order autoregressive models.

There are several possible extensions of the above models. A generalisation to non-parametric intercorrelated models is discussed in Hjellvik et al. (2004). However this is beyond the scope of the present thesis, as we here contend ourselves to parametric models. Besides including more than one common factor in the GICM as in Forni et al. (2000), one could assume that the autoregressive parameters of the residual processes are not fixed but for example normally distributed with $a^{(i)} \sim \mathrm{N}(a, \Sigma_a)$. Similarly one could investigate a model with a common autoregressive parameter where the residual processes are allowed to have distinct variances σ_i^2, $i = 1, \ldots, n$. Furthermore we have not included explanatory variables in the model. However the proofs cannot directly be generalised to any of those cases. Another question is whether we can omit the condition of Gaussianity. This should be possible, but then the conditional log-likelihood function in the ICM would lose its convenient structure. The estimator of Hjellvik and Tjøstheim (1999a) is consistent for any distribution of η_t as they treat η_t as a nuisance parameter which is eliminated in the analysis (remark 2.4.8). Thus in the non-Gaussian case the possible gain in information by including $\bar{X}_t$ in the ICM procedure is outweighed by the additional complexity.

Throughout the thesis we have assumed that the order of the autoregressive process is known. For practical applications we need a model selection criterion. This can be obtained from the residual variances as in Hjellvik and Tjøstheim (1999b). Furthermore, e.g. in population dynamics, clustering is an important aspect. Yao et al. (2000) propose a method for detecting common structure in panels of uncorrelated time series and use it for clustering mink and muskrat data. If this could be extended to the intercorrelated case, it would allow for a broader modelling of biological processes where geographic conditions, which locally imply a spatial homogeneity, affect the structure of the intercorrelated time series.

Chapter 3

Robust Estimation

3.1 Introduction

The second part of this thesis is concerned with the question of finding robust estimates for parameters in the case of a panel of time series. Here two major types of outliers can occur. The first kind are arbitrary outliers, which are e.g. due to measurement errors. In the panel case we furthermore consider the case that an entire time series is generated by a different model. This is motivated by our applications. In a medical study, outliers can be due to false recordings. But it also can occur that some patient has been wrongly assigned to the treatment group. Then the time series obtained from this patient may be driven by a different dynamical structure and the entire time series can be viewed as outlying. If it is not possible to identify and exclude the outliers prior to a data analysis, we need procedures which remain stable under contamination.

We propose in this chapter several robust methods based on three different concepts. Our main focus is on the second type of outliers. First we describe possible robustifying procedures for the parameter estimator derived in the first chapter and evaluate their performance. Moreover we investigate exemplarily for these estimators how bootstrap can be used to improve the estimates. Then we discuss some methods based on robustifying the autocovariance matrix and the autocovariance vector used in the Yule-Walker / least squares equations. Here the focus is on the parameter estimator derived from the robust scale estimator $Q_{n,T}$ introduced by Rousseeuw and Croux (1993), as it allows us to obtain a breakdown point for the panel estimator. As a reference we include two methods designed for robust regression. Finally we investigate two methods for outlier detection which can be used to find and exclude outliers in a first step before performing a non-robust analysis. The chapter concludes with a comparative evaluation of the different methods described. The various estimators have been evaluated in a simulation study which can be found in the Appendix B.

3.2 Outliers

Robust inference is concerned with estimation in the presence of outliers. Measured data may contain 10 – 15% of outlying data (Hampel 1973), but this proportion can even reach 30% (Huber 1981). In the time series context one can distinguish several types of outliers: *Innovation outliers (IO)* are due to contamination in the innovations, i.e. of the time series $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ which drives the process. In least squares estimation and thus also in conditional maximum likelihood estimation under assumption of Gaussianity, they lead to "good" leverage points: Since a large outlier in the innovations also influences the subsequent values due to the autoregressive structure of the time series, the least squares estimate for an autoregressive parameter can even be improved in the presence of an innovation outlier (Rousseeuw and Leroy 1987). Therefore, we are more concerned with the so-called *additive outliers (AO)*. They appear when one point of the time series itself is changed directly. They can be modelled as genuinely "additive" outliers, i.e. that X_t is replaced by $X_t + W_t$, where e.g. $W_t \sim \mathrm{N}(0, \sigma^2)$ for some $\sigma^2 > 0$. We here regard them as "replacement" outliers, which are obtained by replacing some value X_t by a value from a second time series W_t. This is also called "epsilon-contamination" as the observed process $\tilde{X}_t$ is generated according to $\tilde{X}_t = (1-\delta_t)\, X_t + \delta_t W_t$, where $\mathbb{P}(\delta_t = 1) = \varepsilon = 1 - \mathbb{P}(\delta_t = 0)$. Detection becomes more complicated if there are patches of additive outliers, i.e. if the outliers are dependent. For a brief discussion see Rousseeuw and Leroy (1987).

In the panel case, one also can assume that the single time series in the panel do not contain outliers, but that entire time series may be generated by another model. This depends on the application intended. In econometrics, the standard procedure is testing panels of time series for homogeneity first (Hsiao 1986). If this test does not reject the null hypothesis of homogeneity, one common parameter is estimated for all time series of the panel. In the heterogeneous case, the parameters are estimated separately for each time series. Depending on the context it is however justified to make the above assumption, for example if the model states that time series of patients having the same affliction follow the same dynamics. Here a patient who suffers from a different disease could have been wrongly assigned to the therapy group. Since we are not interested in the dynamical structure of the outlying time series, we focus on robust procedures for parameter estimation. Thus we now assume that outliers can either be single replacement outliers or generated by replacing complete time series by time series following another model.

3.2.1 ASSUMPTION

Let $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, be a panel of time series and let $\{\tilde{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, be the observed panel. Outliers are generated either

1. by replacing single points of the data:

$$\tilde{X}_t^{(i)} = (1-\delta_{1,t,i})\, X_t^{(i)} + \delta_{1,t,i}\, V_t^{(i)}\,, \quad t \in \mathbb{Z},\, i = 1, \dots, n,$$

   where the processes $\{V_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, are independent of the $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$. We assume that they are independently and identically distributed with $V_t^{(i)} \sim \mathrm{N}(0, \sigma_V^2)$ for all $t \in \mathbb{Z}$, $i = 1, \dots, n$. For $i = 1, \dots, n$, the processes $\{\delta_{1,t,i}\}_{t\in\mathbb{Z}}$ are independent Bernoulli processes which are independent of $\{X_t^{(j)}\}_{t\in\mathbb{Z}}$ and $\{V_t^{(j)}\}_{t\in\mathbb{Z}}$ for all $j = 1, \dots, n$. They are identically distributed with $\mathbb{P}(\delta_{1,t,i} = 1) = 1 - \mathbb{P}(\delta_{1,t,i} = 0) = \varepsilon_1$ for some $\varepsilon_1 > 0$.

2. or by replacing entire time series:

$$\{\tilde{X}_t^{(i)}\}_{t\in\mathbb{Z}} = (1-\delta_{2,i})\, \{X_t^{(i)}\}_{t\in\mathbb{Z}} + \delta_{2,i}\, \{W_t^{(i)}\}_{t\in\mathbb{Z}}\,, \quad i = 1, \dots, n,$$

where the $\{W_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1,\dots,n$, are independent autoregressive processes which are independent of the $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1,\dots,n$. The random variables $\delta_{2,i}$, $i = 1,\dots,n$, are independent of the processes $\{X_t^{(j)}\}_{t\in\mathbb{Z}}$ and $\{W_t^{(j)}\}_{t\in\mathbb{Z}}$, $j = 1,\dots,n$, and fulfil $\mathbb{P}(\delta_{2,i} = 1) = 1 - \mathbb{P}(\delta_{2,i} = 0) = \varepsilon_2 > 0$.

All further investigations on contaminated data are based on this assumption.

3.3 Robustifying the ICM Parameter Estimator

In sections 2.4.2 and 2.4.3, parameter estimators have been derived for the intercorrelated time series model ICM (definition 2.2.2) and its generalisation, the GICM (definition 2.3.2). In the first case we employ an iterative procedure for estimating, whereas in the GICM the parameters can be obtained in a single step. In order to simplify the notation in the subsequent considerations, we introduce the following:

3.3.1 NOTATIONS
For ease of notation let

$$\hat{A}_1 = \sum_{t=p+1}^{T}\sum_{i=1}^{n} \mathring{X}_t^{(i)}\, \mathring{\mathbf{x}}_{t-1}^{(i)}\,, \qquad \hat{B}_1 = \sum_{t=p+1}^{T}\sum_{i=1}^{n} \mathring{\mathbf{x}}_{t-1}^{(i)}\, \mathring{\mathbf{x}}_{t-1}^{(i)\prime}\,,$$

$$\hat{A}_2 = \sum_{t=p+1}^{T} \bar{X}_t\, \bar{\mathbf{x}}_{t-1} \quad \text{and} \quad \hat{B}_2 = \sum_{t=p+1}^{T} \bar{\mathbf{x}}_{t-1}\, \bar{\mathbf{x}}_{t-1}'\,.$$

where $\mathring{\mathbf{x}}_{t-1}^{(i)} = (\mathring{X}_{t-1}^{(i)},\dots,\mathring{X}_{t-p}^{(i)})'$, $i = 1,\dots,n$, and $\bar{\mathbf{x}}_{t-1} = (\bar{X}_{t-1},\dots,\bar{X}_{t-p})'$, $t \in \mathbb{Z}$.

3.3.2 REMARK

1. In the ICM the parameter estimator $\hat{a} = (\hat{a}_1,\dots,\hat{a}_p)'$ of the autoregressive parameter a given the estimates $\hat{\sigma}^2$ and $\hat{\omega}_n^2$ of σ^2 and ω_n^2 is

$$\hat{a} = \left(\frac{1}{\hat{\sigma}^2}\,\hat{B}_1 + \frac{1}{\hat{\omega}_n^2}\,\hat{B}_2\right)^{-1} \times \left(\frac{1}{\hat{\sigma}^2}\,\hat{A}_1 + \frac{1}{\hat{\omega}_n^2}\hat{A}_2\right).$$

Conditional on $\hat{a}$, the variances σ^2 and ω_n^2 can be estimated by

$$\hat{\sigma}^2 = \frac{1}{(n-1)}\,\frac{1}{T-p}\sum_{i=1}^{n}\sum_{t=p+1}^{T}\left(\hat{a}(\mathrm{L})\mathring{X}_t^{(i)}\right)^2 \quad \text{and}$$

$$\hat{\omega}_n^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\left(\hat{a}(\mathrm{L})\bar{X}_t\right)^2.$$

Starting with $\hat{\sigma}_0^2 = \hat{\omega}_{n,0}^2 = 1$, these steps are repeated until convergence (see section 2.4.2).

In the GICM, we get for the parameters of the residual processes and the mean process, respectively, that (see section 2.4.3)

$$\hat{a} = \hat{B}_1^{-1}\,\hat{A}_1\,, \qquad \hat{\sigma}_n^2 = \frac{1}{(n-1)\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \left(\hat{a}(\mathrm{L})\mathring{X}_t^{(i)}\right)^2,$$

$$\hat{b} = \hat{B}_2^{-1}\,\hat{A}_2\,, \qquad \text{and} \qquad \hat{\omega}_n^2 = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(\hat{b}(\mathrm{L})\bar{X}_t\right)^2.$$

2. As systematic errors can only occur when a complete time series is outlying, a heuristic approach is to replace each cross sectional mean by the corresponding median. Thus $\bar{X}_t$ is substituted by $X_t^m = \mathrm{med}_{i=1,\ldots,n} X_t^{(i)}$, $\mathring{X}_t^{(i)}$ by the process $\breve{X}_t^{(i)} = X_t^{(i)} - X_t^m$, and the mean $\frac{1}{n}\sum_{i=1}^{n}\sum_{t=p+1}^{T} \mathring{X}_{t-k}^{(i)}\mathring{X}_{t-l}^{(i)}$ by the median $\mathrm{med}_{i=1,\ldots,n}\sum_{t=p+1}^{T} \breve{X}_{t-k}^{(i)}\breve{X}_{t-l}^{(i)}$. This causes problems since taking medians over $\sum_{t=p+1}^{T} \breve{X}_{t-k}^{(i)}\breve{X}_{t-l}^{(i)}$ means taking componentwise medians over matrices. Therefore the resulting matrix is not necessarily positive definite any more. So the procedure will not be numerically stable, in particular if the order p of the autoregressive process grows.

We use the following procedure for robustifying. $\hat{A}_1$, $\hat{A}_2$, and $\hat{B}_1$ are estimated as in the heuristic approach mentioned in the preceding remark. $\hat{B}_2$, however, is obtained by a recursive algorithm. The underlying idea is that the transformed matrices $\hat{B}^{(i)}$, $i = 1, \ldots, n$, are diagonally dominated such that the median is essentially taken over their eigenvalues. It turns out that in practice very few iterations (often only two) are needed until the procedure converges.

Robust estimation of $\hat{a}$

1. For $t \in \mathbb{Z}$ let $X_t^m = \mathrm{med}_{i=1,\ldots,n} X_t^{(i)}$ and $\mathbf{x}_{t-1}^m = \left(X_{t-1}^m, \ldots, X_{t-p}^m\right)'$. Replace

$$\hat{A}_2 \text{ by } \hat{A}_2^m = \sum_{t=p+1}^{T} X_t^m\,\mathbf{x}_{t-1}^m \quad \text{and} \quad \hat{B}_2 \text{ by } \hat{B}_2^m = \sum_{t=p+1}^{T} \mathbf{x}_{t-1}^m\,\mathbf{x}_{t-1}^{m\,\prime}\,.$$

2. For $i = 1, \ldots, n$, let $\breve{X}_t^{(i)} = X_t^{(i)} - X_t^m$ and $\breve{\mathbf{x}}_{t-1}^{(i)} = (\breve{X}_{t-1}^{(i)}, \ldots, \breve{X}_{t-p}^{(i)})'$.

   $\hat{A}_1$ is robustified by taking the componentwise median:

$$\breve{A}_1 = \mathrm{med}_{i=1,\ldots,n} \sum_{t=p+1}^{T} \breve{X}_t^{(i)}\,\breve{\mathbf{x}}_{t-1}^{(i)}\,.$$

3. Replace $\hat{B}_1$ by a robust covariance matrix obtained from the following algorithm:

   (a) Let $\nu = 0$ and $\Sigma_\nu = \mathrm{med}_{i=1,\ldots,n}\sum_{t=p+1}^{T} \breve{\mathbf{x}}_{t-1}^{(i)}\,\breve{\mathbf{x}}_{t-1}^{(i)\prime}$.

   (b) Let U be the orthonormal matrix consisting of the eigenvectors of Σ_ν.

(c) Transform the sample covariance matrices of each single time series $\breve{\mathbf{x}}_{t-1}^{(i)}$ separately with U:

$$\hat{B}^{(i)} = U' \Big(\sum_{t=p+1}^{T} \breve{\mathbf{x}}_{t-1}^{(i)} \, \breve{\mathbf{x}}_{t-1}^{(i)\prime} \Big) U \qquad \text{for } i = 1, \ldots, n.$$

(d) Take the componentwise median of the transformed covariance matrices:

$$\hat{B} = \text{med}_{i=1,\ldots,n} \, \hat{B}^{(i)} \, .$$

(e) Transform back: $\Sigma_{\nu+1} = U \, \hat{B} \, U'$.

(f) Iterate step (b) to (e) until convergence is attained, e.g. until

$$||\Sigma_{\nu+1} - \Sigma_{\nu}|| < \varepsilon \qquad \text{for some given } \varepsilon > 0.$$

(g) Let $\breve{B}_1 = \Sigma_{\nu+1}$.

4. Estimate the parameters using the robustified vectors and matrices.

This means that in the case of the ICM these matrices together with the robust variance estimates described below are inserted in the recursive algorithm described in section 2.4.2. If there is no contamination and the number of observations per time series T tends to infinity, each element $\frac{1}{T-p} \sum_{t=p+1}^{T} \mathring{X}_{t-k}^{(i)} \mathring{X}_{t-l}^{(i)}$ of the sample covariance matrix $\hat{B}^{(i)}$ is asymptotically normal. This was the motivation for taking the median over the robustified sample covariance matrices.

The variances could be derived as mentioned in remark 3.3.2. For estimating $\hat{\sigma}^2$, we however employ a reweighted estimator.

Robust estimation of the variances

We use the above notations X_t^m and $\breve{X}_t^{(i)}$. Given $\hat{a}$ and $\hat{b}$, the variances are estimated as follows:

1. For $\hat{\omega}_n^2$, replace $\bar{X}_t$ by X_t^m:

$$\hat{\omega}_n^2 = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(\hat{b}(\mathrm{L}) X_t^m \right)^2 .$$

   In the ICM, $p = q$ and $a_k = b_k$ for all $k = 1, \ldots, p$. Thus here we use the common estimator $\hat{a}$ instead of $\hat{b}$.

2. $\hat{\sigma}_n^2$ is estimated using a reweighting step following Rousseeuw and Leroy (1987). Denote the residuals by $r_t^{(i)} = \breve{X}_t^{(i)} - \sum_{k=1}^{p} \hat{a}_k \breve{X}_{t-k}^{(i)}$.

   First, let $\sigma_0^2 = \text{med}_{i=1,\ldots,n} \frac{1}{T-p} \sum_{t=p+1}^{T} r_t^{(i)\,2}$.

   Compute a finite sample correction factor as $s_0 = 1.4826 \, (1 + \frac{5}{n-1}) \, \sigma_0$. Here the constant is for consistency and $\frac{5}{n-1}$ is a correction term for the small sample bias (Rousseeuw and Leroy 1987, p. 44).

As cutoff value we chose $2.5 \times s_0$: For $i = 1, \ldots, n$ let $w_t^{(i)} = 1$ if $|r_t^{(i)}/s_0| \leq 2.5$, else $w_t^{(i)} = 0$. Then

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n \sum_{t=p+1}^T w_t^{(i)} r_t^{(i)\,2}}{\sum_{i=1}^n \sum_{t=p+1}^T w_t^{(i)} - p},$$

where p is the order of the autoregressive process.

We propose two variations of this procedure:

(i) Another possibility for estimating σ_n^2 is to let $\sigma_0^2 = \mathrm{med}_{\{i=1,\ldots,n;t=p+1,\ldots,T\}}\, r_t^{(i)\,2}$ in the first step, leaving the procedure otherwise unchanged. The resulting estimator is called $\hat{\theta}_{oa}$ ("overall median").

(ii) As an improvement, we repeat the weighting step with the estimated $\hat{\sigma}$ replacing s_0, i.e. for $i = 1, \ldots, n$, $t = p+1, \ldots, T$, we let $\tilde{w}_t^{(i)} = 1$ if $|r_t^{(i)}/\hat{\sigma}| \leq 2.5$, else we let $\tilde{w}_t^{(i)} = 0$. For $t = 1, \ldots, p$, we let furthermore $\tilde{w}_t^{(i)} = 1$ for all $i = 1, \ldots, n$. From these weights we determine the weight w_i of the individual time series i, $i = 1, \ldots, n$. If $\sum_{t=1}^T \tilde{w}_t^{(i)} \leq c\,T$, where c is a preliminarily chosen constant, we let $\mathrm{w}_i = 0$, otherwise $\mathrm{w}_i = 1$. These weights can be used on the original data. Then we perform a second, non-robust estimation on the remaining time series. This method is the panel analogue to Rousseeuw's reweighted least squares estimator (Rousseeuw and Leroy 1987). We call the estimator $\hat{\theta}_{rw}$.

(iii) A modification of the last procedure also allowing for arbitrary outliers is to form weights $\tilde{\mathrm{w}}_t^{(i)} = \min\{\mathrm{w}_i, w_t^{(i)}\}$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$. Then we exclude all time points $X_t^{(i)}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$, having weight $\tilde{\mathrm{w}}_t^{(i)} = 0$ from the original panel before performing the second, non-robust estimation. The corresponding estimator is called $\hat{\theta}_{rw_2}$.

3.3.3 REMARK

Note that the above described procedure is a compromise between averaging over single entries of a matrix and estimating one common robust matrix. It enables us to exploit the characteristics of the ICM even in the contaminated case. There are no standard methods for replacing the componentwise median of the matrices. Usually robust covariance matrix estimators are used instead. For taking the median over vectors, however, componentwise medians are commonly used. There exist a variety of more sophisticated methods such as Oja's median (Oja 1983), but in practice the componentwise median often performs well. This also is the case in the above parameter estimation procedures, where the main point was deriving the robust covariance matrix.

Simulation results for the above described estimators can be found in section B.1 of the Appendix B and are discussed in section B.4.

Bootstrap corrections

In parameter estimation, where the true values are not known, bootstrap can be used for deriving the empirical bias. Thus we apply bootstrap procedures in order to compensate for the bias. For comparison, we have implemented three versions: first the parameters were estimated with a residual bootstrap. In a second step this has been adapted to the panel situation. As the innovations are normal by assumption 2.2.1, we also generated samples of a simpler structure where the data was obtained from normally distributed errors with variances $\hat{\sigma}_n^2$ and $\hat{\omega}_n^2$. The procedure is based on a residual bootstrap for autoregressions, whose properties are known (Kreiss 1997).

(i) In the ICM let $\hat{\theta} = \left(\hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}^2, \hat{\tau}^2\right)'$, where $\hat{\tau}^2$ is given by $\hat{\tau} = \hat{\sigma}^2 - \hat{\omega^2}_n$. Denote the residuals obtained from the $\breve{X}_t^{(i)}$ by $\breve{r}_t^{(i)}$, $t = p+1, \ldots, T$, $i = 1, \ldots, n$, and the residuals obtained from the X_t^m by r_t^m, $t = p+1, \ldots, T$ (notations as above). The bootstrap data then is generated by

$$X_t^{*\,(i)} = \sum_{k=1}^{p} \hat{a}_k X_{t-k}^{*\,(i)} + \breve{r}_t^{(i)} + r_t^m\,, \quad t = p+1, \ldots, T,\ i = 1, \ldots, n\,,$$

where $X_1^{*\,(i)} = \cdots = X_p^{*\,(i)} = 0$ for $i = 1, \ldots, n$. The residuals are sampled with replacement. In order to get stationary time series, $T + 500$ points are simulated for each time series. For the estimation, we disregard the first 500 points.

In the GICM, the samples are generated analogously.

Comparing the estimates $\hat{\theta}$ (e.g. $\hat{\theta}_{rob}$) from the original procedure and $\hat{\theta}_{bs}$ derived from a bootstrap procedure based on $\hat{\theta}$ and the corresponding residuals, we obtain the empirical bias $\hat{\theta} - \hat{\theta}_{bs}$ of the robust estimator. This then is used as approximation for the true bias $\mathbb{E}\,\hat{\theta} - \theta_0$, where θ_0 denotes the true parameter of the model. In order to adjust for the size of the underlying parameter, we multiply each component $\hat{\theta}_k$ of $\hat{\theta}$ by a factor derived from the relative size of the bias. Thus we get $\hat{\theta}_{k,2}$ by $\hat{\theta}_{k,2} = \hat{\theta}_k / \left(\hat{\theta}_{bs;k}/\hat{\theta}_k\right) = \hat{\theta}_k^2/\hat{\theta}_{bs;k}$.

(ii) The above procedure does not take the correlation structure of the panel into account. For example in the ICM, the residuals are not independent, but correlated with $\operatorname{cov}\left(\breve{r}_t^{(i)}, \breve{r}_t^{(j)}\right) = -\frac{1}{n}$ for $i \neq j$. Thus the method can be modified such that the residuals $\breve{r}_t^{(i)}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$, are sampled from the set of vectors $\left(\breve{r}_t^{(1)}, \ldots, \breve{r}_t^{(n)}\right)'$, $t = 1, \ldots, T$. All further steps are performed as in the preceding method.

(iii) For the third, simplified, procedure we sample from independent normally distributed innovations: $\varepsilon_t^{(i)} \sim \mathrm{N}(0, \hat{\sigma}^2)$, $t = 1, \ldots, 500 + T$, $i = 1, \ldots, n$, and $\eta_t \sim \mathrm{N}(0, \hat{\omega}_n^2)$, $t = 1, \ldots, 500 + T$.

Let

$$X_t^{*\,(i)} = \sum_{k=1}^{p} \hat{a}_k X_{t-k}^{*\,(i)} + \varepsilon_t^{(i)} + \eta_t\,, \quad t = p+1, \ldots, T,\ i = 1, \ldots, n\,,$$

where $X_1^{*\,(i)} = \cdots = X_p^{*\,(i)} = 0$ for $i = 1, \ldots, n$. Moreover we simulate again $T + 500$ points per time series and disregard the first 500.

In the GICM, the generation of the sample is analogous.

As before, the empirical bias is used for correcting the estimates.

Simulations show that the latter estimator, where we use the information on the distribution of the innovations, performs better than the other two. However these procedures, being non-robust in nature, cannot completely compensate for the bias in the presence of outliers. For details, see the Appendix B, subsection B.1.1, and the discussion in section B.4.

3.4 The Robust Panel Autocovariance Estimator

As mentioned in the introduction, a second possibility is to derive robust parameter estimators from the identity $\mathrm{cov}(X,Y) = \frac{1}{4}\big(\mathrm{var}(X+Y) - \mathrm{var}(X-Y)\big)$. We here focus on one robust autocovariance estimator that can easily be adapted to the panel situation. It is based on the robust scale estimator Q_n which has been suggested by Rousseeuw and Croux (1993). Ma and Genton (2000) have generalised Q_n to the time series case and have used it for deriving a robust autocovariance estimator as follows.

3.4.1 DEFINITION (MA AND GENTON (2000))
Let $X = (X_1, \ldots, X_n)'$ be an observation from a stationary time series. Define the robust scale estimator Q_n as the following kth order statistic:

$$Q_n(X) = c \times \{|X_t - X_s|, 1 \leq s < t \leq n\}_{(k)} .$$

Here $c = 2.219$ is a factor for consistency and $k = \left\lfloor \frac{\binom{n}{2}+2}{4} \right\rfloor + 1$.

This means that we sort the set of all $\binom{n}{2}$ inter-point distances in increasing order and then compute its kth order statistic, which is approximately the 1/4-quantile for large n. This scale estimator can be used to define a robust autocovariance estimator.

3.4.2 DEFINITION (MA AND GENTON (2000))
Let $\mathbf{x} = (X_1, \ldots, X_n)$ be a sample from a stationary time series.
For $h \in \{1, \ldots, n-1\}$ let $\mathbf{u_h} = (X_1, \ldots, X_{n-h})$, and $\mathbf{v_h} = (X_{h+1}, \ldots, X_n)$.
Then the robust autocovariance estimator $\hat{\gamma}_{Q,n}$ is defined by

$$\hat{\gamma}_{Q,n}(h, \mathbf{x}) = \frac{1}{4}\Big(Q^2_{n-h}(\mathbf{u_h} + \mathbf{v_h}) - Q^2_{n-h}(\mathbf{u_h} - \mathbf{v_h})\Big) .$$

3.4.3 REMARK
We now summarise the main properties of these estimators.

1. Rousseeuw and Croux (1993) have investigated Q_n for independently and identically distributed data. They have shown that in the Gaussian case the estimator is Fisher-consistent, i.e. that $\mathbb{E}(Q_n) = \sigma$ if $X_t \sim \mathrm{N}(0, \sigma^2)$. It has a smooth influence function and the efficiency at Gaussian distributions is 82.27%. Also, it achieves the maximal possible asymptotic breakdown point of 50%. The efficiency can be improved up to 91% at a trade-off for a lower breakdown point. Furthermore Q_n is asymptotically normal. This follows from a result of Serfling (1984) since Q_n is a special case of Serfling's generalised L-statistics. For their computations, Rousseeuw and Croux (1993) introduce an empirical correction

factor of $n/(n+1.4)$ for n even. We use this correction factor in our simulations, too. Q_n is still consistent and asymptotically normal when the dependence in the process is not too strong, e.g. if the process is α-mixing (Ma and Genton 2000).

2. The robust autocovariance estimator $\hat{\gamma}_{Q,n}$ is consistent since Q_n is consistent. Moreover it is asymptotically normal. The asymptotic variance can be derived from the influence function $IF(\gamma, Q, F)$ of $\gamma_{Q,n}$, where F is the distribution function of the process $\{X_t\}_{t\in\mathbb{Z}}$. However, numerical integration is necessary for the computation. For details, see Ma and Genton (2000).

3. Note that Q_n does not rely on any location knowledge, it is therefore said to be location-free. Thus also the robust covariance estimators based on Q_n are location-free.

The above estimators can easily be generalised to the panel case.

3.4.4 DEFINITION
Let $T \geq 2$ *and* $\mathbf{x} = \{X_t^{(i)};\, t = 1, \ldots, T, i = 1, \ldots, n\}$ *be a panel of time series. The* panel scale estimator $Q_{n,T}$ *is defined as the* k*th order statistic*

$$Q_{n,T} = c \times \left\{ |X_s^{(i)} - X_t^{(i)}|; 1 \leq s < t \leq T, i = 1, \ldots, n \right\}_{(k)},$$

where $c = 2.219$ *is a factor for consistency as in Rousseeuw and Croux (1993) and* $k = \left\lfloor \frac{n\binom{T}{2}+2}{4} \right\rfloor + 1$.
For $h \in \{1, \ldots, T-1\}$ *let* $\mathbf{u} = (X_1^{(1)}, \ldots, X_{T-h}^{(1)}, \ldots, X_1^{(n)}, \ldots, X_{T-h}^{(n)})$*, and analogously* $\mathbf{v} = (X_{h+1}^{(1)}, \ldots, X_T^{(1)}, \ldots, X_{h+1}^{(n)}, \ldots, X_T^{(n)})$*. Then the autocovariance estimator obtained from* $Q_{n,T}$ *is*

$$\hat{\gamma}_{n,T}(h, \mathbf{x}) = \frac{1}{4}\Big(Q^2_{n,T-h}(\mathbf{u_h} + \mathbf{v_h}) - Q^2_{n,T-h}(\mathbf{u_h} - \mathbf{v_h}) \Big).$$

The order k is chosen to guarantee a fast convergence to the 1/4-quantile.

3.4.5 REMARK

1. Note that the order k of the statistic $Q_{n,T}$ fulfils $k \approx n\binom{T}{2}/4$. To be more specific, elementary calculations show that

$$\frac{k}{n\binom{T}{2}} = \frac{1}{4} + O\left(\frac{1}{n\,T^2}\right).$$

2. The correlation in an autoregressive time series decreases exponentially. If we exclude the differences of time points X_s, X_t with $|s-t|$ small in the computation of $Q_{n,T}$, we thus can eliminate effects due to a high correlation. The corresponding panel scale estimator for a panel $\mathbf{x} = \{X_t^{(i)};\, t = 1, \ldots, T, i = 1, \ldots, n\}$ of time series then is

$$Q^d_{n,T} = c \times \left\{ |X_s^{(i)} - X_t^{(i)}|; 1 \leq s < t \leq T, |s-t| > d, i = 1, \ldots, n \right\}_{(k)},$$

where $c = 2.219$ and $k = \left\lfloor \frac{n\binom{T-d}{2}+2}{4} \right\rfloor + 1$. The scale estimator $Q^d_{n,T}$ can be used for constructing a modified panel autocovariance estimator $\hat{\gamma}^d_{n,T}$ as in the preceding definition. We investigate both $\hat{\gamma}_{n,T}$ and $\hat{\gamma}^d_{n,T}$ in the simulations (section B.2.1 in the Appendix B). There we choose $d = 0.1\,T$.

In the panel case, two kinds of breakdown points can be defined according to the type of possible outliers. We use Huber's view of the breakdown point as the maximal fraction of outliers the estimator can cope with (Huber 1981).

3.4.6 DEFINITION
Let $\mathbf{x}$ *be a sample from a time series panel as above and suppose that outliers are generated as in assumption 3.2.1.*
If $\tilde{\mathbf{x}}$ *is derived from* $\mathbf{x}$ *by replacing* m *entire time series, the sample breakdown point of a scale estimator* $S_{n,T}(\mathbf{x})$ *is*

$$\epsilon^\star(S_{n,T}(\mathbf{x})) = \max\left\{\frac{m}{n} : \sup_{\tilde{\mathbf{x}}}(S_{n,T}(\tilde{\mathbf{x}})) < \infty \text{ and } \inf_{\tilde{\mathbf{x}}}(S_{n,T}(\tilde{\mathbf{x}})) > 0\right\}.$$

In the case of arbitrary outliers, where $\tilde{\mathbf{x}}$ *is generated by replacing* m *observations of* $\mathbf{x}$ *by arbitrary values, the sample breakdown point is given by*

$$\epsilon^\circ(S_{n,T}(\mathbf{x})) = \max\left\{\frac{m}{n\,T} : \sup_{\tilde{\mathbf{x}}}(S_{n,T}(\tilde{\mathbf{x}})) < \infty \text{ and } \inf_{\tilde{\mathbf{x}}}(S_{n,T}(\tilde{\mathbf{x}})) > 0\right\}.$$

Thus we can derive the sample breakdown points of $Q_{n,T}$.

3.4.7 LEMMA
With the above notations, we obtain for the panel scale estimator
if complete time series are outlying

$$\epsilon^\star(Q_{n,T}(\mathbf{x})) = \left\lfloor \frac{n}{4} \right\rfloor / n$$

and for arbitrary outliers if $T > 2$

$$\frac{1}{n\,T}\min\left\{\left\lfloor \frac{n}{4} \right\rfloor \times (T-1)\right\} \le \epsilon^\circ(Q_{n,T}(\mathbf{x})) \le \frac{1}{n\,T}\min\left\{\left(\left\lfloor \frac{n}{4} \right\rfloor + \frac{1}{2}\right) \times (T-1)\right\},$$

and if $T = 2$

$$\epsilon^\circ(Q_{n,T}(\mathbf{x})) = \frac{\left\lfloor \frac{n}{2} \right\rfloor + 2}{n\,T}.$$

PROOF:
The breakdown point is the maximal proportion of observations that can be changed with $\sup_{\tilde{\mathbf{x}}}(S_{n,T}(\tilde{\mathbf{x}})) < \infty$ and $\inf_{\tilde{\mathbf{x}}}(S_{n,T}(\tilde{\mathbf{x}})) > 0$. Since $Q_{n,T}$ is a kth order statistic, $S_{n,T}(\tilde{\mathbf{x}}) = 0$ ("implosion") if k differences are zero. $S_{n,T}(\tilde{\mathbf{x}}) = \infty$ ("explosion") occurs if $n\binom{T}{2} - k + 1$ differences are allowed to become arbitrarily large. k was defined as $k = \left\lfloor \frac{n\binom{T}{2}+2}{4} \right\rfloor + 1$ (see definition 3.4.4).
Thus the first case of entire time series as outliers is clear. In the second case, we have to investigate what the highest impact of m outliers can be. The largest number

of differences vanishes (implosion) if all outliers have the same value and occur in the same time series i. Write $m = a\,(T-1) + x$ with $0 \le a \le n$ and $0 \le x \le (T-2)$. Then m outliers can influence up to $k_m = a\binom{T}{2} + \frac{x\,(x+1)}{2}$ differences. For $T > 2$ the condition $k_m < k$, which implies that $\inf_{\tilde{\mathbf{x}}}(S_{n,T}(\tilde{\mathbf{x}})) > 0$ is preserved, is equivalent to $m \le \left\lfloor \frac{n}{4} \right\rfloor \times (T-1) + x_{max}$.
Here x_{max} is the maximal x such that the inequality

$$\frac{x\,(x+1)}{2} \le \frac{\tilde{n}}{4}\binom{T}{2} - \frac{\tilde{x}}{4} + \left\lfloor \frac{\tilde{x}+2}{4} \right\rfloor,$$

where $\tilde{n} \equiv n \mod (4)$ and $\tilde{x} \equiv n\binom{T}{2} \mod (4)$, holds. For all $T > 2$, it fulfils $0 \le x_{max} < (T-1)/2$. Thus $\left\lfloor \frac{n}{4} \right\rfloor \times (T-1) \le \epsilon^\circ(Q_{n,T}(\mathbf{x})) < \left\lfloor \frac{n}{4} \right\rfloor \times (T-1) + \frac{T-1}{2}$.
In the case of an explosion, the highest impact is reached if each time series contains the same number of outliers. If $m = a\,n + x$ with $0 \le a \le (T-1)$ and $0 \le x < n$, the maximal number of changed differences is $k_m = n\,a\,T - n\,\frac{a\,(a+1)}{2} + x\,T - (a+1)\,x$. As the outliers here induce very large differences, up to $\binom{T}{2} - k_m$ differences can be affected without the estimator breaking down. From the inequality $k \le \binom{T}{2} - k_m$ we obtain for $T > 2$ by distinguishing several cases

- if T is odd, $m \le \left\lfloor \frac{T-1}{2} \right\rfloor n + \left\lfloor \frac{n-1}{4} \right\rfloor$
- if T is even, $m \le \left\lfloor \frac{T-1}{2} \right\rfloor n + \left\lfloor \frac{3n-1}{4} \right\rfloor$
- exceptions have to be made if $T = 3$ and $\tilde{n} = 1$ or 2, or if $T = 5$ and $\tilde{n} = 1$. Then $m \le \left\lfloor \frac{T-1}{2} \right\rfloor n + \left\lfloor \frac{n-1}{4} \right\rfloor - 1$.

All these m fulfil $m \le \epsilon^\circ(Q_{n,T}(\mathbf{x})) < \left\lfloor \frac{T+1}{2} \right\rfloor \times n$.
Since for all $T > 2$ we have that $\left\lfloor \frac{n}{4} \right\rfloor \times (T-1)$ is smaller than the right hand sides of the above inequalities and $\left(\left\lfloor \frac{n}{4} \right\rfloor + \frac{1}{2}\right) \times (T-1) \le \left\lfloor \frac{T+1}{2} \right\rfloor \times n$, we get the stated result.
If $T = 2$, the worst cases for implosion and explosion coincide. It is easy to see that here $m \le \left\lfloor \frac{n+2}{2} \right\rfloor$. □

In analogy to the above breakdown points one could calculate breakdown points for the autocovariance estimator $\hat{\gamma}_{n,T}$ derived from $Q_{n,T}$ by adding the number of components of $\mathbf{u_h} + \mathbf{v_h}$ and $\mathbf{u_h} - \mathbf{v_h}$ which can be replaced without causing $\hat{\gamma}_{n,T}$ to explode or to implode. But this is not consistent with the character of $\hat{\gamma}_{n,T}$ as autocovariance estimator, since $\mathbf{u_h}$ and $\mathbf{v_h}$ are both derived from the same sample $\mathbf{x}$. This led Ma and Genton (2000) to define a breakdown point depending on the size of the original sample and on the lag h.

3.4.8 DEFINITION (MA AND GENTON (2000))
Let $\mathbf{x} = (X_1, \ldots, X_n)$, be a sample from a stationary time series and $\tilde{\mathbf{x}}$ a contaminated version. Furthermore let $\tilde{\mathbf{u}}_\mathbf{h}$ and $\tilde{\mathbf{v}}_\mathbf{h}$ be vectors as derived in definition 3.4.2.

The temporal breakdown point of a autocovariance estimator $\hat{\gamma}(h, \mathbf{x})$ *derived from a scale estimator* S_T *is*

$$\begin{aligned}\epsilon^t(\hat{\gamma}(h,\mathbf{x})) = \max\Big\{\frac{m}{T} :& \sup_{I_m}\sup_{\tilde{\mathbf{x}}}(S_{T-h}(\tilde{\mathbf{u}}_\mathbf{h}+\tilde{\mathbf{v}}_\mathbf{h})) < \infty \quad \textit{and} \\ & \inf_{I_m}\inf_{\tilde{\mathbf{x}}}(S_{T-h}(\tilde{\mathbf{u}}_\mathbf{h}+\tilde{\mathbf{v}}_\mathbf{h})) > 0 \quad \textit{and} \\ & \sup_{I_m}\sup_{\tilde{\mathbf{x}}}(S_{T-h}(\tilde{\mathbf{u}}_\mathbf{h}-\tilde{\mathbf{v}}_\mathbf{h})) < \infty \quad \textit{and} \\ & \inf_{I_m}\inf_{\tilde{\mathbf{x}}}(S_{T-h}(\tilde{\mathbf{u}}_\mathbf{h}-\tilde{\mathbf{v}}_\mathbf{h})) > 0\Big\},\end{aligned}$$

where the first supremum is over all sets I_m *of* m *points which are to be replaced in* $\tilde{\mathbf{x}}$*. This definition is valid for the panel autocovariance estimator* $\hat{\gamma}_{n,T}$ *if* $\mathbf{x}$*,* $\tilde{\mathbf{x}}$*,* $\tilde{\mathbf{u}}_\mathbf{h}$ *and* $\tilde{\mathbf{v}}_\mathbf{h}$ *are as in definition 3.4.4 and if* $\frac{m}{T}$ *is substituted by* $\frac{m}{nT}$*.*

3.4.9 REMARK

The temporal breakdown point of $\gamma_Q(h, \mathbf{x})$ is optimal: for each lag h and each T Ma and Genton (2000) have derived the maximal number $\nu_{max}(h, m, T)$ of differences affected by m outliers. Asymptotically, the temporal breakdown point of $\hat{\gamma}_Q(h, \mathbf{x})$ is $\epsilon^t(\gamma_Q(h, \mathbf{x})) = 25\%$, which is the maximal possible value. The procedure is analogous in the panel case though the calculations become more cumbersome.

The robust autocovariance estimator $\hat{\gamma}_Q$ is very robust against outliers, especially against arbitrary outliers (Ma and Genton (2000)). Thus also $\hat{\gamma}_{n,T}$ is robust against randomly distributed contamination, which is not the case for the previously discussed estimator. We use $\hat{\gamma}_{n,T}$ for estimating the covariance matrix and the autocovariance vector robustly. The method is described in the next section. Although the autocovariance estimators are moderately biased, the corresponding parameter estimates are satisfyingly close to the true value. The simulations can be found in section B.2.1 of the Appendix B, see also the discussion in section B.4.

3.5 Parameter Estimation via Robust Autocovariances

There are several possibilities for deriving parameter estimators which are robustified versions of the least squares or Yule-Walker equations $\theta = \hat{\Gamma}^{-1}\hat{\gamma}$. Each element of the covariance matrix $\hat{\Gamma}$ and the autocovariance vector $\hat{\gamma}$ can be estimated separately or the complete matrix can be replaced by a robust estimator.
We compare two procedures. First, we estimate each entry of $\hat{\Gamma}$ and $\hat{\gamma}$ separately by the robust panel autocovariance estimator $\hat{\gamma}_{n,T}$ of the last section. As alternative we employ the minimum covariance determinant (MCD) method for estimating the covariance matrix robustly.

Robust panel autocovariance estimator

The robust panel autocovariance estimator $\hat{\gamma}_{n,T}$ derived from the scale estimator Q_n has been defined in definition 3.4.4. It can be employed for robustifying the parameter estimation in the ICM and GICM as follows.

Let $T \geq 2$ and $\mathbf{x} = \{X_t^{(i)};\, t = 1, \ldots, T, i = 1, \ldots, n\}$. Using the notations of section 3.3, denote the panel of robust residuals by $\breve{\mathbf{x}} = \{\breve{X}_t^{(i)};\, t = 1, \ldots, T, i = 1, \ldots, n\}$ and the median vector by $\mathbf{x}^m = \{X_t^m\}_{t=1,\ldots,T}$. For $h = 0, \ldots, p$, compute

$$g_r(h) = \hat{\gamma}_{n,T}(h, \breve{\mathbf{x}}) \quad \text{and} \quad g_m(h) = \hat{\gamma}_{n,T}(h, \mathbf{x}^m)\,.$$

From these we obtain robust versions of the sample covariance matrices $\hat{B}_1$ and $\hat{B}_2$ and the sample autocovariance vectors $\hat{A}_1$ and $\hat{A}_2$ which are used in the parameter estimation (see remark 3.3.1): let

$$\begin{aligned} \hat{A}_{1,Q} &= \big(g_r(1), \ldots, g_r(p)\big)', & \hat{B}_{1,Q} &= \left(g_r(i-j)\right)_{i,j=1,\ldots,p}\,, \\ \hat{A}_{2,Q} &= \big(g_m(1), \ldots, g_m(p)\big)' \quad \text{and} & \hat{B}_{2,Q} &= \left(g_m(i-j)\right)_{i,j=1,\ldots,p}\,. \end{aligned}$$

As the method is based on componentwise robustification, the obtained matrices are not necessarily positive definite.

Minimum covariance determinant

For comparison, we derive robust versions of $\hat{A}_1$, $\hat{B}_1$, $\hat{A}_2$ and $\hat{B}_2$ directly from robust covariance matrices. Let $\breve{\mathbf{x}}$ and $\mathbf{x}^m$ be as above. For both we estimate a robust covariance matrix separately in the following way: if the order of the underlying process is p, we split each time series in the panel $\breve{\mathbf{x}}$ or $\mathbf{x}^m$ itself into consecutive blocks of length $p+1$, thus obtaining phase space representations of the time series. Such a set of $(p+1)$-dimensional vectors then can be used for estimating a robust $(p+1)$-dimensional covariance matrix $\hat{\Gamma}$. If its entries are denoted by $g_{i,j} = \hat{\gamma}(i-j)$, $i, j = 1, \ldots, p+1$, we get a robust p-dimensional covariance matrix by $\hat{\Gamma}_p = (g_{i,j})_{i,j=1,\ldots,p}$. A robust autocovariance vector can be obtained from its first column as $\hat{\gamma}_p = (g_{2,1}, \ldots, g_{p+1,1})'$. In this way we estimate robust versions of $\hat{B}_1$ and $\hat{A}_1$ from the phase space representation of the $\breve{\mathbf{x}}$, and of $\hat{B}_2$ and $\hat{A}_2$ from the phase space representation of the $\mathbf{x}^m$. These matrices are then used instead of $\hat{A}_1$, $\hat{B}_1$, $\hat{A}_2$ and $\hat{B}_2$ in the parameter estimation (remark 3.3.2). Note that in the GICM, the matrices derived from $\breve{\mathbf{x}}$ and from $\mathbf{x}^m$ can have different dimensions p and q as the orders of the autoregressive processes may differ.

We have decided to consider the covariance estimator obtained from the minimum covariance determinant (MCD) method, as this is reported to be more stable than the minimum volume ellipsoid and also is more efficient in high dimensions (Croux and Haesbroeck 1999). Both estimators have an asymptotic breakdown point of 50%. They are described in Rousseeuw and Leroy (1987) and have been implemented in R. For the MCD, the fast algorithm suggested in Rousseeuw and Driessen (1999) is employed. In principle any robust estimator of multivariate scatter such as the generalised M-estimators proposed by Maronna (1974), Tyler's estimator (Tyler 1987) or a method based on projections (Maronna et al. 1992) could be used instead. However, the computation of these estimators is more complicated and they have not been included in R yet.

Simulation results comparing the estimators $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$ described above can be found in section B.2 of the Appendix B. They are also discussed in section B.4.

3.6 Robust Regression

In order to evaluate the proposed methods, we compare them to two procedures derived from standard robust regression methods. The first one is based on an M-estimator as proposed by Huber (1981), the second one on the least trimmed squares method (Rousseeuw and Leroy 1987). Due to the nature of the estimators, which have been designed for regression problems, we cannot mimic the procedure of the ICM estimation any more. Either we have to perform an estimation on the original data, which is then not robust against outlying time series, or we have to content ourselves with estimation as in the GICM. To be more specific, we again transform the data in order to obtain the robust residual processes $\{\breve{X}_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$ and $\{X_t^m\}_{t\in\mathbb{Z}}$ (notations as in section 3.3), which then are used for the estimation. This procedure yields estimators $\hat{\theta}_a = (\hat{a}', \hat{\sigma}_n^2)'$ and $\hat{\theta}_b = (\hat{b}', \hat{\omega}_n^2)'$, respectively. The two methods proposed subsequently have been discussed in Rousseeuw and Leroy (1987) as robust estimators for time series analysis.

M-estimation

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a set of observations. Any pair of statistics (T_n, S_n) determined by two equations of the form

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - T_n}{S_n}\right) = 0 \quad \text{and} \quad \sum_{i=1}^{n} \chi\left(\frac{x_i - T_n}{S_n}\right) = 0$$

is called simultaneous M-estimate of location and scale. In most cases, ψ will be an odd and χ an even function. A popular choice is Huber's proposal 2, i.e.

$$\begin{aligned} & \psi(x) = \max\big(-k, \min(k, x)\big) \\ \text{and} \quad & \chi(x) = \psi(x)^2 - \beta(k) \quad \text{with} \quad \beta(k) = \int \psi(x)^2 \, \Phi(dx)\,, \end{aligned}$$

where Φ is the distribution function of the standard normal distribution (Huber 1981). For the estimation we use the iterated re-weighted least squares procedure implemented in R. There k is chosen to be $k = 1.345$. The estimators based on this method are called $\hat{\theta}_{M;a}$ and $\hat{\theta}_{M;b}$ (GICM procedure) and $\hat{\theta}_{M;dir}$ (direct procedure using the original data).

Least Trimmed Squares

Least squares estimates are obtained by minimising the sum of squared residuals. Many robust estimators, e.g. M-estimators, are defined by replacing the square by another function of the residuals. Rousseeuw's approach is however to replace the sum in the least squares approach by a more robust function. If the sum is exchanged with the median, this leads to the least median of squares (LMS) method first described in Rousseeuw (1984). Another possibility is to omit the largest residuals in the estimation. The least trimmed squares estimator (LTS) is given by minimising $\sum_{i=1}^{h} (r^2)_{i:n}$ with $h < n$, where $(r^2)_{1:n}, \dots, (r^2)_{n:n}$ are the ordered squared residuals (Rousseeuw and Leroy 1987). The LMS estimation is of low efficiency, the estimator converges at the rate $n^{1/3}$, whereas the LTS converges at the rate of $\sqrt{n}$. Furthermore, the computational difficulties, which led Rousseeuw and Leroy to recommend using LMS as preliminary estimator in a reweighted least squares procedure instead of employing the LTS, are overcome. Both estimators have the same maximal breakdown value which is attained e.g. if h is chosen to be $h = \lfloor \frac{n}{2} \rfloor + \lfloor \frac{p+1}{2} \rfloor$. In this case the asymptotic breakdown point is 50%. We use this choice of the LTS in the simulations. The resulting estimators are denoted by $\hat{\theta}_{LTS;a}$ and $\hat{\theta}_{LTS;b}$ (GICM procedure) and $\hat{\theta}_{LTS;dir}$ (direct procedure).

The simulation results for these estimators are shown section B.2.3 of the Appendix B and discussed in section B.4.

3.7 Outlier Detection

In this section we discuss two more methods for robust estimation in the panel model. They both are concerned with outlier identification. After eliminating the outliers, the parameters can be estimated with a non-robust method.
The first procedure is the heuristic approach of first identifying outliers by a (non-robust) likelihood ratio test. We include this as a comparison. Since non-robust methods can be substantially influenced by outliers (Rousseeuw and Leroy 1987, Becker and Gather 1999), we want to investigate whether it is still possible to estimate the panel parameters in this way.
The second method proposed is a new method of outlier identification. The idea is to represent the time series in the phase space, as it is done in Gather, Bauer and Fried (2002). By computing Mahalanobis distances for each time point, they were able to discover outliers inside a single time series. We generalise this concept to panels of independent time series and investigate its reliability.

3.7.1 Likelihood Ratio Test

A likelihood ratio test for homogeneity has been proposed by Basawa et al. (1984). Their setting can be specialised to a panel of independent autoregressive processes.

3.7.1 ASSUMPTION
Let $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, be a panel of independent autoregressive time series such that

$$X_t^{(i)} = a^{(i)\prime} x_{t-1}^{(i)} + \varepsilon_t^{(i)} \qquad \text{for all } t \in \mathbb{Z},$$

with $\varepsilon_t^{(i)}$ independently and identically distributed as $\varepsilon_t^{(i)} \sim \mathrm{N}(0, \sigma^2)$ for all $t \in \mathbb{Z}$, $i = 1, \dots, n$. Here $a^{(i)} = (a_1^{(i)}, \dots, a_p^{(i)})'$ and $\mathbf{x}_{t-1}^{(i)} = \big(X_{t-1}^{(i)}, \dots, X_{t-p}^{(i)}\big)'$.
Then, conditional on the initial observations $X_1^{(i)}, \dots, X_p^{(i)}$, $i = 1, \dots, n$, the corresponding likelihood function is

$$L_{n,T}(\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{t=1}^{T} \big(X_t^{(i)} - a^{(i)\prime}\mathbf{x}_{t-1}^{(i)}\big)^2 \Big).$$

The vector of unknown parameters $\theta = (a^{(1)\prime}, \dots, a^{(n)\prime}, \sigma^2)'$ can be partitioned into $\theta_1 = (a^{(1)\prime}, \sigma^2)'$ and $\theta_2 = (\alpha^{(2)\prime}, \dots, \alpha^{(n)\prime})'$, where $\alpha^{(i)} = a^{(i)} - a^{(1)}$ for $i = 2, \dots, n$.

We consider testing the composite hypothesis $H : \theta_2 = 0$ against the sequence of local alternatives $K_n : \theta_2 = \theta_{2,n}$, where $\theta_{2,n} = \frac{1}{\sqrt{n}} h$ with h an $(n-1)p$-dimensional vector of fixed real numbers. The likelihood ratio statistic then is

$$Q_{LR} = -2 \log \left(\frac{L_{n,T}(\hat{\theta}_H)}{L_{n,T}(\hat{\theta})} \right) ,$$

where $\hat{\theta}$ and $\hat{\theta}_H$ are the maximum likelihood estimators of θ in the unrestricted case and restricted by H.

Under these assumptions, the above panel belongs to a locally asymptotically normal family. Thus it is possible to derive the asymptotic distribution of Q_{LR}.

3.7.2 THEOREM (BASAWA ET AL. (1984))
Let the panel of independent autoregressive time series $\mathbf{x}$*, the hypotheses* H*,* K_n *and the likelihood ratio statistic* Q_{LR} *be as in the above assumption.*
Then the limit distribution of Q_{LR} *under* H *is* $\chi^2((n-1)p)$*, and under* K_n *it is noncentral* $\chi^2((n-1)p, \delta^2)$*. The noncentrality parameter* δ^2 *can be derived from the asymptotic covariance matrix of* Q_{LR} *under* H *and* h*, where* h *is the vector which defines the local alternative* K_n*.*

PROOF:
See Basawa et al. (1984). The authors show that the model satisfies the conditions for local asymptotic normality given in Basawa and Koul (1979). There it has been proven that under these conditions the limit distribution of Q_{LR} is as stated in the theorem. □

We apply this test in order to identify outliers in a panel of time series. Since the above result is based on independent time series, we suppose the following in this section:

3.7.3 ASSUMPTION
Let $\{X_t^{(i)}\}_{t \in \mathbb{Z}}$, $i = 1, \ldots, n$, be a panel of independent time series such that for each i

$$X_t^{(i)} = \sum_{k=1}^{p} a_k X_{t-k}^{(i)} + \varepsilon_t^{(i)}$$

with independently and identically distributed innovations $\varepsilon_t^{(i)} \sim \mathrm{N}(0, \sigma^2)$. Outliers are generated by replacing entire time series by independent time series following a different model.

For estimating the parameters, we employ an iterative procedure.

3.7.4 ALGORITHM
First, we test for homogeneity using Q_{LR}. If the hypothesis is not rejected, the parameters are estimated from the conditional log-likelihood given above. Otherwise, the time series with the smallest p-value is deleted from the sample. This is iterated until the hypothesis is not rejected any more or until a certain proportion of the time series has been classified as outlying. For the simulations we assume that not more than 20% of the time series have been replaced.

3.7.5 REMARK

As we use a multiple testing procedure, the significance level α has to be adjusted in advance for obtaining a specified significance for the test. We use the adjustment $\alpha_n = 1 - (1-\alpha)^{\frac{1}{u+1}}$, which is equivalent to $\alpha = 1 - (1-\alpha_n)^{u+1}$. Here α is the significance level we want to achieve and u is the assumed maximal proportion of outliers. Thus we have to fix the expected number of replaced time series beforehand. In the simulations we have set $u = \lfloor n/5 \rfloor$.

The results of the simulations are displayed in section B.3.1 of the Appendix B. We also have applied the method to an intercorrelated panel with entire time series as outliers. In both cases we can observe a massive masking effect if outliers are present. The test statistic rejects the null hypothesis of homogeneity, but is influenced by the outliers such that it identifies wrong time series as outlying with a high probability. For a detailed discussion we refer to section B.4 and the end of section B.3.2, where the performance of $\hat{\theta}_{LR}$ is compared to that of $\hat{\theta}_{PS}$, the estimator introduced in the next subsection.

3.7.2 Phase Space Representation

Gather, Imhoff and Fried (2002) propose a method for identifying outliers in stationary Gaussian time series by deriving a time series Mahalanobis distance. They represent the time series $\{X_t\}_{t=1,\ldots,T}$ in an m-dimensional phase space, i.e. they consider the set of vectors $\mathbf{x}_t = (X_t, X_{t+1}, \ldots, X_{t+m-1})'$, $t = 1, \ldots, T-m+1$. When the order p of the autoregressive time series is known, they choose the dimension m of the phase space as $p+1$. If the order is unknown, they choose it as $m = 1 + \max\{h; |\rho(h)| > 0\}$, where ρ is the partial autocorrelation function. For estimating m from the data, Gather, Imhoff and Fried (2002) use $m = 1 + \max\{h; |\hat{\rho}(h)| > u_{1-\alpha}\sqrt{\frac{1}{n}}\}$, where n is the length of the time series and $u_{1-\alpha}$ the $(1-\alpha)$-quantile of the standard normal distribution. $\hat{\rho}$ is the sample partial correlation function. In this setting, they consider the following analogue of the classical Mahalanobis distance outlier identifier.

3.7.6 DEFINITION

Let $\{X_t\}_{t=1,\ldots,T}$ *be a sample from a stationary Gaussian time series and define the phase space vectors as* $\mathbf{x}_t = (X_t, X_{t+1}, \ldots, X_{t+m-1})'$*,* $t = 1, \ldots, T-m+1$*. Denote their mean by* $\bar{x} = \sum_{t=1}^{T-m+1} x_t$ *and the corresponding* m*-dimensional sample covariance matrix by* $\hat{S}_m$*. The* Mahalanobis distance for time series (MDTS) *at point* t *is defined as*

$$MDTS_t = \sqrt{(\mathbf{x}_t - \bar{x})\, \hat{S}_m^{-1}\, (\mathbf{x}_t - \bar{x})} \quad \text{for } t = 1, \ldots, T-m+1\,.$$

In the above article, the authors also derive its asymptotic distribution.

3.7.7 THEOREM (GATHER, IMHOFF AND FRIED (2002))

Let $(X_1, \ldots, X_T)$ *be a sample from a stationary Gaussian process with absolutely summable autocovariance function* $\gamma(h)$*,* $h \in \mathbb{N}$*, and denote the dimension of the phase space by* m*. Then the asymptotic distribution of the Mahalanobis distance for time series (MDTS) is given by*

$$MDTS_t^2 \Rightarrow Y_t \qquad \text{for } T \to \infty, \qquad \text{where } Y_t \sim \chi^2_m\,.$$

Since the MDTS is defined using a non-robust covariance estimator, the outlier identifier is susceptible to masking effects (Becker and Gather 1999). Therefore (Gather, Imhoff and Fried 2002) suggest replacing the sample covariance matrix and the mean by the robust alternatives obtained from the minimum volume ellipsoid (MVE) method which is not affected by the dependencies in the time series.

We now extend their definition to panels of time series.

3.7.8 DEFINITION

Let $X_t^{(i)}$, $t = 1, \ldots, T$, $i = 1, \ldots, n$, *be observations from a panel of stationary time series* $\{X_t^{(i)}\}_{t \in \mathbb{Z}}$, $i = 1, \ldots, n$, *such that* $\mathbb{E}\, X_t^{(i)} = \mu$ *and* $\operatorname{var} X_t^{(i)} = \sigma^2 > 0$, *and let* m *be the dimension of the chosen phase space. Denote the corresponding phase space vectors by* $\mathbf{x}_t^{(i)} = \left(X_t^{(i)}, \ldots, X_{t+m-1}^{(i)}\right)'$, $t = 1, \ldots, T-m+1$, $i = 1, \ldots, n$. *Let*

$$\hat{\mu} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} X_t^{(i)} \quad \text{and} \quad \hat{S} = \left(\hat{\gamma}_{n,T}(k-l)\right)_{k,l=1,\ldots,m},$$

where $\hat{\gamma}_{n,T}(h) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=h+1}^{T} \left(X_t^{(i)} - \hat{\mu}\right)\left(X_{t-h}^{(i)} - \hat{\mu}\right)$ *is the panel covariance estimator. Moreover let* $\mathbb{1} = (1, \ldots, 1)'$ *be the* m*-dimensional vector consisting of ones and denote* $\underline{\hat{\mu}} = \hat{\mu}\,\mathbb{1}$.
Then the squared Mahalanobis distance for this panel of time series (MDP^2) *at the time series* i *is defined as*

$$MDP_i^2 = \sum_{t=1}^{T-m+1} \left(\mathbf{x}_t^{(i)} - \underline{\hat{\mu}}\right)' \hat{S}^{-1} \left(\mathbf{x}_t^{(i)} - \underline{\hat{\mu}}\right).$$

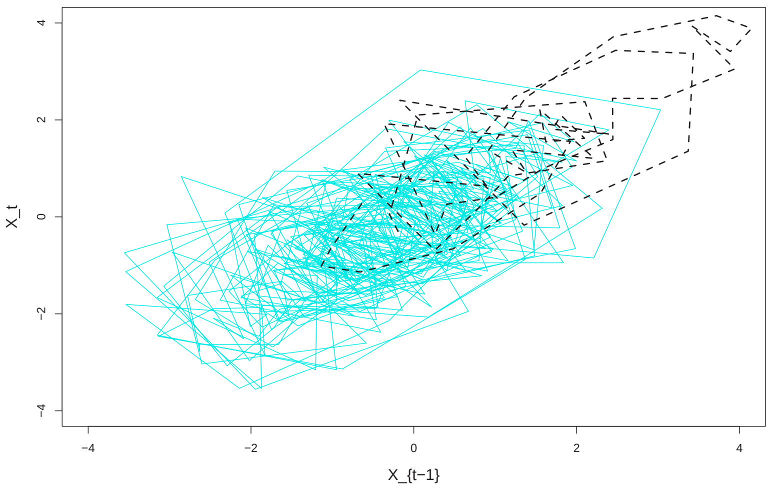

Figure 3.1: Phase space representation of a panel of 9 independent autoregressive time series with autoregressive parameter $a = 0.5$ and variance $\sigma^2 = 1$ (grey lines), and one autoregressive time series with parameter $a_{out} = 0.9$ and variance $\sigma^2_{out} = 1$ (dashed black line). The length of the time series is $T = 72$.

Thus the MDP_i^2 essentially is the sum of the Mahalanobis distances of the ith time series, the only difference being that the univariate estimators are replaced by their panel counterparts. The underlying idea is that the phase space vectors are m-dimensional Gaussian random variables, although not independent; and that time series belonging to different models possess different density ellipsoids (compare figure 3.1). The test then measures the deviation of the single time series from the main behaviour.
If the single phase space vectors were independent, the resulting statistic would be a sum of $T - m + 1$ independent random variables which each are asymptotically distributed as χ_m^2. As we are concerned with causal autoregressive processes where the autocorrelations decay exponentially, the strong mixing property ensures asymptotic normality.

3.7.9 THEOREM
Let $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \dots, n$, be a panel of identically distributed stationary Gaussian time series with absolutely summable autocovariance function $\gamma(h)$. and let m be the dimension of the chosen phase space. Using the notations introduced in the preceding definition, the asymptotic distribution of the Mahalanobis distance for observations of a panel of time series at time series i is given by

$$\sqrt{T}\left(\frac{1}{T}\,MDP_i^2 - 1\right) \Rightarrow Y \qquad \textit{for } n, T \to \infty,$$

where $Y \sim \mathrm{N}(0, \sigma_Y^2)$, the asymptotic variance σ_Y^2 being independent of n and T.
In order to ensure the consistency of the estimators $\hat{\mu}$ and $\hat{S}$ used for calculating the test statistic, we furthermore assume that the intercorrelation in the panel is for all $h \in \mathbb{Z}$ determined by $\mathrm{cov}\big(X_t^{(i)}, X_{t+|h|}^{(j)}\big) = u_n\,\gamma(h)$ for $i \neq j$, where $u_n = O\big(\frac{1}{n}\big)$.

PROOF:
Let $\Sigma = \big(\gamma(k-l)\big)_{k,l=1,\dots,m}$ be the true m-dimensional covariance matrix of the phase space vectors $x_t^{(i)}$ and $\underline{\mu} = \mu\,\mathbb{1}$ their true mean vector. Furthermore let $\hat{S}$ be the sample covariance matrix and $\underline{\hat{\mu}}$ the overall mean as defined above. Due to lemma 1.2.4 the entries of $\hat{S}$ fulfil

$$\mathbb{E}\big(\hat{\gamma}_{n,T}(h) - \gamma(h)\big)^2 = O\left(\frac{1}{n\,T}\right)$$

since $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ and the intercorrelation factor in the panel is $u_n = O\left(\frac{1}{n}\right)$. For the same reasons, also

$$\mathbb{E}\big(\hat{\mu} - \mu\big)^2 = O\left(\frac{1}{n\,T}\right) .$$

Since for each $i = 1, \dots, n$,

$$\begin{aligned}(\mathbf{x}_t^{(i)} - \underline{\hat{\mu}})'\,\hat{S}^{-1}\,(\mathbf{x}_t^{(i)} - \underline{\hat{\mu}}) &= (\mathbf{x}_t^{(i)} - \underline{\mu})'\,\Sigma^{-1}\,(\mathbf{x}_t^{(i)} - \underline{\mu}) + 2\,(\underline{\mu} - \underline{\hat{\mu}})'\,\Sigma^{-1}\,(\mathbf{x}_t^{(i)} - \underline{\mu}) \\ &\quad + (\underline{\mu} - \underline{\hat{\mu}})'\,\Sigma^{-1}\,(\underline{\mu} - \underline{\hat{\mu}}) + (\mathbf{x}_t^{(i)} - \underline{\mu})'\,\big(\hat{S}^{-1} - \Sigma^{-1}\big)\,(\mathbf{x}_t^{(i)} - \underline{\mu}) \\ &\quad + 2\,(\underline{\mu} - \underline{\hat{\mu}})'\,\big(\hat{S}^{-1} - \Sigma^{-1}\big)\,(\mathbf{x}_t^{(i)} - \underline{\mu}) + (\underline{\mu} - \underline{\hat{\mu}})'\,\big(\hat{S}^{-1} - \Sigma^{-1}\big)\,(\underline{\mu} - \underline{\hat{\mu}})\,,\end{aligned}$$

we thus have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-m+1} (\mathbf{x}_t^{(i)} - \underline{\hat{\mu}})'\,\hat{S}^{-1}\,(\mathbf{x}_t^{(i)} - \underline{\hat{\mu}}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-m+1} Y_t^{(i)} + O_P\Big(\frac{1}{\sqrt{n}}\Big)\,,$$

where $Y_t^{(i)} = (\mathbf{x}_t^{(i)} - \underline{\mu})' \Sigma^{-1} (\mathbf{x}_t^{(i)} - \underline{\mu}) \sim \chi^2_m$ due to the assumption of Gaussianity. For ease of notation choose $i \in \{1, \ldots, n\}$ fixed and denote $Y_t = Y_t^{(i)}$. Then $\mathbb{E}\, Y_t = 1$ for all $t = 1, \ldots, T - m + 1$. The Y_t are dependent. Using the fact that for Gaussian processes all cumulants of order larger than two are zero (Shiryayev 1984, p. 291), the variance of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T-m+1} Y_t$ can be calculated as

$$\sigma_T^2 = \operatorname{var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-m+1} Y_t\right) = \operatorname{var}\big(Y_t\big) + 2 \sum_{h=-(T-m)}^{T-m} \left(1 - \frac{|h|}{T}\right) \sum_{k,l=1}^{m} s_{kl}$$
$$\times \big(\gamma(k+h-1), \ldots, \gamma(k+h-m)\big)' \Sigma^{-1} \big(\gamma(l+h-1), \ldots, \gamma(l+h-m)\big)\,,$$

where s_{kl} denotes the (k,l)th entry of Σ^{-1}. Since by assumption $\operatorname{var} X_t^{(i)} > 0$, we have $\sigma_T^2 > 0$. Moreover $\sigma_Y^2 = \lim_{T\to\infty} \sigma_T^2$ exists because we have assumed the autocovariance function to be absolutely summable.
For deriving the asymptotic distribution, we employ strong mixing theory. We regard

$$S_T = \frac{1}{\sqrt{T\,\sigma_T^2}} \sum_{t=1}^{T-m+1} \big(Y_t - 1\big) = c_T \sum_{t=1}^{T-m+1} Z_t\,,$$

i.e. $c_T = \big(T\,\sigma_T^2\big)^{-\frac{1}{2}}$ and $Z_t = Y_t - 1$, $t = 1, \ldots, T - m + 1$.
Thus for all $t = 1, \ldots, T - m + 1$, Z_t is measurable and has mean zero. The choice of the constant gives $\mathbb{E}\, S_T^2 = 1$. Furthermore, stationary Gaussian AR-processes are strong mixing (Davidson 1994, p. 210). Therefore $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$ is strong mixing for each $i = 1, \ldots, n$. By theorem 14.1 of Davidson (1994), thus Z_t is strong mixing of the same order. Moreover it trivially is near-epoch dependent (Davidson 1994, definition 17.1) in L_2-norm. Since $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$ is assumed to be Gaussian for all $i = 1, \ldots, n$, all higher moments of $X_t^{(i)}$ and thus of Y_t and Z_t exist. Therefore, $\mathbb{E}\, Z_t^r = M_r < \infty$ for all $t = 1, \ldots, T - m + 1$. This leads to

$$\sup_{t=1,\ldots,T-m+1} \mathbb{E}\left(\frac{Z_t^r}{c_T^r}\right)^{\frac{1}{r}} = c_T\, M_r^{\frac{1}{r}} < \infty$$

for all fixed $r > 0$. As the coefficient c_T is constant for fixed T, we moreover have $\sup_{t=1,\ldots,T-m+1} T\,c_T^2 = T\,c_T^2 = \big(\sigma_T^2\big)^{-1} < \infty$. Thus, all conditions of theorem 24.6 of Davidson (1994) are fulfilled. It follows that for $T \to \infty$

$$S_T = c_T \sum_{t=1}^{T-m+1} Z_t \;\Rightarrow\; Z \qquad \text{where } Z \sim \mathrm{N}(0,1).$$

Equivalently,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T-m+1} \big(Y_t - 1\big) \;\Rightarrow\; Y, \qquad \text{where } Y \sim \mathrm{N}(0, \sigma_Y^2).$$

Altogether this means that in case of $T \to \infty$ and $n \to \infty$

$$\sqrt{T}\left(\frac{1}{T}\, MDP_i^2 - 1\right) \Rightarrow Y. \qquad \square$$

The most important point in this proof is that we only have used the fact that $\hat{S}$ and $\underline{\hat{\mu}}$ are $\sqrt{n\,T}$-consistent estimators of the covariance matrix and the mean vector. Thus any other $\sqrt{n\,T}$-consistent estimator can be used for defining a Mahalanobis-type distance, leading to the same asymptotic distribution. We employ the above result for deriving a robust test for outliers in a panel of time series.

Robust outlier identification in the panel case

In the simulations we generate panels of Gaussian time series, up to $m_{max} = \lfloor 0.2\,n \rfloor$ time series being replaced by outlying time series. These are assumed to be independent of the time series in the panel and of each other.
For robustly testing whether a time series is an outlier and for eliminating these from the data set we use the following procedure.

3.7.10 ALGORITHM
In order to obtain a robust estimate $rMDP_i^2$, we replace the sample covariance matrix and the mean vector used in the computation of MDP_i^2 by the covariance matrix obtained from the minimum covariance determinant (MCD) method and the overall median $\text{med}_{t=1,\ldots,T;i=1,\ldots,n} X_t^{(i)}$. A time series $\{X_t^{(i)}\}_{t=1,\ldots,T}$ is classified as an outlier if

$$\left| \sqrt{T} \left(\frac{1}{T} rMDP_i^2 - 1 \right) \right| > c_\alpha \, ,$$

where c_α is the $\left(1 - \frac{\alpha_n}{2}\right)$-quantile of the $\mathrm{N}(0, \sigma_Y^2)$-distribution. The adjusted significance level α_n is obtained from α by $\alpha = 1 - (1-\alpha)^{1/n}$. For approximating σ_Y^2 use a robust estimate of the empirical variance of the Y_t, $i = 1, \ldots, n$:

$$\hat{\sigma}_Y^2 = \text{med}_{i=1,\ldots,n} \frac{1}{T-m+1} \sum_{t=1}^{T-m+1} \left(Y_t^{(i)} - \bar{Y}^{(i)} \right)^2 \, ,$$

where $\bar{Y}^{(i)} = \frac{1}{T-m+1} \sum_{t=1}^{T-m+1} Y_t^{(i)}$. These outliers are deleted from the data set. The parameter estimation then can be performed in a second step using a non-robust estimator.

The significance level α has to be adjusted as the algorithm implies multiple testing: for each time series we compute the test statistic, where the estimate of the covariance matrix is based on all observations, and decide whether it is outlying or not. This adjustment is the same as chosen in the likelihood ratio procedure described in the previous subsection (see remark 3.7.5). As we here test for each time series separately, we have to adjust α with n, the number of time series in the panel, instead of u.

3.7.11 REMARK

1. Replacing the term $1 = \mathbb{E}\,Y_t$ by a robust estimate of this expectation in the test statistic leads to more reliable results. In fact we have used the estimator $\text{med}_{i=1,\ldots,n} MDP_i^2$ for calculating $\hat{\theta}_{PS;rob}$ in the simulations shown in the Appendix B.3.2.

2. The above choice of the robust covariance matrix facilitates the comparison with the parameter estimator $\hat{\theta}_{MCD}$ of section 3.5 which is based on the MCD. In fact, the two methods are closely connected. Robust methods allow to estimate the parameters and to identify outliers as the data with the largest residuals in one step. Here the MCD method is used to classify the outliers in order to be able to perform a non-robust estimation on a smaller data set in a second step, whereas $\hat{\theta}_{MCD}$ provides a robust estimator based on all data.

3. The definition of the Mahalanobis distance implicitly includes a transformation of the Gaussian m-dimensional random vectors to standard normally distributed random variables. Thus the test statistic yields a distance of the transformed time series to a standard normally distributed process. This idea is similar to that in Hallin and Puri (1988), where the authors test one ARMA against another ARMA model by checking whether a transformed ARMA processes is a white noise process. However here the focus is different. Hallin and Puri develop the asymptotic properties of the procedure with unspecified densities. Our main interest lies in methods adapted to the panel case which are robust and easily applicable.

For the performance of the method see the simulation study in section B.3.2 of the Appendix B and the discussion in the next section. In the simulations we also have included the non-robust outlier identification procedure based on the sample covariance and the sample mean for comparison. There $\hat{\sigma}_Y^2$ is estimated as

$$\hat{\sigma}_Y^2 = \frac{1}{n\,(T-m+1)} \sum_{t=1}^{T-m+1} \sum_{i=1}^{n} \left(Y_t^{(i)} - \bar{Y}^{(i)} \right)^2 .$$

In these cases the parameters are estimated by eliminating all identified outliers from the panel and then performing a non-robust estimation using the ICM parameter estimator of chapter 2. We call the resulting parameter estimators $\hat{\theta}_{PS}$ (non-robust method) and $\hat{\theta}_{PS;rob}$. In order to compare the estimators with those obtained using the likelihood ratio procedure, we further include in the simulations a (non-robust) modification where the time series are eliminated iteratively. The estimators obtained from this procedure are denoted by $\hat{\theta}_{PS;rec}$.

3.8 Conclusion and Outlook

We have investigated several approaches for obtaining robust panel covariance estimators and have evaluated their behaviour in a simulation study.
The first method is to robustify the conditional maximum likelihood estimators of the first chapter by replacing all non-robust parts with a robust method as it has been done e.g. in Haddad (2000). This leads to robust estimates of the autoregressive parameter. However, the simulations show that the estimators of the residual variances are biased (see section B.4), which is a known problem in robust estimation (Rousseeuw and Leroy 1987). As bootstrap methods are non-robust in character, they reflect the

empirical behaviour of the underlying parameter estimator and thus improve the estimates only moderately. Here the method which exploits the normality assumption of the model performs best (see section B.4).

As a second aspect we have investigated the effects of replacing the covariance matrix and the autocovariance vector in the Yule-Walker equations by a robust counterpart. We have focused on the panel scale estimator $Q_{n,T}$ which generalises the robust time series scale estimator Q_n proposed by Ma and Genton (2000). It was possible to define and compute panel breakdown points for $Q_{n,T}$. Some estimators with a high breakdown point such as the minimum volume ellipsoid (MVE) estimator have a large bias which may be so high as to make the estimator unreliable, even for small amounts of contamination (see Maronna et al. 1992). This seems not to be the case with $Q_{n,T}$. The simulations in the Appendix B.2 suggest that the original standardisation constant is not appropriate in the panel case, though. As this factor cancels out in the calculation of the autoregressive parameter it does not affect the parameter estimation as such. The drawback of estimating the separate components of the covariance matrix robustly is that the estimate is not necessarily positive definite. The fast algorithm of Croux and Rousseeuw (1992) could not be transferred to the panel case since it relies on a procedure for efficient partial sorting of a single vector of observations. But the computational speed posed no problem in the simulations.

Positive definiteness is ensured if we replace the entire covariance matrix by a robust counterpart. We chose the minimum covariance determinant (MCD) covariance estimator because it is implemented in R, whereas other covariance estimators cited in the introduction are not yet available. The MCD is reported to be more stable than the minimum volume ellipsoid (MVE) method, and it is more efficient in high dimensions (Croux and Haesbroeck 1999).

Although the approach of first identifying and then deleting outliers is very intuitive, it poses two problems. If the procedure used for identification is not robust, this method can lead to a masking effect (Rousseeuw and Leroy 1987, Becker and Gather 1999). This means that the test statistic is influenced by the outliers such that they are not recognised as outlying, whereas some of the original data may wrongly be identified as outliers. This masking effect is evident for the non-robust likelihood ratio test of section 3.7 (see the simulation study in section B.3.1 of the Appendix B). Secondly, iterating the procedure implies multiple testing. Thus the significance level of the test has to be adjusted. For this, the maximal proportion of possible outliers must be specified in advance. Furthermore this implies that the local tests are performed at a much higher significance level, which makes rejections for the single time series less probable. Thus Rousseeuw and Leroy (1987) prefer genuinely robust estimators which allow estimating the parameters and identifying outliers in the same time. There the outliers are characterised by their large residuals. For example the least trimmed squares estimator and the reweighted least squares estimator discussed in section 3.3 downweight the observations which belong to the largest residuals in the estimation procedure. The phase space method of outlier recognition is related to this class of estimators. The procedure starts with computing a robust covariance matrix using the MCD method. This can then either be used directly in the the Yule-Walker equations (thus leading to the estimator discussed in section 3.5) or in the robust outlier test treated in section 3.7.2.

For a more detailed comparison of the different properties of the above estimators we refer to the discussion of the simulation study in section B.4 in the Appendix. The final recommendation is to use the reweighted robustified ICM parameter estimator $\hat{\theta}_{rw}$ or the estimator obtained after a preliminary outlier detection using the the phase space test, $\hat{\theta}_{PS;rob}$, for the estimation if no arbitrary outliers are present. In the case of arbitrary outliers, one should use the estimators derived from the robust covariance matrices $\hat{\theta}_{Q}$ or $\hat{\theta}_{MCD}$, depending on the order of the autoregressive processes.

As the aim of this chapter was to survey which robust methods could be used for the panel case, the character of our investigations is exploratory and we have not sought to improve the single estimators as much as possible. Certainly there exist modifications of some of the above parameter estimators which perform better than these. In particular it seems that the bias of the robust autocovariance estimator $\hat{\theta}_{Q}$ can be lowered by adapting the choice of the order statistic and the standardisation to the panel case.

For a more detailed discussion of the empirical behaviour of the proposed estimators, we refer to the Appendix B, and in particular to section B.4.

Chapter 4

Real Data Examples

4.1 Introduction

This chapter is concerned with the analysis of data collected in experiments. We investigate how our methods can be applied in practice. The panels of intercorrelated time series we analyse have already been mentioned several times throughout the thesis. The first one is the data set which was the motivation for the present thesis. It originates from a therapy process study conducted at the Medical University Hospital of Heidelberg, Department of Internal and Psychosomatic Medicine. Fibromyalgia Syndrome (FMS) patients were undergoing a treatment consisting of several modules, including a psychotherapeutic group therapy. Therefore the assumption that the time series obtained from these patients are independent cannot be made initially. A second example, where the presence of intercorrelation is predominant, is the grey-sided voles data set presented in section 2.1. It already served Hjellvik and Tjøstheim (1999a) and Fu et al. (2002) as an example. These two data sets allow us to elaborate the features of the parameter estimators in typical applications.

The chapter is structured as follows. We start with estimating the autoregressive parameters for the voles data. Then we investigate the FMS data. In such studies (and this indeed has been the case) some patient may have been wrongly assigned to the therapy group. Thus we employ an outlier identification step before analysing the remaining data. All of the analyses are exploratory in character as we primarily want to illustrate the properties of the different estimators. The chapter concludes with a discussion of the obtained results.

4.2 Population Dynamics

The grey-sided voles data has already been briefly introduced in section 2.1. The data set is plotted there, in figure 2.1. The data, which are also investigated in Hjellvik and Tjøstheim (1999a) and Fu et al. (2002) as an example of intercorrelated time series, consist of the yearly catches (from 1962 to 1992) of grey-sided voles at 41 different locations on Hokkaido, Japan. They are measured on a logarithmic scale: if the number of voles trapped each year is denoted by $\{V_t^{(i)}, 1962 \leq t \leq 1992, 1 \leq i \leq 41\}$, we consider the transformed data $X_t^{(i)} = \log(V_{1961+t}^{(i)} + 1)$, $1 \leq t \leq 31$, $1 \leq i \leq 41$.

As we want to use the methods discussed in this thesis for the analysis, we first must decide whether the assumptions 2.2.1 of the ICM or 2.3.1 of the GICM are fulfilled. In the present case it is however justified to use only the GICM procedure, even if the true model fulfils the assumptions of the ICM, since here $n = 41$. Thus the difference between the ICM estimator $\hat{\theta}_{n,T}$ (calculated using the iterative algorithm 2.4.4) and the GICM estimator obtained in lemma 2.4.7, is small (see remark 2.6.10). Note that $\hat{\theta}_a$ coincides with the estimator of Hjellvik and Tjøstheim (1999a) (see remark 2.4.8).

Preprocessing the data

Since the empirical mean of the data is $\hat{\mu} = 1.7$, we must preprocess the data prior to the analysis. The main question is whether the mean term is common to all time series in the panel or whether we have to subtract different means from the individual time series. Regarding the empirical data both possibilities are reasonable. Hjellvik and Tjøstheim (1999a) assume a common mean, stating that the data has been chosen from a larger data set such that the difference in the individual means was minimised. This assumption is sufficient for being able to employ their estimator, as then the residual time series $\mathring{X}_t^{(i)} = X_t^{(i)} - \bar{X}_t$, $t \in \mathbb{Z}$, $i = 1, \ldots, n$, have zero mean.
We here however do not want to make that assumption and thus preprocess the data in the following way.

- ◦ We subtract the individual sample means $\lambda_i = \frac{1}{31} \sum_{t=1}^{31} X_t^{(i)}$ from the observations of the corresponding single time series $\{X_t^{(i)}\}_{t=1,\ldots,31}$, $i = 1, \ldots, 41$. The resulting data set is called V_{ind}.

- ◦ For comparing our results to those of Hjellvik and Tjøstheim, we only subtract the overall sample mean $\mu = \frac{1}{31\times 41} \sum_{t=1}^{31} \sum_{i=1}^{41} X_t^{(i)}$ from the set of observations $\{X_t^{(i)};\ t = 1, \ldots, 31, i = 1, \ldots, 41\}$. We denote this data set by V_{HT}.

Data analysis

The data set has been used already in several studies (see e.g. Hjellvik and Tjøstheim 1999a). Thus we as well assume that the analyses are not influenced by outliers and therefore employ non-robust estimators. This is further supported by the fact that robust analyses here lead to qualitatively the same results. In order to simplify the presentation we omit these here. We proceed as follows. For both of the two transformed data sets V_{HT} and V_{ind} we compute parameter estimates using GICM parameter estimators $\hat{\theta}_a$ and $\hat{\theta}_b$. Furthermore, we give the results of the ICM parameter estimator $\hat{\theta}_{n,T}$ for comparison. Following the procedure of Hjellvik and Tjøstheim (1999a), we fit autoregressive processes of different orders to the data. The quality of the fits is assessed by computing the residual processes and performing diagnostic checks on these. If there are several competing models, those with a lower order are preferred. We first consider the individual processes. The results of the estimation are displayed in table 4.1. It turns out that for the model V_{HT} indeed an AR(4) process yields the best fit. This corresponds to the findings of Fu et al. (2002). In Hjellvik and Tjøstheim (1999a), the authors only investigate processes up to order three. However they also do not propose

data set		model	estimated model parameters
V_{HT}	$\hat{\theta}_a$	AR(4)	$\hat{a} = (0.136, -0.007, 0.082, 0.132)'$, $\hat{\sigma}^2 = 0.581$
	$\hat{\theta}_{n,T}$	AR(4)	$\hat{a} = (0.131, -0.012, 0.085, 0.136)'$, $\hat{\sigma}^2 = 0.581$
V_{ind}	$\hat{\theta}_a$	WN	$\hat{\sigma}^2 = 0.553$
	$\hat{\theta}_{n,T}$	WN	$\hat{\sigma}^2 = 0.553$

Table 4.1: Coefficients of the individual processes obtained from the estimators $\hat{\theta}_a$ and $\hat{\theta}_{n,T}$, computed for each of the two data sets V_{ind} and V_{HT}. The third column gives the type of model chosen (WN=white noise).

a white noise model but an AR(3) process for fitting the data. Thus we see that the results obtained for the differently transformed data set V_{ind}, where we get a white noise model, are clearly distinct from those based upon V_{HT}.
For the estimation of the background process using theorem 2.4.14 we take the empirical mean process $\{\frac{1}{41}\sum_{i=1}^{41} X_t^{(i)}\}_{t=1,\dots,31}$. This process is the same for both data sets V_{HT} and V_{ind}. The analyses yield that it is best approximated by a white noise process with the variance $\tau^2 = 0.444$.

Implications for modelling

The above analysis yield similar parameter estimates for the ICM estimator $\hat{\theta}_{n,T}$ and the GICM estimator $\hat{\theta}_a$. However we have to take the structure of the background process into account, which is estimated from the empirical mean process $\{\frac{1}{41}\sum_{i=1}^{41} X_t^{(i)}\}_{t=1,\dots,31}$. If the ICM model (definition 2.2.2) is true, then the mean process fulfils

$$\bar{X}_t = \sum_{k=1}^{p} a_k \, \bar{X}_{t-k} + \bar{\varepsilon}_t + \eta_t \qquad \text{for } t \in \mathbb{Z},\ i = 1, \dots, n,$$

thus having the same autoregressive parameters as the individual processes. Note that here $\{\eta_t\}_{t\in\mathbb{Z}}$ is assumed to be a white noise process. This property is not fulfilled in the case of the data set V_{HT}. Indeed modelling the process with the parameters obtained from the analysis of the individual processes leads to a fit which is much worse than adapting a white noise model. Using a GICM model (definition 2.3.2) for the data, we get that for all $t \in \mathbb{Z}$, $i = 1, \dots, n$,

$$X_t^{(i)} = Z_t^{(i)} + Y_t = \sum_{u=0}^{\infty} \psi_u \varepsilon_{t-u}^{(i)} + Y_t = \sum_{u=0}^{\infty} \psi_u \left(\varepsilon_{t-u}^{(i)} + a(\mathrm{L})\, Y_{t-u} \right),$$

where a is the autoregressive parameter, $a(\mathrm{L})$ the backward shift operator and $\{\psi_u\}_{u\geq 0}$ are the MA(∞) coefficients corresponding to a (see section 1.1). This is equivalent to

$$X_t^{(i)} = \sum_{k=1}^{p} a_k \, X_{t-k}^{(i)} + \varepsilon_t^{(i)} + a(\mathrm{L})\, Y_t \qquad \text{for } t \in \mathbb{Z},\ i = 1, \dots, n.$$

Thus the background process $\{Y_t\}_{t\in\mathbb{Z}}$ being a white noise process corresponds to $\{\eta_t\}_{t\in\mathbb{Z}}$ having an autoregressive structure, whereas in the ICM $\{\eta_t\}_{t\in\mathbb{Z}}$ is required to be a white noise process. This has already been discussed in the examples 2.3.4, but the grey-sided voles data set illustrates this fact nicely. Indeed Hjellvik and Tjøstheim (1999a), who estimate $\{\hat{\eta}_t\}_{t\in\mathbb{Z}}$ from the residual process $\hat{\eta}_t = \hat{a}(\mathrm{L})\,\bar{X}_t$ for $t \in \mathbb{Z}$, also conclude that it may be autocorrelated, which contradicts the ICM assumption.

4.3 Fibromyalgia Syndrome Therapy Study

Our second real-data example comes from a therapy process study on fibromyalgia syndrome (FMS) patients conducted at the University Hospital of Heidelberg, Department of Internal and Psychosomatic Medicine. This study was the original motivation for the present thesis. FMS is a chronic pain disease which is characterised by widespread pain and a reduced pain threshold (Wolfe et al. 1990). The therapy based on a psycho-bio-social approach consists of several modules, combining information, medication, physical therapy and a psychotherapeutic group therapy (Eich et al. 1998). This implies that assuming independence is not justified in this setting, which led us to investigate panels of intercorrelated time series.

FMS being a chronic pain disease, the therapy's main focus is on helping patients to cope better in their daily life. FMS patients often display a number of physical and psychosomatic attendant symptoms, among these are sleep disorders, anxiety and an elevated level of depressivity. Thus the parameters of main interest are, besides pain intensity, the levels of depressivity (mood) and self-efficacy. The latter is a measure of how much a patient believes that he or she can influence the symptoms of the disease himself (Müller et al. 2003). Using graphical models for time series (Dahlhaus 2000a), it has been shown that self-efficacy plays a central role in the therapy process and is supposed to serve as a mediator between other parameters such as pain intensity, sleep quality, anxiety and depression (Feiler et al. 2005).

Data

58 female patients participated in the study. They entered the data themselves into a handheld computer (Psion 3mx) which served as an electronic diary. The data were measured using visual analogue scales ranging from 1 to 10.

The data comprise 72 daily entries, i.e. they cover the span from the beginning of the therapy until two weeks after its termination. As the patients were divided into separate therapy groups, we here analyse the data of 11 patients participating in the same group (group 1). Outlying values in the individual time series, which were e.g. due to retrospective entries, were identified and eliminated preliminary to the analysis. Each univariate time series was detrended using 5th order polynomial trends and standardised with its empirical standard deviation. Missing values were replaced by a weighted average of forward and backward predictions using univariate autoregressive processes. For the analysis of the parameter depressivity we exclude the data of two patients from the analysis who exhibit virtually no variation in this parameter over long stretches of time. As an example, the data obtained for the parameters "pain intensity" and "self-efficacy" are displayed in figure 4.1.

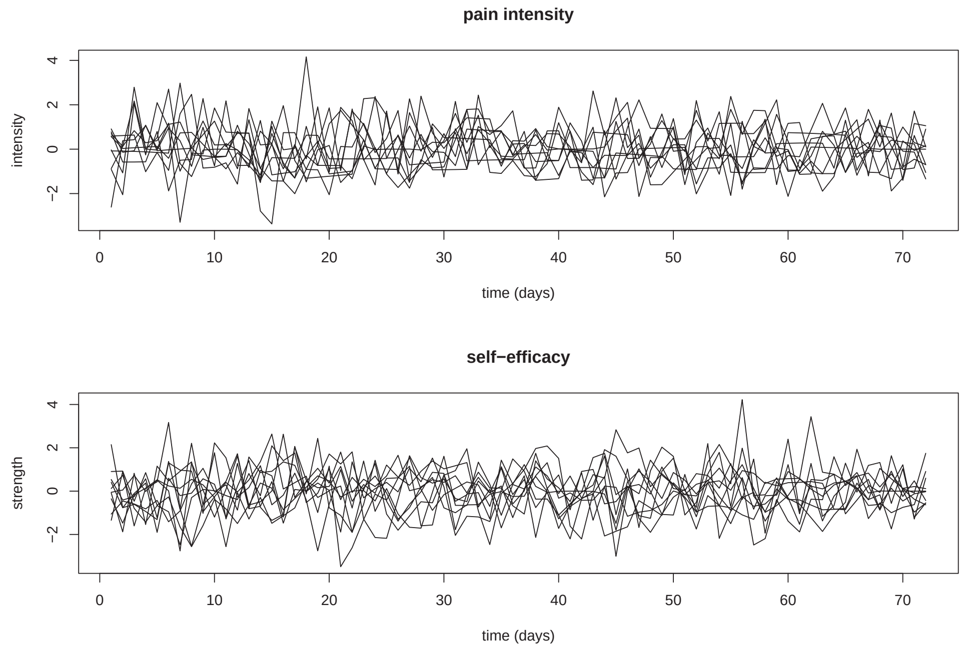

Figure 4.1: Standardised data for the parameters "pain intensity" and "self-efficacy" (group 1), measured over 72 days.

Intercorrelation

The plots of the data shown in figure 4.1 do not exhibit an obvious intercorrelation pattern as it is the case for the voles data shown in section 2.1. Nevertheless we cannot exclude intercorrelation due to theoretical reasons. The asymptotic results in this thesis are valid whether or not intercorrelation is present. Thus we test for intercorrelation first in order to get a clearer picture. For testing we use the method given in Brillinger (1973). This is based on the spectral representation of the data and tests whether the the spectrum of the random variable causing the common influence is different from zero. Applied for the parameters of interest with significance level $\alpha = 0.05$, it only is significant for the self-efficacy. For the other parameters it does not reject the null hypothesis of independence. The corresponding plots, again restricted to the parameters "pain intensity" and "self-efficacy", are displayed in figure 4.2.

Testing for outliers

Next we test for outliers in the data. We know that one person in the group (patient no. 11) did not suffer from FMS. But as she already had been accidently admitted to the study she was allowed to participate in the therapy group. For testing we use the robust test based on the phase space representation described in section 3.7.

The test performed at the 5% significance level does not give any significant results. In table 4.2 we list the p-values for those cases where the p-values are below 20%.

The results for the parameter "sleep quality" are not significant since we have to adjust for the multiple testing. Thus the individual tests have to be performed at a local significance level of 0.005 in order to guarantee the 5% significance level for the test.

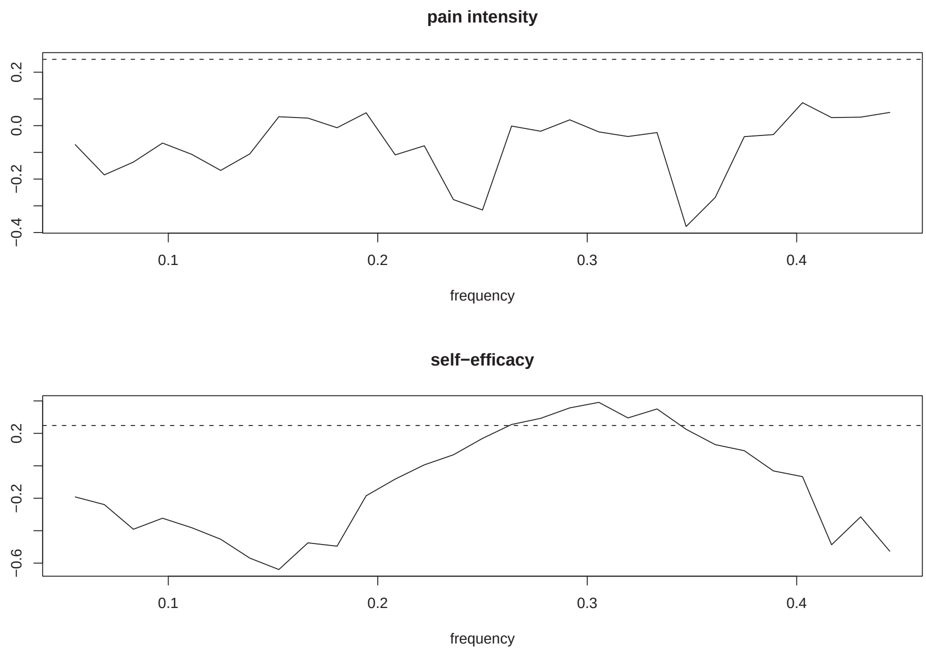

Figure 4.2: Results of the intercorrelation test for the parameters “pain intensity” and “self-efficacy” (group 1). Solid line: test statistic. Dashed horizontal line: 5%-bound for the test statistic. The curves were smoothed using a window of width 7.

For comparison we have performed the test also in its non-robust version, using the sample covariance matrix instead of the robustly estimated one for testing. This leads to qualitatively the same results.

Fitting autoregressive processes

In the case of independent time series we can estimate the parameters using ordinary least squares, which corresponds to a conditional maximum likelihood procedure in the case of Gaussian distributions. The intercorrelation test above did in most cases not reject the hypothesis of independence. However this does not prove that these time series indeed are independent. Therefore we now compare the estimates obtained under the assumption of independence to those obtained using the GICM estimator $\hat{\theta}_a$ given in proposition 2.4.7. Since the previous test does not indicate the presence of outliers (with a possible exception in the case of sleep quality), it is sufficient to use non-robust estimators. As in the last section, we fit autoregressive processes of different orders to the univariate data and check which of these model the data best. We first present the results of the direct least squares fit in table 4.3. Estimating the parameters with the ICM procedure yields virtually the same parameters, thus we have omitted them form this presentation. Table 4.3 furthermore shows that there is essentially no difference between the estimates for sleep quality with or without the data of patients 7 and 8.

The models fitted using the GICM procedure are given in table 4.4.

parameter	
pain and depressivity	no incidents
self-efficacy	patient 11 (p=15.86%)
anxiety	patient 5 (p=16.46%)
sleep quality	patients 7 (p=1.62%) and 8 (p=7.06%)

Table 4.2: Results of the robust outlier tests. Cases with p-values $\leq$ 20%.

parameter	model	estimated model parameters
pain intensity	AR(4)	$\hat{a} = (0.093, -0.003, -0.063, -0.141)'$, $\hat{\sigma}^2 = 0.949$
self-efficacy	AR(1)	$\hat{a} = 0.152$, $\hat{\sigma}^2 = 1.019$
depressivity	AR(1)	$\hat{a} = 0.141$, $\hat{\sigma}^2 = 1.009$
anxiety	AR(1)	$\hat{a} = 0.117$, $\hat{\sigma}^2 = 0.938$
sleep quality	WN	$\hat{\sigma}^2 = 1.007$
sleep quality without 7 and 8	WN	$\hat{\sigma}^2 = 1.063$

Table 4.3: Coefficients of the autoregressive processes fitted using a least squares procedure. The second column gives the type of model chosen (WN=white noise).

Simulations

In order to explore whether the more complicated models obtained from the GICM procedure are due to overfitting, which implies that the models which have been identified for the background processes are artefacts, we perform a small simulation.
We generate panels of independent autoregressive processes with identical parameters $a = (0.093, -0.003, -0.063, -0.141)'$ and $\sigma^2 = 0.949$ (pain intensity) and with parameters $a = 0.141$ and $\sigma^2 = 1.009$ (depressivity). Thus the variance of the mean process will be approximately 0.1. Furthermore we simulate panels of white noise processes with variance $\sigma^2_{WN} = 1$ and a panel of intercorrelated autoregressive time series with parameters $\alpha = 0.141$ and $\sigma^2 = \tau^2 = 0.5$. As size of the panel we choose $n = 10$ and $T = 72$, such that the results are compatible to the above analyses. Thus we have in the last model that $\omega_n^2 = 0.55$. For each type of panel we compute the parameters estimated from the mean process over 1,000 iterations. The results are

- *in the AR(4) case:* in 154 of the 1,000 iterations indeed an AR(4) process has been chosen. However, the average over the estimated autoregressive parameters is $\hat{a} = (0.088, -0.027, -0.105, -0.289)'$ (mean taken over those cases where an AR(4) process has been fitted), having a componentwise standard deviation $\text{sd}(\hat{a}) = (0.125, 0.152, 0.112, 0.064)'$. This means that the variance in the estimation is very high. The average variance is $\sigma^2 = 0.111$ ($\text{sd}(\hat{\sigma}^2) = 0.019$). Moreover in 541 cases the chosen model was a white noise model with similar variance.

parameter		model	estimated model parameters
pain intensity	$\hat{\theta}_a$	AR(4)	$\hat{a} = (0.073, -0.014, -0.075, -0.154)'$, $\hat{\sigma}^2 = 0.97$
	$\hat{\theta}_b$	AR(1)	$\hat{b} = 0.227$, $\hat{\omega}_n^2 = 0.08$
self-efficacy	$\hat{\theta}_a$	AR(4)	$\hat{a} = (0.187, 0.090, -0.050, -0.121)'$, $\hat{\sigma}^2 = 0.99$
	$\hat{\theta}_b$	WN	$\hat{\omega}_n^2 = 0.11$
depressivity	$\hat{\theta}_a$	AR(6)	$\hat{a} = (0.121, -0.003, -0.081, -0.086, 0.003, -0.125)'$, $\hat{\sigma}^2 = 0.99$
	$\hat{\theta}_b$	AR(2)	$\hat{b} = (0.309, -0.173)'$, $\hat{\omega}_n^2 = 0.11$
anxiety		AR(1)	ICM: $\hat{a} = 0.117$, $\hat{\sigma}^2 = 0.967$, $\hat{\omega}_n^2 = 0.097$
sleep quality	$\hat{\theta}_a$	WN	$\hat{\sigma}^2 = 1.01$
	$\hat{\theta}_b$	AR(4)	$\hat{b} = (0.165, 0.229, 0.072, -0.372)'$, $\hat{\omega}_n^2 = 0.08$
sleep quality, without 7 and 8	$\hat{\theta}_a$	WN	$\hat{\sigma}^2 = 1.07$
	$\hat{\theta}_b$	AR(5)	$\hat{b} = (0.193, 0.164, 0.085, -0.300, -0.213)'$, $\hat{\omega}_n^2 = 0.08$

Table 4.4: Coefficients of the autoregressive processes fitted using the GICM procedure. $\hat{\theta}_a$ and $\hat{\theta}_b$ are the parameters of the individual processes and the background process, respectively. The third column gives the type of model chosen (WN=white noise).

- *in the AR(1) case:* in 626 of the 1,000 iterations the process was correctly identified as an AR(1) process and in 2 a white noise process has been chosen. However there were for example 146 cases where an AR(4) model was fitted. The average over the estimated autoregressive parameters (mean taken over those cases where an AR(1) process has been fitted) is $\hat{a} = 0.380$ ($\mathrm{sd}(\hat{a}) = 0.140$), the average variance is $\hat{\sigma}^2 = 0.109$ ($\mathrm{sd}(\hat{\sigma}^2) = 0.018$).

- *in the white noise case:* only in 733 cases the mean process was correctly identified as a white noise process. The mean of the estimated autoregressive parameters is $\hat{a} = -0.010$ ($\mathrm{sd}(\hat{a}) = 0.095$) and of the mean of the estimated variances it is $\hat{\sigma}^2 = 0.109$ ($\mathrm{sd}(\hat{\sigma}^2) = 0.018$).

- *in the intercorrelated AR(1) case:* only in 90 cases a model of order three or larger was fitted. In 73 case an AR(2) model and in 574 cases a white noise model was fitted. The average over the estimated autoregressive parameters is $\hat{a} = 0.038$ (mean taken over those cases where an AR(4) process has been fitted), with standard deviation $\mathrm{sd}(\hat{a}) = 0.091$. The variance has been estimated as $\hat{\omega}^2 = 0.550$ ($\mathrm{sd}(\hat{\omega}^2) = 0.095$).

This shows that the variance of the true mean process is always estimated quite accurately, whereas the parameter estimations exhibit a large variation, in particular if the time series in the panel are independent. In the case of the intercorrelated panel however the estimated order is not varying as much as in the panel of independent AR(1) processes.

Conclusion

The above analyses show that the intercorrelation in the present data sets is only weak. Furthermore the simulations illustrate that the GICM procedure leads to an overfit in this case. We therefore conclude that the FMS data are best modelled as panels of independent time series with the parameters as given in table 4.3. There the pain intensity is modelled as an AR(4) process. Indeed the therapy sessions showed that patients tend to be more active in periods with a lower pain intensity, which often leads to an overload and thus to a higher pain level a few days later. This is an example which justifies fitting a higher order process to the parameter “pain intensity”. The other parameters are modelled as AR(1) processes, with exception of the sleep quality. This means that the levels of the parameters on one day have some influence on their value the next day, which is very plausible. The fact that sleep quality is best modelled as a white noise process may have its reasons in physiological reality. However it could also be due to the self-recording. It has been shown that self-recorded sleep quality may differ from the actual one (Wilson et al. 1998). Testing robustly for outliers has not detected the data of the patient not suffering from FMS as outlying. There are several possible explanations for this fact. The data set could have been too small, such that the test was not able to discover the (existing) differences. Another possibility is that the dynamical structures of the univariate time series do not differ. Then still the interaction structure (Feiler et al. 2005) of the parameters can exhibit a distinct dynamic behaviour, which is however not captured by the univariate analyses.

4.4 Discussion

The above examples illustrate various features of our methods. Applying the ICM and GICM parameter estimator to the grey-sided voles data investigated by Hjellvik and Tjøstheim (1999a), we can confirm the results of Fu et al. (2002) that at least four autoregressive parameters should be included in the model of the individual processes. We have however seen that the results depend critically on the assumptions on the data set. If we allow the time series to possess different individual means, it is better to transform the data by subtracting the individual means from each time series. Then the above analysis yields a white noise model. Hjellvik and Tjøstheim (1999a) state that the grey-sided voles data set already has been chosen from a larger one in order to minimise individual differences. This justifies their procedure which assumes a common mean for all time series in the panel.

For a robust analysis we cannot use the transformed data sets V_{HT} and V_{ind} employed above, as these are obtained by subtracting sample means which are not robust. We here have to subtract the respective medians for generating the transformed data sets. It has been shown in section B.1 of the simulation study in the Appendix B that robust estimators as e.g. $\hat{\theta}_{rob}$ furthermore have a larger variance and therefore are less reliable. In the present case these analyses lead however to qualitatively the same results as the non-robust procedure. Thus we have omitted them here in order to simplify the presentation.

At the end of section 4.2 we have discussed that forcing the data into the ICM scheme yields misleading results. It implies that the background process has the same au-

toregressive structure as the individual processes or in other words that the common influence is a white noise process. In the presence of a strong background process it is therefore not recommendable to employ the ICM parameter estimator alone. The advantage of employing the GICM estimator $\hat{\theta}_a$ also used by Hjellvik and Tjøstheim (1999a) is that we need no assumptions on the structure of the background process. It may even be deterministic (see remark 2.4.8).
The second example addresses the aspects of testing in panels of intercorrelated time series and of the behaviour of the ICM and GICM parameter estimators in the case of weak intercorrelation. In particular it illustrates that the results obtained from smaller data sets have to be carefully interpreted. For example, the intercorrelation test of Brillinger (1973) identifies the data set belonging to the parameter "self-efficacy" as intercorrelated. This is not confirmed by the further analyses. As the test is an asymptotic test, the reason may be that the data set was too small for obtaining reliable results. The robust test for outlying time series does not yield significant results either although we know that one of the patients did not suffer from FMS. One problem, which is also seen from section B.3.2 in the Appendix B, is that the test is rather conservative if the difference in the dynamic structures in not large. Because of the implicit multiple testing we must adjust the significance level (see remark 3.7.5). This implies that the single tests have to be performed at a very high significance level in order to guarantee a nominal significance level of 5%. Furthermore σ_Y^2, the variance of the asymptotic distribution, can only be approximated and the covariance matrix used for the testing also has to be estimated from the data (see theorem 3.7.9 and algorithm 3.7.10). Thus the sample size is probably too small to yield significant results. It however can also be the case that the time series obtained from the patient not suffering from FMS do not differ in their univariate structure from those of the other patients. Nevertheless the interaction structure of the parameters may be different (see e.g. Feiler et al. 2005), which cannot be detected from the univariate analyses. For a more detailed discussion of the interpretations of the results we refer to the last section.
The simulations performed in the last section demonstrate that using the GICM procedure indeed leads to an overfit if there is no or only a weak background process. If the n processes in the panel are independent, each point of the mean process converges almost surely to zero for $n \to \infty$ due to the strong law of large numbers. Thus the autoregressive parameter of the mean process is asymptotically not identifiable. In the situation of small panels we can however still infer about the mean process, which makes the above overfit possible. If the variances given by the GICM are small (around $\hat{\sigma}^2/n$ if $\hat{\sigma}^2$ is the estimated variance of the individual process and n the number of time series in the panel), the consequence is therefore that the analysis should be performed using the ordinary least squares procedure or the ICM method instead. Employing the ICM estimators implicitly means that the common influence is modelled as a white noise process with variance $\tau^2 > 0$ or does not exist at all. In contrast, least squares estimation is entirely based on the assumption on independence. Thus it is advisable to use the ICM estimator if an intercorrelation cannot be excluded by theoretical arguments but the GICM analysis indicates that this intercorrelation is weak. As can be seen from the analysis of the FMS data, in the case of no intercorrelation this leads to the same results as the least squares estimation.

Appendix A

Simulation Results for the ICM and GICM parameter estimators

The following simulation study compares the performance of the ICM estimator obtained using the minimisation algorithm of section 2.4.2 with the estimator of Hjellvik and Tjøstheim (1999a). There $\{\eta_t\}_{t\in\mathbb{Z}}$ is treated as a nuisance parameter. The estimator $\hat{a}_{HT}$ is obtained by minimising $\mathcal{L}^{\circ}_{n,T}$ under the restrictions of the ICM as described in proposition 2.4.7; $\hat{\sigma}^2_{HT}$ and $\hat{\omega}^2_{HT}$ then can be derived from the corresponding residuals (see remark 2.4.8).
The data is simulated from the ICM model (assumption 2.2.1). As variance of the processes $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, we always fix $\operatorname{var} X_t^{(i)} = 1$, i.e. $\sigma^2 + \tau^2 = 1$. The examples considered correspond to those treated in Hjellvik and Tjøstheim (1999a). The estimates are obtained by calculating the mean and standard deviation for each parameter over 5,000 independent realisations per model.
In the tables, the upper rows contain the estimates obtained from the minimisation algorithm. Subsequently follow the estimates from Hjellvik and Tjøstheim's procedure, indexed by HT. The empirical standard deviations are displayed in brackets below the estimated parameters.
We regard first small panels consisting of $n = 2$ and $n = 4$ time series. Then we investigate the behaviour for $T \to \infty$ more closely, for a small ($n = 3$) and a large ($n = 128$) panel. Finally we regard an AR(6) process. The section concludes with a brief discussion of the results.

A.1 Small Panels

Here we investigate the behaviour of the estimator in panels consisting of a small number of time series ($n = 2$ or $n = 4$). The size of the data is $nT = 200$, $nT = 2000$ and $nT = 20000$; the intercorrelation varies from no ($\tau^2 = 0$) to strong intercorrelation ($\tau^2 = 0.9$).
We regard an AR(1) process with parameter $a = 0.5$. Three models with different strengths of intercorrelation are simulated:

- $\tau^2 = 0$, i.e. $\sigma^2 = 1$. Thus $\omega_n^2 = 0.5$ ($n = 2$) or $\omega_n^2 = 0.25$ ($n = 4$).
- $\tau^2 = 0.5$, i.e. $\sigma^2 = 0.5$. Here $\omega_n^2 = 0.75$ ($n = 2$) or $\omega_n^2 = 0.625$ ($n = 4$).

		$nT = 200$		$nT = 2{,}000$		$nT = 20{,}000$	
		$n = 2$	$n = 4$	$n = 2$	$n = 4$	$n = 2$	$n = 4$
$\tau^2 = 0$	$\hat{a}$	0.4960 (0.0620)	0.4957 (0.0618)	0.4993 (0.0195)	0.4994 (0.0197)	0.4999 (0.0062)	0.5000 (0.0061)
	$\hat{\sigma}^2$	0.9944 (0.1423)	0.9967 (0.1178)	0.9996 (0.0439)	0.9998 (0.0361)	1.0000 (0.0141)	1.0001 (0.0116)
	$\hat{\omega}_n^2$	0.4961 (0.0718)	0.2483 (0.0507)	0.4994 (0.0227)	0.2502 (0.0161)	0.5000 (0.0070)	0.2500 (0.0050)
	$\hat{a}_{HT}$	0.4909 (0.0870)	0.4947 (0.0710)	0.4989 (0.0277)	0.4993 (0.0228)	0.4999 (0.0088)	0.5000 (0.0072)
	$\hat{\sigma}^2_{HT}$	0.9895 (0.1418)	0.9951 (0.1176)	0.9992 (0.0439)	0.9996 (0.0361)	1.0003 (0.0145)	0.9999 (0.0117)
	$\hat{\omega}^2_{HT}$	0.5036 (0.0739)	0.2511 (0.0513)	0.5002 (0.0227)	0.2505 (0.0162)	0.5001 (0.0072)	0.2501 (0.0051)
$\tau^2 = 0.5$	$\hat{a}$	0.4953 (0.0614)	0.4945 (0.0617)	0.4999 (0.0191)	0.5001 (0.0193)	0.4999 (0.0062)	0.5000 (0.0061)
	$\hat{\sigma}^2$	0.4966 (0.0701)	0.4971 (0.0580)	0.4998 (0.0224)	0.4997 (0.0182)	0.4999 (0.0071)	0.4998 (0.0058)
	$\hat{\omega}_n^2$	0.7491 (0.1065)	0.6210 (0.1273)	0.7493 (0.0340)	0.6252 (0.0400)	0.7497 (0.0107)	0.6247 (0.0125)
	$\hat{a}_{HT}$	0.4905 (0.0878)	0.4930 (0.0707)	0.4995 (0.0273)	0.4998 (0.0223)	0.4998 (0.0088)	0.4998 (0.0071)
	$\hat{\sigma}^2_{HT}$	0.4942 (0.0698)	0.4963 (0.0579)	0.4996 (0.0224)	0.4996 (0.0182)	0.4997 (0.0071)	0.5001 (0.0057)
	$\hat{\omega}^2_{HT}$	0.7604 (0.1092)	0.6283 (0.1290)	0.7505 (0.0341)	0.6259 (0.0400)	0.7500 (0.0105)	0.6248 (0.0126)
$\tau^2 = 0.9$	$\hat{a}$	0.4961 (0.0624)	0.4968 (0.0619)	0.4996 (0.0194)	0.4996 (0.0194)	0.5000 (0.0061)	0.5000 (0.0062)
	$\hat{\sigma}^2$	0.0994 (0.0141)	0.0993 (0.0116)	0.0998 (0.0044)	0.0997 (0.0037)	0.0998 (0.0014)	0.0998 (0.0011)
	$\hat{\omega}_n^2$	0.9429 (0.1325)	0.9222 (0.1845)	0.0997 (0.0425)	0.9231 (0.0586)	0.9502 (0.0133)	0.9249 (0.0183)
	$\hat{a}_{HT}$	0.4894 (0.0877)	0.4957 (0.0716)	0.4990 (0.0273)	0.4992 (0.0223)	0.5001 (0.0087)	0.5000 (0.0070)
	$\hat{\sigma}^2_{HT}$	0.0989 (0.0140)	0.0991 (0.0116)	0.0998 (0.0044)	0.0997 (0.0037)	0.0999 (0.0014)	0.0999 (0.0012)
	$\hat{\omega}^2_{HT}$	0.9575 (0.1364)	0.9326 (0.1874)	0.9518 (0.0427)	0.9242 (0.0587)	0.9498 (0.0134)	0.9249 (0.0188)

Table A.1: Simulation results obtained from the parameter estimators $\hat{\theta}_{n,T}$ and $\hat{\theta}_a$ in intercorrelated panels of AR(1)-processes of various sizes, where $\tau^2 = 0$, $\tau^2 = 0.5$, $\tau^2 = 0.9$ ($a = 0.5$).

		$nT = 200$		$nT = 2{,}000$		$nT = 20{,}000$	
		$n=2$	$n=4$	$n=2$	$n=4$	$n=2$	$n=4$
$\text{eff}_{rel}(\hat{a}, \hat{a}_{HT})$	$\tau^2 = 0$	0.5079	0.7576	0.4956	0.7466	0.4964	0.7178
	$\tau^2 = 0.5$	0.4890	0.7616	0.4895	0.7490	0.4964	0.7381
	$\tau^2 = 0.9$	0.5063	0.7474	0.5050	0.7568	0.4916	0.7845

Table A.2: Empirical relative efficiencies $\text{eff}_{rel}(\hat{a}, \hat{a}_{HT})$ of the estimators $\hat{\theta}_{n,T}$ and $\hat{\theta}_a$ in intercorrelated panels of AR(1)-processes of various sizes, where $\tau^2 = 0$, $\tau^2 = 0.5$, $\tau^2 = 0.9$ ($a = 0.5$).

- $\tau^2 = 0.9$, i.e. $\sigma^2 = 0.1$. Then $\omega_n^2 = 0.95$ ($n = 2$) or $\omega_n^2 = 0.925$ ($n = 4$).

The simulation results are displayed in table A.1. It can be seen that the ICM parameter estimator $\hat{\theta}_{n,T}$ performs equally well if the panel consists of independent time series or if they are intercorrelated. It is obvious that the variance of the estimators decreases for $nT \to \infty$. We further can read off this table that the estimators of $\hat{a}$ and $\hat{a}_{HT}$ are not affected by the strength of intercorrelation. The standard deviation of the variance estimators however changes depending on the true variances. This corresponds to the theorems 2.5.20 and 2.5.35, where we have derived the asymptotic distributions of the ICM parameter estimator $\hat{\theta}_{n,T}$ and $\hat{\theta}_a$. The theoretical asymptotic variance of $\hat{a}$ only depends on n via the sample size. In our case it is $0.75/\left(n\,(T-p)\right)$, which corresponds well to the simulated values.
The theoretical asymptotic variance of $\hat{\sigma}^2$ is $2\,(n-1)\,\sigma_0^4/\left(n^2\,(T-p)\right)$, which is also the asymptotic variance of $\hat{\sigma}^2_{HT}$. The asymptotic variance of $\hat{\omega}_n^2$ is $\frac{2}{T-p}\,\omega_n^4$. For $\hat{\sigma}^2$ we get for $n = 2$ that $\text{sd}\,(\hat{\sigma}^2) = 0.707\,\sigma_0^2/\sqrt{T-p}$ and for $n = 4$ that $\text{sd}\,(\hat{\sigma}^2) = 0.612\,\sigma_0^2/\sqrt{T-p}$. Again, the simulated values correspond well to the theoretical ones. We can see from the simulations that the standard deviations of $\hat{\sigma}^2$ and $\hat{\sigma}^2_{HT}$ and of $\hat{\omega}_n^2$ and $\hat{\omega}^2_{HT}$ are

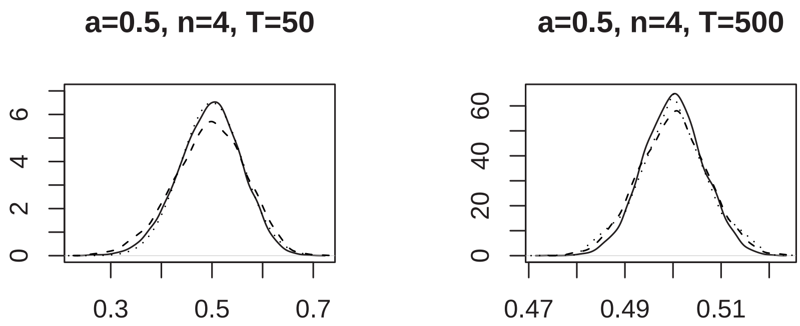

Figure A.1: Empirical densities for $\hat{a}$ (solid lines), $\hat{a}_{HT}$ (dashed lines) and $\text{N}(0.5, \sqrt{0.75}/\sqrt{nT})$, which is the theoretical density of $\hat{a}$, (dotted lines) in intercorrelated panels of AR(1)-processes with true values $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$ ($n = 4, T = 50$ and $n = 4, T = 500$).

compatible, which is also the case for their bias. Finally, the results for $nT = 200$ illustrate that the ICM parameter estimator $\hat{a}$ has a smaller bias than $\hat{a}_{HT}$, which is seen more clearly if the time series are short.
Table A.2 shows the empirical asymptotic efficiency for each model. As the theoretical relative efficiency is given by $\text{eff}_{rel}(\hat{a}, \hat{a}_{HT}) = \frac{n-1}{n}$ (remark 2.5.36), the theoretical values are 0.5 for $n = 2$ and 0.75 for $n = 4$. The simulations come close to these values. The values become less exact for $nT = 20,000$, which is mostly due to rounding effects.
In figure A.1 we see the empirical densities of $\hat{a}$ (solid line) and $\hat{a}_{HT}$ (dotted line) for the two cases $n = 4, T = 50$ and $n = 4, T = 5000$, which again illustrate the higher relative efficiency of $\hat{a}$.

A.2 Increasing Length of the Time Series

Now we investigate the properties of the estimators $\hat{a}$ and $\hat{a}_{HT}$ dependent on the length of the time series. Here $a = 0.5$ and $\sigma^2 = \tau^2 = 0.5$ are fixed. We regard one small and one large panel and various values of the time series length T:

- $n = 3$, thus $\omega_n^2 = 2/3$. The length of the time series increases from $T = 8$ to $T = 500$.

- $n = 128$, i.e. $\omega_n^2 = 0.504$. T increases from $T = 2$ to $T = 100$.

It is obvious that the variance of the estimators decreases substantially for T growing. If the number of time series in the panel is small ($n = 3$, see table A.3), one can again see that the ICM estimator is more efficient than $\hat{a}_{HT}$, in that the ratio of the variances

	$T=8$	$T=16$	$T=32$	$T=64$	$T=125$	$T=250$	$T=500$
$\hat{a}$	0.4623 (0.1989)	0.4823 (0.1332)	0.4921 (0.0912)	0.4947 (0.0628)	0.4976 (0.0453)	0.4988 (0.0318)	0.4993 (0.0220)
$\hat{\sigma}^2$	0.4776 (0.1883)	0.4876 (0.1259)	0.4943 (0.0904)	0.4979 (0.0633)	0.4990 (0.0451)	0.4994 (0.0314)	0.4993 (0.0223)
$\hat{\omega}_n^2$	0.6524 (0.3652)	0.6610 (0.2438)	0.6585 (0.1703)	0.6633 (0.1218)	0.6642 (0.0870)	0.6654 (0.0615)	0.6660 (0.0417)
$\hat{a}_{HT}$	0.4407 (0.2368)	0.4726 (0.1612)	0.4866 (0.1112)	0.4920 (0.0767)	0.4970 (0.0557)	0.4983 (0.0391)	0.4988 (0.0270)
$\hat{\sigma}^2_{HT}$	0.4664 (0.1836)	0.4823 (0.1246)	0.4917 (0.0900)	0.4967 (0.0632)	0.4983 (0.0451)	0.4991 (0.0314)	0.4991 (0.0223)
$\hat{\omega}^2_{HT}$	0.7326 (0.4394)	0.6986 (0.2665)	0.6766 (0.1775)	0.6719 (0.1240)	0.6687 (0.0880)	0.6676 (0.0619)	0.6670 (0.0418)
eff_{rel}	0.7055	0.6828	0.6726	0.6704	0.6614	0.6614	0.6639

Table A.3: Simulation results for the parameter estimators $\hat{\theta}_{n,T}$ and $\hat{\theta}_a$ in intercorrelated panels of AR(1)-processes for T increasing ($a = 0.5, n = 3, \sigma^2 = \tau^2 = 0.5$) last row: empirical relative efficiency $\text{eff}_{rel}(\hat{a}, \hat{a}_{HT})$.

	$T = 2$	$T = 10$	$T = 100$
$\hat{a}$	0.5012 (0.0830)	0.4992 (0.0261)	0.4998 (0.0077)
$\hat{\sigma}^2$	0.4956 (0.0621)	0.4991 (0.0207)	0.4998 (0.0062)
$\hat{\omega}_n^2$	0.5169 (0.7435)	0.4990 (0.2367)	0.5050 (0.0729)

	$T = 2$	$T = 10$	$T = 100$
$\hat{a}_{HT}$	0.5015 (0.0769)	0.4992 (0.0262)	0.4998 (0.0078)
$\hat{\sigma}_{HT}^2$	0.4950 (0.0619)	0.4991 (0.0207)	0.4998 (0.0062)
$\hat{\omega}_{HT}^2$	0.5244 (0.7438)	0.4997 (0.2370)	0.5051 (0.0729)

Table A.4: Simulation results for the parameter estimators $\hat{\theta}_{n,T}$ and $\hat{\theta}_a$ in intercorrelated panels of AR(1)-processes for T increasing ($a = 0.5, n = 128, \sigma^2 = \tau^2 = 0.5$).

tends to $2/3$ (see table A.3). However, for n large ($n = 128$), there is virtually no difference between the two estimators. This can be directly read off table A.4.

A.3 AR(6) Process

The process Hjellvik and Tjøstheim (1999a) use for investigating the effects of the intercorrelation more closely for two estimators of theirs is the AR(6) process with $a = (1, -0.6, 0.2, -0.2, 0, 0.4)$. Hjellvik and Tjøstheim (1999a) fix $T = 100$ and regard $n = 3, 4, 5$ and $\rho = \frac{\tau^2}{\sigma^2+\tau^2} = 1/(n-1)$. Furthermore they investigate for $n = 4$ also the cases $\rho = 0.5$ and $\rho = 0.25$.
The values for σ^2, τ^2 and ω_n^2 are in those cases

- $\sigma^2 = \tau^2 = 0.5$, $\omega_n^2 = 2/3$
- $\sigma^2 = 1/3$, $\tau^2 = 2/3$, $\omega_n^2 = 0.75$
- $\sigma^2 = 0.25$, $\tau^2 = 0.75$, $\omega_n^2 = 0.8125$
- $\sigma^2 = \tau^2 = 0.5$, $\omega_n^2 = 0.625$
- $\sigma^2 = 0.25$, $\tau^2 = 0.75$, $\omega_n^2 = 0.8125$

The tables A.5 and A.6 display the results for the ICM estimator $\hat{\theta}_{n,T}$ and the estimator of Hjellvik and Tjøstheim, $\hat{\theta}_a$, respectively. For $\hat{a}$ and $\hat{a}_{HT}$ we give the mean-square error in order to facilitate the comparison. As $T = 100$, the bias is small in both cases. The ratio of the mean squared errors (see table A.6) again illustrates that the ICM parameter estimator has a higher relative efficiency, and that this does not depend on the strength of the intercorrelation. Both estimators perform well.

A.4 Summary

As we have seen above, the simulation results are close to the true asymptotic values given in theorems 2.5.20 (ICM) and 2.5.34 (GICM), even when the time series are rather short. Moreover the simulations show that $\hat{a}$ has an smaller bias than $\hat{a}_{HT}$ if $n\,T$

	$n = 3$, $\tau^2 = 0.5$	$n = 4$, $\tau^2 = 1/3$	$n = 5$, $\tau^2 = 0.25$	$n = 4$, $\tau^2 = 0.5$	$n = 4$, $\tau^2 = 0.25$
$\hat{a}_1$	0.9980	0.9974	0.9980	0.9999	0.9988
$\hat{a}_2$	-0.6035	-0.6005	-0.6017	-0.6039	-0.6024
$\hat{a}_3$	0.2032	0.2005	0.2012	0.2022	0.2015
$\hat{a}_4$	-0.2060	-0.2035	-0.2019	-0.2034	-0.2039
$\hat{a}_5$	0.0043	0.0024	0.0005	0.0025	0.0031
$\hat{a}_6$	0.3895	0.3932	0.3950	0.3922	0.3925
MSE($\hat{a}$)	0.0353	0.0257	0.0210	0.0262	0.0259
$\hat{\sigma}^2$	0.4894 (0.0517)	0.3286 (0.0278)	0.2469 (0.0181)	0.4922 (0.0419)	0.2464 (0.0212)
$\hat{\omega}_n^2$	0.6532 (0.0981)	0.7381 (0.1089)	0.7935 (0.1163)	0.6162 (0.0905)	0.7993 (0.1194)

Table A.5: Behaviour of $\hat{\theta}_{n,T}$ in an intercorrelated panel of AR(6) processes with true parameter: $a = (1, -0.6, 0.2, -0.2, 0, 0.4)'$ ($T = 100$, intercorrelation and variances varying with n).

	$n = 3$, $\tau^2 = 0.5$	$n = 4$, $\tau^2 = 1/3$	$n = 5$, $\tau^2 = 0.25$	$n = 4$, $\tau^2 = 0.5$	$n = 4$, $\tau^2 = 0.25$
$\hat{a}_{HT,1}$	0.9966	0.9974	0.9977	0.9992	0.9980
$\hat{a}_{HT,2}$	-0.6053	-0.6019	-0.6024	-0.6040	-0.6024
$\hat{a}_{HT,3}$	0.2047	0.2013	0.2018	0.2024	0.2017
$\hat{a}_{HT,4}$	-0.2088	-0.2053	-0.2030	-0.2043	-0.2050
$\hat{a}_{HT,5}$	0.0059	0.0040	0.0013	0.0031	0.0039
$\hat{a}_{HT,6}$	0.3840	0.3900	0.3933	0.3900	0.3903
MSE($\hat{a}_{HT}$)	0.0533	0.03484	0.0261	0.0344	0.0350
$\hat{\sigma}^2_{HT}$	0.4841 (0.0511)	0.3269 (0.0277)	0.2462 (0.0180)	0.4896 (0.0416)	0.2451 (0.0211)
$\hat{\omega}^2_{HT}$	0.6939 (0.1107)	0.7680 (0.1158)	0.8175 (0.1208)	0.6410 (0.0962)	0.8323 (0.1273)
$\frac{MSE(\hat{a})}{MSE(\hat{a}_{HT})}$	0.6623	0.7385	0.8046	0.7616	0.7400

Table A.6: Behaviour of $\hat{\theta}_a$ in an intercorrelated panel of AR(6) processes with true parameter $a = (1, -0.6, 0.2, -0.2, 0, 0.4)'$ ($T = 100$, intercorrelation and variances varying with n). The last row displays the relative efficiency of $hata$ compared to $\hat{a}_H T$.

is small. If $nT \to \infty$, the differences between the two estimators $\hat{\theta}_{n,T} = (\hat{a}', \hat{\sigma}^2, \hat{\omega}_n^2)'$ and $\hat{\theta}_a = (\hat{a}_{HT}, \hat{\sigma}^2_{HT}, \hat{\omega}^2_{HT})$ vanish. For small n and if T is not very large, however, the higher relative efficiency of $\hat{a}$ compared to $\hat{a}_{HT}$ becomes important. In the last section we have shown that even in the case of a higher order autoregressive process both estimators perform well.

In practice, both estimators are feasible. Although the ICM parameter estimator $\hat{\theta}_{n,T}$ is calculated using an iterative algorithm, convergence is usually attained after 6 to 7 iterations. As the computational speed is high, this does not have any practical implications when the sample sizes are as investigated above.

Appendix B

Simulation Study (Robust Estimators)

We now study the performance of the above described robust estimators in simulations. In order to make a comparison of the estimators possible, we employ a fixed set of models. However the focus may vary in the different sections according to the specific properties of the estimators, so we sometimes include further simulations or do not display the full set of results. The chapter concludes with a comparative evaluation of the estimators regarded. Before starting with the comparisons, we give a brief summary of the estimators investigated and the main models used for the simulations.

Parameter Estimators

The various parameter estimators discussed in this thesis are summarised in table B.1.

Simulations

As models we choose intercorrelated AR(1) and AR(6) models (see table B.2). We always let $\sigma^2 = \tau^2 = 0.5$, i.e. the variance of the innovations is $\mathrm{var}\left(a_\theta(\mathrm{L})\, X_t^{(i)}\right) = \sigma^2 + \tau^2 = 1$ for all $t \in \mathbb{Z}$, $i = 1, \ldots, n$. The choice of the AR(6) model is as in Hjellvik and Tjøstheim (1999a). They let $a = (1, -0.6, 0.2, -0.2, 0, 0.4)'$ (see the Appendix A). These models are denoted by M_1 and M_6.
Moreover we investigate the performance of the estimators under contamination (see assumption 3.2.1). In the case of entire time series outlying, the outliers are independent AR(1) processes with parameters $a_{out} = 0.9$ and $\sigma^2_{out} = 1$ and Gaussian white noise processes with variance 1. The corresponding models, where two time series in the panel are replaced by the outlying time series, are called TS_1 and TS_n. For generating arbitrary outliers, we employ independent normally distributed random variables $V_t^{(i)} \sim \mathrm{N}(0, \sigma_V^2)$ where $\sigma_V^2 = 9$. The Bernoulli panel is such that $\mathbb{P}(\delta_{1,t,i} = 1) = 0.1$, i.e. the proportion of outliers is approximately 10%. Such models are for example regarded in Ma and Genton (2000). We denote the latter models by AO_1 and AO_6. An overview on these models is given in table B.2.
Each simulated panel consists of $n = 10$ time series of length $T = 50$ in the AR(1) case and $T = 100$ for AR(6) processes. Thus the true value of ω_n^2 is $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n} = 0.55$ and in the case of entire time series outlying, the ω_n^2 estimated only from the uncontaminated data would be $\omega_n^2 = 0.625$.

Non-robust parameter estimators described in the first chapter:	
$\hat{\theta}_{n,T}$	ICM parameter estimator (section 2.4.2)
$\hat{\theta}_a$, $\hat{\theta}_b$	GICM estimators (section 2.4.3)
Robust version of these estimators, with modifications (section 3.3):	
$\hat{\theta}_{rob}$	algorithm given in section 3.3
$\hat{\theta}_{oa}$	similar to $\hat{\theta}_{rob}$; variance determined using the overall median
$\hat{\theta}_{rw}$	reweighted version
$\hat{\theta}_{rw_2}$	reweighted version also allowing for arbitrary outliers
Bootstrap approximations (section 3.3):	
$\hat{\theta}_{RB}$	residual bootstrap
$\hat{\theta}_{PB}$	residual bootstrap adapted for panels
$\hat{\theta}_{NB}$	sampling from normal distributions
Covariance estimators (section 3.5):	
$\hat{\theta}_Q$	covariance matrix derived from the robust scale estimator $Q_{n,T}$
$\hat{\theta}_{MCD}$	minimum covariance determinant method
Robust regression (section 3.6):	
$\hat{\theta}_M$	M-estimator
$\hat{\theta}_{LTS}$	least trimmed squares estimator
Preliminary outlier detection (section 3.7):	
$\hat{\theta}_{LR}$	non-robust likelihood ratio test
$\hat{\theta}_{PS}$	non-robust phase space method
$\hat{\theta}_{PS;rob}$	robust phase space method based on the MCD
$\hat{\theta}_{PS;rec}$	phase space method, iterative elimination of the outliers

Table B.1: Overview of the parameter estimators compared in the simulation study.

Basic models:	
M_1	AR(1), $a = 0.5$, $n = 10$, $T = 50$, $\sigma^2 = \tau^2 = 0.5$
M_6	AR(6), $a = (1, -0.6, 0.2, -0.2, 0, 0.4)'$, $n = 10$, $T = 100$, $\sigma^2 = \tau^2 = 0.5$
Entire time series outlying:	
TS_1	M_1, two time series replaced by ind. AR(1) with $a_{out} = 0.9$, $\sigma^2_{out} = 1$
TS_n	M_1, two time series replaced by Gaussian WN with $\sigma^2_{out} = 1$
TS_6	M_6, two time series replaced by ind. AR(1) with $a_{out} = 0.9$, $\sigma^2_{out} = 1$
Arbitrary outliers:	
AO_1	M_1, 10% outliers, Gaussian WN with $\sigma^2_{out} = 9$
$AO_{1;100}$	M_1, 10% outliers, Gaussian WN with $\sigma^2_{out} = 100$
AO_6	M_6, 10% outliers, Gaussian WN with $\sigma^2_{out} = 9$

Table B.2: Models used for generating panels of intercorrelated time series.

If not otherwise stated, we compute the mean and standard deviation (in brackets) over 5000 iterations. Exceptions are for example the bootstrap procedures which are not applied to every model and where we have to restrict ourselves to 100 iterations due to the high computation time.

B.1 Robustifying the ICM Parameter Estimator

We here compare the four robustifications of the parameter estimator, $\hat{\theta}_{rob}$, $\hat{\theta}_{oa}$, $\hat{\theta}_{rw}$ and $\hat{\theta}_{rw_2}$ which have been derived in the section 3.3. We investigate these estimators in various situations.

model		$\hat{\theta}_{n,T}$	$\hat{\theta}_{rob}$	$\hat{\theta}_{oa}$	$\hat{\theta}_{rw}$	$\hat{\theta}_{rw_2}$
M_1	$\hat{a}$	0.4984 (0.0394)	0.4661 (0.0511)	0.4664 (0.0521)	0.4978 (0.0403)	0.4625 (0.0463)
	$\hat{\sigma}^2$	0.4989 (0.0338)	0.5930 (0.0405)	0.4489 (0.0413)	0.4971 (0.0350)	0.4460 (0.0375)
	$\hat{\omega}_n^2$	0.5487 (0.1100)	0.5820 (0.1195)	0.5769 (0.1161)	0.5511 (0.1122)	0.5060 (0.1130)
TS_n	$\hat{a}$	0.3912 (0.0541)	0.4188 (0.0666)	0.4175 (0.0672)	0.4725 (0.0551)	0.4465 (0.0550)
	$\hat{\sigma}^2$	0.7252 (0.0594)	0.8561 (0.0722)	0.5596 (0.0563)	0.5446 (0.0787)	0.4880 (0.0582)
	$\hat{\omega}_n^2$	0.3875 (0.0779)	0.4618 (0.0967)	0.4610 (0.0970)	0.5187 (0.1279)	0.4699 (0.1186)
TS_1	$\hat{a}$	0.7219 (0.0651)	0.5018 (0.0542)	0.5015 (0.0559)	0.5074 (0.0552)	0.4830 (0.0536)
	$\hat{\sigma}^2$	0.7469 (0.0623)	1.0248 (0.1616)	0.5765 (0.0600)	0.5131 (0.0562)	0.4784 (0.0533)
	$\hat{\omega}_n^2$	0.4058 (0.0835)	0.5203 (0.1098)	0.5220 (0.1067)	0.5480 (0.1187)	0.5089 (0.1168)
AO_1	$\hat{a}$	0.2027 (0.0545)	0.2171 (0.0613)	0.1922 (0.0626)	0.2333 (0.0808)	0.3665 (0.0913)
	$\hat{\sigma}^2$	1.5046 (0.2392)	1.7802 (0.2876)	0.5919 (0.0654)	1.2909 (0.3486)	0.5225 (0.1164)
	$\hat{\omega}_n^2$	0.6030 (0.1316)	0.6161 (0.1318)	0.6296 (0.1439)	0.6319 (0.1831)	0.5013 (0.1473)

Table B.3: Simulation results for the robustified ICM parameter estimators $\big(\hat{\theta}_{rob}$, $\hat{\theta}_{oa}$, $\hat{\theta}_{rw}$, $\hat{\theta}_{rw_2}$; $\hat{\theta}_{n,T}$ included for reference$\big)$ in an intercorrelated panel of AR(1) processes ($n = 10$, $T = 50$). True parameters: $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$, $\omega_n^2 = 0.55$.

From table B.3 we see that the non-robust estimator performs well if there are no outliers, but in presence of outliers it is much influenced by these. In this case it is advisable to use a robust estimator. Here the reweighted procedure ($\hat{\theta}_{rw}$) performs uniformly best as long as there are no arbitrary outliers present (AO_1). In the uncontaminated case its behaviour is similar to that of the non-robust estimator. This is not surprising as the reweighting method coincides with the non-robust estimation if no time series are identified as outlying. The estimator $\hat{\theta}_{rob}$ underestimates a, but overestimates σ^2. This is due to the method chosen for estimating the variance σ^2. We use the robust scale estimator employed in the least median of squares procedure, which is known to overestimate the variance. However it remains bounded in situations where the variance obtained from the least squares procedure explodes (see Rousseeuw and Leroy 1987, p. 212). Estimating the variance with the overall median leads to the estimator θ_{oa}, which underestimates a but which prevents $\hat{\sigma}^2$ from exploding. As the method has been developed for estimation in the presence of entire time series outlying and is based on a transformation involving the median process, it is not suited if arbitrary outliers may occur. Only $\hat{\theta}_{rw_2}$ is constructed for coping with this kind of outliers, as it allows to eliminate single outlying time points as well as entire time series from the estimation procedure. In the other cases the latter estimator performs in an acceptable way. It is a better estimator than both $\hat{\theta}_{rob}$ and $\hat{\theta}_{oa}$, but is clearly outperformed by $\hat{\theta}_{rw}$ if the only kind of outliers occuring are entire time series outlying. In practice it is preferable to identify and eliminate the arbitrary outliers in the single time series in a first step, which

model		$\hat{\theta}_{n,T}$	$\hat{\theta}_{rob}$	$\hat{\theta}_{oa}$	$\hat{\theta}_{rw}$	$\hat{\theta}_{rw_2}$
M_6	$\hat{a}_1$	0.9994	0.7759	0.7782	0.9989	0.6664
	$\hat{a}_2$	-0.6004	-0.2454	-0.2474	-0.5999	-0.1529
	$\hat{a}_3$	0.1993	-0.0682	-0.0638	0.1997	-0.1416
	$\hat{a}_4$	-0.2001	-0.1558	-0.1518	-0.2002	-0.1506
	$\hat{a}_5$	0.0001	0.0341	0.0294	-0.0003	-0.0103
	$\hat{a}_6$	0.3971	0.3896	0.3947	0.3979	0.4104
	MSE($\hat{a}$)	0.0105	0.4268	0.4716	0.0103	0.4540
	$\hat{\sigma}^2$	0.4968	2.5514	0.9518	0.4963	1.0781
		(0.0244)	(0.4516)	(0.1845)	(0.0239)	(0.2582)
	$\hat{\omega}_n^2$	0.5470	1.0202	1.0581	0.5462	0.7913
		(0.0802)	(0.2184)	(0.2400)	(0.0806)	(0.2564)
TS_6	$\hat{a}_1$	1.0254	1.1359	1.2445	1.0197	0.7602
	$\hat{a}_2$	-0.5237	-0.5998	-0.7174	-0.5336	-0.1896
	$\hat{a}_3$	0.1031	0.2137	0.2969	0.1147	-0.1104
	$\hat{a}_4$	-0.1386	0.0295	0.0789	-0.1469	-0.1247
	$\hat{a}_5$	0.0457	-0.0885	-0.1287	0.0382	0.0269
	$\hat{a}_6$	0.3540	0.5139	0.5454	0.3599	0.3988
	MSE($\hat{a}$)	0.0412	0.7844	1.2246	0.0367	0.3580
	$\hat{\sigma}^2$	0.7975	5.6531	2.2776	0.7488	1.2279
		(0.0519)	(3.5547)	(1.6206)	(0.0982)	(0.2585)
	$\hat{\omega}_n^2$	0.3934	1.4185	1.7648	0.4161	0.6159
		(0.0577)	(0.7901)	(1.2226)	(0.0800)	(0.2451)

Table B.4: Simulation results for the robustified ICM parameter estimators $\big(\hat{\theta}_{rob}$, $\hat{\theta}_{oa}$, $\hat{\theta}_{rw}$, $\hat{\theta}_{rw_2}$; $\hat{\theta}_{n,T}$ included for reference$\big)$ in an intercorrelated panel of AR(6) processes ($n = 10$, $T = 100$). True parameters: $\sigma^2 = \tau^2 = 0.5$, $\omega_n^2 = 0.55$.

is possible (Rousseeuw and Leroy 1987, Gather, Bauer and Fried 2002).
The simulations for the AR(6) case displayed in table B.4 again illustrate the effects seen in the AR(1) case. $\hat{\theta}_{rob}$, $\hat{\theta}_{oa}$ and also $\hat{\theta}_{rw_2}$ become more unreliable. However the values obtained indicate that $\hat{\theta}_{rw}$ seems not to detect the outlying time series, as it performs very similar to the the non-robust estimator. This perhaps can be overcome by adjusting the tuning factor used in the estimation according to the order of the underlying process.
In order to see how the behaviour of the estimators changes when n or T become large, we simulate the models M_1 and TS_1 with $n = 10$, $T = 100$ and $n = 100, T = 50$. In the latter case, we furthermore replace 20 out of the 100 time series by samples from independent AR(1) processes with parameter $a_{out} = 0.9$ and the innovations' variance $\sigma^2_{out} = 1$. The results displayed in table B.5 confirm the properties of the estimators shown in table B.3. Again, $\hat{\theta}_{n,T}$ is drawn to the parameter of the outliers and $\hat{\theta}_{rw}$ performs best. The variances of the estimators decrease if n or T increase. The simulations show that the increase in T has a stronger relative effect than n increasing. This is due the the estimation procedure. After subtracting the median process, which is better estimated if n is large, a is estimated from the transformed processes using a modified least squares method. Thus the estimate improves with T growing. The same can be seen from the model TS_1. Here the proportion of outliers is 20% in each case. The variance of the estimators improves with n growing. But the larger absolute number of outliers in the case of $n = 100$ leads to a higher bias. However, the variance estimators $\hat{\sigma}^2$ and $\hat{\omega}^2_n$ are improved for both $\hat{\theta}_{rob}$ and $\hat{\theta}_{oa}$ although they still are biased. Finally, it seems that if a larger absolute number of outliers are present, they are not all detected by $\hat{\theta}_{rw}$, resulting in a slightly worse estimate for $n = 100$, $T = 50$ than in the case of fewer observations. $\hat{\theta}_{rw_2}$ is instable for the higher order autoregressive process.
We complete this section considering the behaviour of the GICM estimators. In the first chapter we have seen that the non-robust ICM parameter estimator has a smaller bias and is asymptotically more efficient than the GICM estimator if n is not too large (remark 2.5.36, see also the simulations in the Appendix A). We now investigate whether this effect is also visible in the robust estimators. As example we use $\hat{\theta}_{rob}$. The simulation results are displayed in table B.6. Indeed the estimates of $\hat{\theta}_{rob}$ are better than those by $\hat{\theta}_{rob;a}$ in the case of the uncontaminated model M_1 and for TS_1. In these cases the bias of $\hat{a}$ is smaller and the estimator is more efficient. As the estimator $\hat{\theta}_{rob;b}$ is only based on a single time series of length $T = 50$, the median process, its variance is higher. Nevertheless, the estimate of a in the model M_1 is comparable to the one obtained from $\hat{\theta}_{rob;a}$. For TS_1 the estimate of the variance is close to the true one, whereas $\hat{\theta}_{rob;a}$ overestimates σ^2. In the case of AO_1 however the results for $\hat{\theta}_{rob;b}$ are comparable to those in the uncontaminated model M_1, since taking the median compensates for the influence of the arbitrary outliers.

B.1.1 Improvement by Bootstrap Procedures

Most of the above robustified versions of the original ICM parameter estimator are biased. Thus we now investigate whether a bootstrap procedure can be used for assessing the empirical bias and thus for improving the estimator. As example we take the estimator $\hat{\theta}_{rob}$ which underestimates a and exhibits a large bias in $\hat{\sigma}^2$. The empirical bias

model			$\hat{\theta}_{n,T}$	$\hat{\theta}_{rob}$	$\hat{\theta}_{oa}$	$\hat{\theta}_{rw}$	$\hat{\theta}_{rw_2}$
M_1	$n = 10$, $T = 100$	$\hat{a}$	0.4994 (0.0273)	0.4742 (0.0360)	0.4740 (0.0373)	0.4991 (0.0278)	0.4641 (0.0325)
		$\hat{\sigma}^2$	0.4994 (0.0237)	0.5924 (0.0284)	0.4485 (0.0297)	0.4991 (0.0236)	0.4477 (0.0263)
		$\hat{\omega}_n^2$	0.5479 (0.0783)	0.5780 (0.0817)	0.5806 (0.0831)	0.5509 (0.0783)	0.4991 (0.0775)
	$n = 100$, $T = 50$	$\hat{a}$	0.4999 (0.0124)	0.4811 (0.0201)	0.4813 (0.0199)	0.4997 (0.0127)	0.4747 (0.0154)
		$\hat{\sigma}^2$	0.4997 (0.0101)	0.5098 (0.0101)	0.4570 (0.0115)	0.4994 (0.0103)	0.4640 (0.0118)
		$\hat{\omega}_n^2$	0.5072 (0.1027)	0.5096 (0.1033)	0.5121 (0.1042)	0.5074 (0.1007)	0.4593 (0.1016)
TS_1	$n = 10$, $T = 100$	$\hat{a}$	0.7259 (0.0474)	0.5015 (0.0379)	0.5018 (0.0389)	0.5010 (0.0340)	0.4789 (0.0354)
		$\hat{\sigma}^2$	0.7483 (0.0441)	1.0299 (0.1179)	0.5737 (0.0417)	0.5008 (0.0325)	0.4688 (0.0329)
		$\hat{\omega}_n^2$	0.4072 (0.0602)	0.5201 (0.0772)	0.5224 (0.0749)	0.5578 (0.0807)	0.5123 (0.0812)
	$n = 100$, $T = 50$	$\hat{a}$	0.7459 (0.0219)	0.5116 (0.0219)	0.5129 (0.0218)	0.5249 (0.0240)	0.5000 (0.0209)
		$\hat{\sigma}^2$	0.7436 (0.0262)	0.8638 (0.0462)	0.5802 (0.0174)	0.5257 (0.0223)	0.5000 (0.0207)
		$\hat{\omega}_n^2$	0.3531 (0.0745)	0.4365 (0.0887)	0.4380 (0.0900)	0.4828 (0.1056)	0.4462 (0.1018)

Table B.5: Large sample behaviour of the robustified ICM parameter estimators $\big(\hat{\theta}_{rob}$, $\hat{\theta}_{oa}$, $\hat{\theta}_{rw}$, $\hat{\theta}_{rw_2}$; $\hat{\theta}_{n,T}$ included for reference$\big)$ in an intercorrelated panel of AR(1) processes ($n = 10, T = 100$ and $n = 100, T = 50$). True parameters: $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$, $\omega_n^2 = 0.55$ ($n = 10$), $\omega_n^2 = 0.505$ ($n = 100$). TS_1: 2 ($n = 10$) and 20 ($n = 100$) time series replaced by independent AR(1) processes with $a_{out} = 0.9$, $\sigma_{out}^2 = 1$.

obtained by the bootstrap procedure is then used to calculate the factor by which the parameter estimates are adjusted. As described in section 3.3, we compare three methods: a residual bootstrap for autoregressions (RB), a modification thereof, where the structure of the panel is conserved (PB), and sampling from normal distributions (NB). Because of the computation time of the bootstrap procedure, the empirical mean and variance are calculated only from 100 samples. Preliminary simulations have shown that both bootstrap estimates remain stable if the bootstrap is iterated $n_{sim} = 300$ times.

Table B.7 displays the estimated means and standard deviations (in brackets) of $\hat{\theta}_{rob}$ and the improved versions using the bootstrap procedures. The panel bootstrap ($\hat{\theta}_{PB}$) reflects the properties of $\hat{\theta}_{rob}$. In the estimator $\hat{\theta}_{RB}$, $\hat{a}$ is increased, but also slightly $\hat{\sigma}^2$, whereas $\hat{\omega}_n^2$ is downweighted. The estimator $\hat{\theta}_{NB}$ increases $\hat{a}$ and downweights $\hat{\sigma}^2$. In the uncontaminated case it leads to satisfying results. In the presence of outliers, this estimator however also is biased. Nevertheless it gives the best results of these four estimators. For arbitrary outliers all estimators are not valid.

Theoretically, $\hat{\theta}_{PB}$ should perform better that $\hat{\theta}_{RB}$ in the case of intercorrelated time series. However in the ICM procedure the intercorrelation does not have a strong effect, since the correlation of the residual processes is σ^2/n, where n is the number of time series in the panel. The better performance of $\hat{\theta}_{RB}$ could also be due to the fact that the set which is used for sampling only consists of T residual vectors in the case of $\hat{\theta}_{PB}$, whereas we sample from nT individual residuals in the first case. Thus we moreover evaluate the behaviour of the estimators for larger panels. From table B.8 we see that again $\hat{\theta}_{RB}$ outperforms $\hat{\theta}_{PB}$, which shows that the intercorrelation does not play a strong role. $\hat{\theta}_{PB}$ still has a tendency to downweight the estimate of a. Altogether, $\hat{\theta}_{NB}$ is again preferable to $\hat{\theta}_{PB}$ and $\hat{\theta}_{RB}$.

model		$\hat{\theta}_{rob;a}$	$\hat{\theta}_{rob;b}$
M_1	$\hat{a}$	0.4637 (0.0562)	0.4657 (0.1263)
	var	0.5924 (0.0404)	0.5675 (0.1162)
TS_1	$\hat{a}$	0.5022 (0.0594)	0.4739 (0.1265)
	var	1.0313 (0.1640)	0.5112 (0.1060)
AO_1	$\hat{a}$	0.1784 (0.0649)	0.4509 (0.1291)
	var	1.7803 (0.2938)	0.5631 (0.1157)

Table B.6: Comparison of the GICM estimators $\hat{\theta}_{rob;a}$ and $\hat{\theta}_{rob;b}$ obtained from the residual processes and the median process. The panels are formed of intercorrelated AR(1) processes with $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$, $n = 10$, $T = 50$. The true values for the variances are $\text{var}_a = \sigma^2 = 0.5$ and $\text{var}_b = \omega_n^2 = 0.55$.

B.2 Robust Autocovariances

B.2.1 The Robust Panel Autocovariance Estimator $\hat{\gamma}_{n,T}$

First we study the behaviour of $\hat{\gamma}_{n,T}$ as autocovariance estimator. The performance of the resulting parameter estimator is discussed in more detail in the next subsection. We compare the estimator $\hat{\gamma}_{n,T}$ derived from the robust scale estimator $Q_{n,T}$ with its modification $\hat{\gamma}^d_{n,T}$ obtained from $Q^d_{n,T}$, where only the differences of time points whose distance is at least $0.1 \times T$ are taken into account (see remark 3.4.5). The model used for the comparison is an AR(1) process with parameter $a = 0.5$ and variance $\sigma^2 = 1$. For these preliminary considerations, we compute mean and standard deviation (given in brackets) over 1,000 iterations.
Table B.9 shows that for $T = 100$ and $T =$ 1,000 the behaviour of the two estimators is similar but that $\hat{\gamma}^d_{n,T}$ is slightly less efficient. In the case of small T, the estimator $\hat{\gamma}^d(0)$ obtained from $Q^d_{n,T}$ is less biased. However here the variance of $\hat{\gamma}^d_{n,T}$ is very large. It

model		$\hat{\theta}_{rob}$	$\hat{\theta}_{RB}$	$\hat{\theta}_{PB}$	$\hat{\theta}_{NB}$
M_1	$\hat{a}$	0.4661 (0.0511)	0.5067 (0.0537)	0.4718 (0.0574)	0.4954 (0.0549)
	$\hat{\sigma}^2$	0.5930 (0.0405)	0.6178 (0.0405)	0.5954 (0.0406)	0.5101 (0.0317)
	$\hat{\omega}^2$	0.5820 (0.1195)	0.5228 (0.1184)	0.5891 (0.1268)	0.5446 (0.1290)
TS_n	$\hat{a}$	0.4188 (0.0666)	0.4422 (0.0751)	0.4037 (0.0754)	0.4577 (0.0706)
	$\hat{\sigma}^2$	0.8561 (0.0722)	0.8938 (0.0651)	0.8512 (0.0603)	0.7025 (0.0533)
	$\hat{\omega}^2$	0.4618 (0.0967)	0.3874 (0.0797)	0.4530 (0.1060)	0.4132 (0.0983)
TS_1	$\hat{a}$	0.5018 (0.0542)	0.5320 (0.0583)	0.4870 (0.0595)	0.5324 (0.0638)
	$\hat{\sigma}^2$	1.0248 (0.1616)	1.0796 (0.1670)	1.0531 (0.1860)	0.8483 (0.1490)
	$\hat{\omega}^2$	0.5203 (0.1098)	0.4433 (0.1138)	0.5201 (0.1194)	0.4651 (0.1083)
AO_1	$\hat{a}$	0.2171 (0.0613)	0.2332 (0.0586)	0.2139 (0.0619)	0.2439 (0.0654)
	$\hat{\sigma}^2$	1.7802 (0.2876)	1.8254 (0.3029)	1.7882 (0.3382)	1.5800 (0.2961)
	$\hat{\omega}^2$	0.6161 (0.1318)	0.5167 (0.1201)	0.6184 (0.1420)	0.5587 (0.1360)

Table B.7: Comparison of the bootstrap estimators $\big(\hat{\theta}_{RB}, \hat{\theta}_{PB}, \hat{\theta}_{NB}$; $\hat{\theta}_{rob}$ included for reference$\big)$ in an intercorrelated panel of AR(1) processes ($n = 10, T = 50$). True parameters: $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$, $\omega^2_n = 0.55$, $n = 10$, $T = 50$.

leads to estimates of a which are more biased than those obtained from $Q_{n,T}$.

In the next step we investigate how these estimators perform in the panel case. For better comparison we start with a panel of independent time series and estimate the autocovariances directly without the preliminary transformation used in the ICM-type estimation. Comparing the simulation results of table B.10 with the case of a single time series treated above, we can see that introducing more time series into the panel reduces the variance of the estimator appreciably. However the autocovariance estimators are even more biased downwards. The bias in the estimation of the autocovariance function might be due to the choice of the quantile in the panel case, which is $\lfloor \frac{n\binom{T}{2}+2}{4} \rfloor + 1$. This problem can probably be overcome by approaching the 1/4 quantile from above for $nT \to \infty$. The effect that for $\hat{\gamma}_{n,T}$ the bias of $\hat{a}$ is much smaller than that of the autocovariance estimators themselves can be due to the underlying scale estimator. This

model			$\hat{\theta}_{rob}$	$\hat{\theta}_{RB}$	$\hat{\theta}_{PB}$	$\hat{\theta}_{NB}$
M_1	$n = 10$, $T = 100$	$\hat{a}$	0.4742 (0.0360)	0.4977 (0.0347)	0.4884 (0.0387)	0.5012 (0.0387)
		$\hat{\sigma}^2$	0.5924 (0.0284)	0.6189 (0.0314)	0.5966 (0.0333)	0.4971 (0.0202)
		$\hat{\omega}_n^2$	0.5780 (0.0817)	0.5089 (0.0835)	0.5835 (0.0880)	0.5525 (0.0936)
	$n = 100$, $T = 50$	$\hat{a}$	0.4811 (0.0201)	0.5009 (0.0184)	0.4794 (0.0208)	0.5007 (0.0193)
		$\hat{\sigma}^2$	0.5098 (0.0101)	0.5118 (0.0105)	0.5090 (0.0095)	0.5001 (0.0109)
		$\hat{\omega}_n^2$	0.5096 (0.1033)	0.5104 (0.0909)	0.5223 (0.1065)	0.4946 (0.0941)
TS_1	$n = 10$, $T = 100$	$\hat{a}$	0.5015 (0.0379)	0.5198 (0.0408)	0.4977 (0.0395)	0.5341 (0.0383)
		$\hat{\sigma}^2$	1.0299 (0.1179)	1.0622 (0.1116)	1.0566 (0.1300)	0.8296 (0.0913)
		$\hat{\omega}_n^2$	0.5201 (0.0772)	0.4229 (0.0655)	0.5455 (0.0906)	0.4606 (0.0614)
	$n = 100$, $T = 50$	$\hat{a}$	0.5116 (0.0219)	0.5343 (0.0216)	0.4838 (0.0256)	0.5335 (0.0211)
		$\hat{\sigma}^2$	0.8638 (0.0462)	0.8581 (0.0461)	0.8772 (0.0430)	0.8451 (0.0435)
		$\hat{\omega}_n^2$	0.4365 (0.0887)	0.4342 (0.0874)	0.4565 (0.0908)	0.4313 (0.0841)

Table B.8: Large sample behaviour of the bootstrap estimators $\big(\hat{\theta}_{RB}$, $\hat{\theta}_{PB}$, $\hat{\theta}_{NB}$; $\hat{\theta}_{rob}$ included for reference$\big)$ in an intercorrelated panel of AR(1) processes. True parameters: $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$, $\omega_n^2 = 0.55$ ($n = 10$), $\omega_n^2 = 0.505$ ($n = 100$). TS_1: 2 ($n = 10$) and 20 ($n = 100$) time series replaced by independent AR(1) processes with $a_{out} = 0.9$, $\sigma_{out}^2 = 1$.

has been standardised by Rousseeuw and Leroy (1987) by an empirical constant, which has not been specifically adapted to the panel case. In the last two cases the estimation is based on 49,000 and 49,500 differences ($n = 40$, $T = 50$ and $n = 10$, $T = 100$),

	$T = 20$	$T = 50$	$T = 100$	$T = 1,000$
$\hat{\gamma}(0)$	1.0119 (0.4896)	1.1620 (0.3345)	1.2284 (0.2459)	1.3238 (0.0796)
$\hat{\gamma}(1)$	0.5091 (0.4693)	0.5876 (0.3106)	0.6247 (0.2221)	0.6622 (0.0724)
$\hat{\gamma}(2)$	0.1302 (0.3432)	0.2262 (0.2530)	0.2809 (0.1877)	0.3264 (0.0640)
$\hat{a}$	0.4602 (0.3476)	0.4847 (0.1681)	0.4978 (0.1032)	0.4991 (0.0317)
$\hat{\gamma}^d(0)$	1.2347 (0.8054)	1.2421 (0.4020)	1.2764 (0.2610)	1.3229 (0.0850)
$\hat{\gamma}^d(1)$	0.9605 (1.0922)	0.7254 (0.4351)	0.6921 (0.2493)	0.6653 (0.0765)
$\hat{\gamma}^d(2)$	0.5336 (0.9142)	0.3584 (0.3643)	0.3414 (0.2258)	0.3325 (0.0668)
$\hat{a}^d$	0.6804 (0.4960)	0.5525 (0.1986)	0.5304 (0.1134)	0.5016 (0.0315)

Table B.9: Comparison of the robust autocovariance estimators $\hat{\gamma}_{n,T}$ and $\hat{\gamma}^d_{n,T}$. The model is a single AR(1) process with parameters $a = 0.5$ and $\sigma^2 = 1$. T varies from $T = 20$ to $T = 1,000$. The true values of the autocovariance function are $\gamma(0) = 1.333$, $\gamma(1) = 0.667$ and $\gamma(2) = 0.333$.

	$\hat{\gamma}_{n,T}$			$\hat{\gamma}^d_{n,T}$		
	$n = 50$, $T = 20$	$n = 40$, $T = 50$	$n = 10$, $T = 100$	$n = 50$, $T = 20$	$n = 40$, $T = 50$	$n = 10$, $T = 100$
$\hat{\gamma}(0)$	0.8319 (0.0570)	1.0975 (0.0508)	1.2060 (0.0773)	0.9402 (0.0770)	1.1583 (0.0594)	1.2401 (0.0834)
$\hat{\gamma}(1)$	0.3626 (0.0532)	0.5299 (0.0469)	0.5949 (0.0695)	0.5735 (0.0924)	0.6351 (0.0600)	0.6519 (0.0793)
$\hat{\gamma}(2)$	0.0626 (0.0370)	0.1949 (0.0367)	0.2603 (0.0566)	0.2142 (0.0686)	0.2879 (0.0490)	0.3134 (0.0679)
$\hat{a}$	0.4350 (0.0492)	0.4823 (0.0277)	0.4922 (0.0346)	0.6081 (0.0688)	0.5475 (0.0320)	0.5244 (0.0378)

Table B.10: Comparison of the robust autocovariance estimators $\hat{\gamma}_{n,T}$ and $\hat{\gamma}^d_{n,T}$ in the panel situation for various choices of n and T. The model is a panel of independent AR(1) processes with parameters $a = 0.5$ and $\sigma^2 = 1$. The true values of the autocovariance function are $\gamma(0) = 1.333$, $\gamma(1) = 0.667$, $\gamma(2) = 0.333$.

so these sizes are comparable. This comparison again shows that a large number of short time series leads to a higher bias than a smaller number of longer time series. The two estimators exhibit the same differences as in the single time series case: $\hat{\gamma}_{n,T}$ is relatively more efficient than $\hat{\gamma}^d_{n,T}$ and its estimates of a are more accurate, whereas the latter estimator is less biased.
The case of panels of intercorrelated time series and the behaviour of the panel autocovariance estimators in the presence of outliers are included in the next subsection. For all further investigations we concentrate on the estimator $\hat{\gamma}_{n,T}$ derived from $Q_{n,T}$, as here the estimator of the autoregressive parameter, which is our main interest, is more accurate.

B.2.2 Comparison of $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$

Inserting robust autocovariance estimates for the entries of the covariance matrix, we obtain a robustified covariance matrix and a robust autocovariance vector which then can be used in the ICM or GICM procedure. As alternative to the robust autocovariance estimator $\hat{\gamma}_{n,T}$ derived from the robust scale estimator $Q_{n,T}$ we have chosen the covariance estimator obtained from the minimum covariance determinant method (see section 3.5). There the covariance matrix is estimated directly and is positive definite by construction. We compare the performance of the two estimators for our standard examples. In order to illustrate whether the calculation using the ICM method and the GICM procedure differ in the present case, we include the GICM estimators. The estimators are called $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$. The GICM versions thereof are $\hat{\theta}_{Q;a}$ and $\hat{\theta}_{MCD;a}$ (estimators obtained from the residual processes) and $\hat{\theta}_{Q;b}$ and $\hat{\theta}_{MCD;b}$ (obtained from the median process).
Table B.11 compares the behaviour of the estimators $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$ in intercorrelated panels of AR(1) processes of size $n = 10, T = 50$. Here $\mathring{\gamma}_n(0) = \frac{n-1}{n\,(1-a^2)}\,\sigma^2 = 0.6$ and $\bar{\gamma}_n(0) = \frac{1}{1-a^2}\,\omega_n^2 = 0.733$. We have seen above that $\hat{\theta}_Q$ is biased downwards for these sizes of n and T. The same effect is true for the estimator $\hat{\theta}_{MCD}$ derived from the MCD procedure. The estimators perform comparably to the robustified ICM estimators $\hat{\theta}_{rob}$ and $\hat{\theta}_{oa}$ of section B.1; the estimators of the variances are similar to those of $\hat{\theta}_{oa}$. Both methods are however much more stable than these in the presence of arbitrary outliers. In this case the estimates of a are much closer to the true value and the variances are smaller than those of the latter estimators. Although the two methods tend to underestimate $\mathring{\gamma}_n(0)$ as we can see from M_1, the bias is positive in the extreme case of arbitrary outliers with variance 100. This illustrates that here the estimators, being derived from robust scale estimators, are more attracted by the variance of the outliers than by the value of the autoregressive parameter.
The results for the larger sample sizes ($n = 40$, $T = 50$ and $n = 10, T = 100$) are displayed in table B.12. For $n = 40$ the true variances are $\mathring{\gamma}_n(0) = 0.65$ and $\bar{\gamma}_n(0) = 0.683$. As the sample sizes are large, we here compute the mean and standard deviation (in brackets) over 1,000 samples instead of 5,000. The estimators still are biased, $\hat{\theta}_{MCD}$ as well as $\hat{\theta}_Q$. As observed above for the robust panel autocovariance estimator $\hat{\gamma}_{n,T}$ itself, we can see that $\hat{\theta}_Q$ is, the sample size remaining constant, more biased in the case of a large number of short time series than if the time series are longer. $\hat{\theta}_{MCD}$ is improved in the case of $n = 40$, $T = 50$ in comparison to $n = 10$,

model		$\hat{\theta}_Q$	$\hat{\theta}_{MCD}$
M_1	$\hat{a}$	0.4759 (0.0572)	0.4351 (0.0640)
	$\hat{\mathring{\gamma}}_n(0)$	0.4607 (0.0467)	0.5016 (0.0557)
	$\hat{\bar{\gamma}}_n(0)$	0.6581 (0.1874)	0.5934 (0.1845)
TS_n	$\hat{a}$	0.4266 (0.0598)	0.3934 (0.0679)
	$\hat{\mathring{\gamma}}_n(0)$	0.5451 (0.0562)	0.5938 (0.0655)
	$\hat{\bar{\gamma}}_n(0)$	0.5026 (0.1472)	0.4525 (0.1451)
TS_1	$\hat{a}$	0.5196 (0.0568)	0.4759 (0.0640)
	$\hat{\mathring{\gamma}}_n(0)$	0.6044 (0.0652)	0.6289 (0.0780)
	$\hat{\bar{\gamma}}_n(0)$	0.5887 (0.1694)	0.5335 (0.1666)
AO_1	$\hat{a}$	0.4394 (0.0674)	0.3944 (0.0626)
	$\hat{\mathring{\gamma}}_n(0)$	0.6120 (0.0673)	0.5917 (0.0665)
	$\hat{\bar{\gamma}}_n(0)$	0.6394 (0.1819)	0.5743 (0.1785)
$AO_{1;100}$	$\hat{a}$	0.5621 (0.0831)	0.4205 (0.0563)
	$\hat{\mathring{\gamma}}_n(0)$	0.6919 (0.0843)	0.5912 (0.0653)
	$\hat{\bar{\gamma}}_n(0)$	0.6686 (0.1899)	0.5997 (0.1846)

Table B.11: Simulation results for $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$ in an intercorrelated panel of AR(1) processes ($n = 10$, $T = 50$). True parameters: $a = 0.5$, $\mathring{\gamma}_n(0) = 0.6$, $\bar{\gamma}_n(0) = 0.733$.

model		$\hat{\theta}_Q$		$\hat{\theta}_{MCD}$	
		$n = 40$, $T = 50$	$n = 10$, $T = 100$	$n = 40$, $T = 50$	$n = 10$, $T = 100$
M_1	$\hat{a}$	0.4752	0.4897	0.4794	0.4389
		(0.0276)	(0.0373)	(0.0324)	(0.0445)
	$\hat{\mathring{\gamma}}_n(0)$	0.5270	0.5063	0.5789	0.5048
		(0.0251)	(0.0346)	(0.0298)	(0.0392)
	$\hat{\bar{\gamma}}_n(0)$	0.6105	0.7065	0.5409	0.6301
		(0.1788)	(0.1392)	(0.1673)	(0.1388)

Table B.12: Behaviour of $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$ in an intercorrelated panel of AR(1) processes ($n = 40, T = 50$ and $n = 10, T = 100$). True parameters: $a = 0.5$ and $\sigma^2 = \tau^2 = 0.5$. The true values for the autocovariance function are $\mathring{\gamma}_n(0) = 0.6$ and $\bar{\gamma}_n(0) = 0.733$ for $n = 10$ and $\mathring{\gamma}_n(0) = 0.65$ and $\bar{\gamma}_n(0) = 0.683$ for $n = 40$.

$T = 100$. This is due to the fact that the minimum covariance determinant estimator is not based on the number of differences. It directly depends on the sample size nT, and this is twice as large in the first case than in the second one. We omit the results for the contaminated models as there only the effects already discussed are illustrated again.

Next we compare the performance of the two estimators in a panel of AR(6) processes. We already have mentioned above that the matrices $\hat{\theta}_Q$ is based on are not necessarily positive definite. Thus the estimator does not give reliable results if the autoregressive processes are of higher order. This can be seen from table B.13. $\hat{\theta}_{MCD}$ performs comparably to the robustified ICM estimators $\hat{\theta}_{rob}$ and $\hat{\theta}_{oa}$ (see table B.4), with a slightly

	$\hat{\theta}_Q$	$\hat{\theta}_{MCD}$		
	M_6	M_6	TS_6	AO_6
$\hat{a}_1$	6.6044	0.7293	0.7578	0.5386
$\hat{a}_2$	-7.9862	-0.2119	-0.2152	-0.0609
$\hat{a}_3$	1.3502	-0.1023	-0.0996	-0.1933
$\hat{a}_4$	8.0162	-0.1723	-0.1527	-0.1971
$\hat{a}_5$	-8.3292	0.0329	0.0402	0.0114
$\hat{a}_6$	3.2800	0.3799	0.3741	0.3647
MSE($\hat{a}$)	10^6	0.3745	0.3535	0.6853
$\hat{\mathring{\gamma}}_n(0)$	3.9843	5.3759	5.7173	5.6200
	(0.9561)	(1.4334)	(1.4627)	(1.2810)
$\hat{\bar{\gamma}}_n(0)$	7.2453	6.8111	4.9995	6.1022
	(5.0582)	(5.0960)	(4.0681)	(4.6112)

Table B.13: Performance of $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$ in an intercorrelated panel of AR(6) processes ($n = 10$, $T = 100$). True parameters: $a = (1, 0.6, -0.2, 0.2, 0, 0.4)'$, $\gamma_n(0) = 5.747$ and $\bar{\gamma}_n = 7.025$.

model		$\hat{\theta}_{Q;a}$	$\hat{\theta}_{Q;b}$	$\hat{\theta}_{MCD;a}$	$\hat{\theta}_{MCD;b}$
M_1	$\hat{a}$	0.4786 (0.0594)	0.4696 (0.1699)	0.4388 (0.0659)	0.4376 (0.2206)
	var	0.4616 (0.0462)	0.6570 (0.1869)	0.5022 (0.0570)	0.5962 (0.1847)
TS_1	$\hat{a}$	0.5236 (0.0600)	0.4761 (0.1708)	0.4791 (0.0652)	0.4397 (0.2272)
	var	0.6029 (0.0656)	0.5912 (0.1709)	0.6295 (0.0786)	0.5349 (0.1677)
AO_1	$\hat{a}$	0.4365 (0.0723)	0.0298 (0.1877)	0.3932 (0.0644)	0.4219 (0.2311)
	var	0.6085 (0.0666)	0.5187 (0.1230)	0.5934 (0.0652)	0.5780 (0.1795)

Table B.14: Comparison of the GICM estimators obtained from $\hat{\gamma}_{n,T}$ and the MCD procedure, $\hat{\theta}_{Q;a}$, $\hat{\theta}_{Q;b}$, $\hat{\theta}_{MCD;a}$ and $\hat{\theta}_{MCD;b}$. The panels are formed of intercorrelated AR(1) processes ($n = 10$, $T = 50$) with $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$. The true values for the variances are $\gamma_a(0) = 0.6$ and $\gamma_b(0) = 0.733$.

smaller mean squared error. It estimates $\mathring{\gamma}_n(0)$ closely, however with a variance which is much larger than that of the estimator obtained from $\hat{\theta}_Q$. As $\bar{\gamma}_n(0)$ is estimated from a single time series, the sample upon which the estimators are based is much smaller than the one used for estimating $\mathring{\gamma}_n(0)$. This leads to the large variance of the estimators, which is more visible here than in the AR(1) case.
Finally, we evaluate the corresponding GICM estimators. The simulation results are displayed in table B.14. We can see that there is not much difference in the performance of $\hat{\theta}_{Q;a}$ and $\hat{\theta}_{MCD;a}$ to the ICM-type estimators $\hat{\theta}_Q$ and $\hat{\theta}_{MCD}$. In the case of arbitrary outliers, $\hat{\theta}_{Q;a}$ and $\hat{\theta}_{MCD;a}$ perform well themselves. The result for $\hat{\theta}_{Q;b}$ in the AO_1 case is however surprising as the estimators $\hat{\theta}_{Q;b}$ and $\hat{\theta}_{MCD;b}$, being derived from the median process, should be more stable against arbitrary outliers than $\hat{\theta}_{Q;a}$ and $\hat{\theta}_{MCD;a}$. This effect has already been addressed for the robustified GICM parameter estimator $\hat{\theta}_{rob;b}$ in section B.1. A possible explanation is the small sample size which facilitates the implosion of $\hat{\gamma}_{n,T}$.

B.2.3 Robust Regression Methods

We now present the results from the robust regression methods as discussed in section 3.6. There it has moreover been explained that we here only get estimators of the GICM type or can perform a direct estimation without transforming the data. The latter procedure is not adapted for the case of entire time series outlying. This means we obtain in a panel fulfilling the assumptions 2.2.1 of the ICM two separate M-estimators $\hat{\theta}_{M;a} = (\hat{a}', \hat{\sigma}_n^2)'$ of $(a', \sigma_n^2)'$ and $\hat{\theta}_{M;b} = (\hat{a}_b', \hat{\omega}_n^2)'$ of $(a', \omega_n^2)'$. Analogously we get for the two LTS estimators $\hat{\theta}_{LTS;a}$ and $\hat{\theta}_{LTS;b}$. The variances of the transformed time series are $\sigma_n^2 = \frac{n-1}{n}\sigma^2 = 0.5$ and $\omega_n^2 = 0.55$ if the variances in the panel of intercorrelated time series are chosen to be $\sigma^2 = \tau^2 = 0.5$. The estimators obtained from the direct procedure are called $\hat{\theta}_{M;dir}$ and $\hat{\theta}_{LTS;dir}$. There the variance for which we obtain an

model		GICM				*dir.*	
		$\hat{\theta}_{M;a}$	$\hat{\theta}_{M;b}$	$\hat{\theta}_{LTS;a}$	$\hat{\theta}_{LTS;b}$	$\hat{\theta}_{M;dir}$	$\hat{\theta}_{LTS;dir}$
M_1	$\hat{a}$	0.4683 (0.0422)	0.4659 (0.1307)	0.2689 (0.1264)	0.4723 (0.3068)	0.4948 (0.0709)	0.4931 (0.1433)
	var	0.5335 (0.0369)	0.5687 (0.1174)	0.4353 (0.0478)	0.5225 (0.1560)	1.0969 (0.1263)	1.0189 (0.1338)
TS_n	$\hat{a}$	0.3771 (0.0501)	0.4423 (0.1311)	0.2276 (0.1161)	0.4335 (0.3128)	0.4197 (0.0708)	0.4261 (0.1522)
	var	0.7645 (0.0634)	0.4523 (0.0967)	0.5327 (0.0597)	0.4082 (0.1251)	1.1421 (0.1117)	1.0554 (0.1239)
TS_1	$\hat{a}$	0.6863 (0.0707)	0.4734 (0.1298)	0.3653 (0.1604)	0.4801 (0.3180)	0.6825 (0.0653)	0.6505 (0.1481)
	var	0.8205 (0.0728)	0.5127 (0.1089)	0.5490 (0.0631)	0.4715 (0.1409)	1.1897 (0.1258)	1.1026 (0.1377)
AO_1	$\hat{a}$	0.1896 (0.0528)	0.4499 (0.1284)	0.1619 (0.1164)	0.4476 (0.3167)	0.2926 (0.0740)	0.4156 (0.1657)
	var	1.6067 (0.2613)	0.5632 (0.1172)	0.5470 (0.0646)	0.5188 (0.1523)	2.1506 (0.2952)	1.2574 (0.1739)

Table B.15: Comparison of the M- and LTS-estimators (GICM and direct procedure). The model is an intercorrelated panel of AR(1) processes ($n = 10$, $T = 50$) with $a = 0.5$ and $\sigma^2 = \tau^2 = 0.5$. The true values of the variances are $\sigma_n^2 = 0.5$ and $\omega_n^2 = 0.55$.

estimate is $\text{var} = \sigma^2 + \tau^2 = 1$.

Table B.15 displays the results obtained from the simulations. $\hat{\theta}_{LTS;a}$ is not valid at this sample size. The parameter estimator $\hat{\theta}_{LTS;b}$ yields acceptable average values for the autoregressive parameters and downward biased ones for the variances, but the variance is very large. Thus the estimates are not reliable. Only $\hat{\theta}_{LTS;dir}$ yields estimates which are comparable to those of the M-estimators. However it is less efficient than $\hat{\theta}_{M;dir}$. The latter estimator performs best in the uncontaminated case. If outliers are present, it is influenced by these. Comparing $\hat{\theta}_{M;a}$ and $\hat{\theta}_{M;b}$ we see that $\hat{\theta}_{M;b}$, which is based on the median process, is less efficient but more robust than $\hat{\theta}_{M;a}$. This means that taking the median levels out the differences induced by another time series model or by arbitrary outliers. The effect has already been observed in section B.1 for the robustified ICM parameter estimators.

The large sample behaviour of the estimators is shown in table B.16. With the sample size increasing the variance of the estimators is reduced for $\hat{\theta}_M$ and $\hat{\theta}_{LTS;a}$. For $\hat{\theta}_{LTS;b}$ is only decreases noticeably in the case of T growing, not if only n becomes larger. Thus here again the results are not reliable due to the large variance. $\hat{\theta}_{LTS;a}$ is still not valid for $n = 10$, $T = 100$. For $n = 100$, $T = 50$ the estimates improve although they are biased. $\hat{\theta}_{LTS;dir}$ again yields estimates comparable to those of $\hat{\theta}_{M;dir}$, but it is still less efficient. Similar to $\hat{\theta}_{LTS;b}$, $\hat{\theta}_{M;b}$ becomes more efficient if T is increased, whereas the

model		GICM				*dir.*	
		$\hat{\theta}_{M;a}$	$\hat{\theta}_{M;b}$	$\hat{\theta}_{LTS;a}$	$\hat{\theta}_{LTS;b}$	$\hat{\theta}_{M;dir}$	$\hat{\theta}_{LTS;dir}$
M_1; $n = 10$, $T = 100$	$\hat{a}$	0.4687 (0.0297)	0.4765 (0.0888)	0.2525 (0.0895)	0.4736 (0.2367)	0.4975 (0.0483)	0.4969 (0.1066)
	var	0.5327 (0.0257)	0.5758 (0.0831)	0.4372 (0.0341)	0.5305 (0.1061)	1.1037 (0.0893)	1.0167 (0.0932)
M_1; $n = 100$, $T = 50$	$\hat{a}$	0.4961 (0.0128)	0.4777 (0.1271)	0.4740 (0.0478)	0.4771 (0.3090)	0.4952 (0.0610)	0.4949 (0.0849)
	var	0.5041 (0.0103)	0.4997 (0.1016)	0.4567 (0.0120)	0.4644 (0.1380)	0.9967 (0.1029)	0.9189 (0.1006)
TS_1; $n = 10$, $T = 100$	$\hat{a}$	0.6904 (0.0517)	0.4823 (0.0923)	0.3521 (0.1237)	0.4857 (0.2417)	0.6901 (0.0480)	0.6633 (0.1120)
	var	0.8227 (0.0518)	0.5163 (0.0751)	0.5529 (0.0438)	0.4746 (0.0972)	1.2000 (0.0886)	1.0973 (0.0941)
TS_1; $n = 100$, $T = 50$	$\hat{a}$	0.7140 (0.0229)	0.4771 (0.1299)	0.5687 (0.0582)	0.4806 (0.3096)	0.6921 (0.0327)	0.6604 (0.0734)
	var	0.7609 (0.0280)	0.4271 (0.0890)	0.5756 (0.0174)	0.3958 (0.1188)	1.0889 (0.0912)	0.9984 (0.0916)

Table B.16: Comparison of the M- and LTS-estimators (GICM and direct procedure). The model is an intercorrelated panel of AR(1) processes ($n\,T$ large) with $a = 0.5$ and $\sigma^2 = \tau^2 = 0.5$. The true values of the variances are $\sigma_n^2 = 0.45$ and $\omega_n^2 = 0.55$ for the GICM procedure and $\sigma^2 + \tau^2 = 1$ for the direct estimation.

effect is only weak if n is growing. The estimator $\hat{\theta}_{M;a}$ is improved in both cases, but here the effect is stronger for n increasing. In the case of the contaminated panels it is again much influenced by the outlying time series. $\hat{\theta}_{M;b}$ however remains stable under contamination. For the investigated models it is thus the best estimator in this section.
We have included the simulation results for the non-contaminated AR(6) case (GICM procedure) in table B.17. Here we can observe that both the M-estimators and $\hat{\theta}_{LTS}$ are not stable if the order grows. Thus we have omitted displaying the results for the contaminated AR(6) models TS_6 and AO_6. The performance of the estimators is comparable to that of $\hat{\theta}_{oa}$ described in section B.1. For $\hat{\theta}_M$ the mean squared error is smaller, for $\hat{\theta}_{LTS}$ larger than that of $\hat{\theta}_{oa}$. In table B.18 we see the simulation results obtained from the direct estimation procedure. In the uncontaminated case both $\hat{\theta}_{M;dir}$ and $\hat{\theta}_{LTS;dir}$ give good estimates. Here the mean squared error of the autoregressive parameter estimator belonging to the latter one is smaller. If entire time series are outlying, both estimators still perform better than e.g. the robustified ICM parameter estimators. The variances are here more biased than in the uncontaminated case. Only in the presence of arbitrary outliers the estimators become clearly unreliable. In this case the variance estimate of $\hat{\theta}_{LTS;dir}$ is less biased than that of $\hat{\theta}_{M;dir}$. Altogether this indicates that $\hat{\theta}_{LTS}$ improves if the autoregressive order of the time series in the panel increases. The result that the estimates in the case of entire time series outlying are not as much influenced by the outliers as in the AR(1) case is surprising and probably due

to the actually chosen models. This effect can also be observed in the behaviour of the non-robust ICM parameter estimator $\hat{\theta}_{n,T}$ itself (see section B.1 and the discussion in section B.4).

B.3 Outlier Detection

B.3.1 Likelihood Ratio Test

Finally we consider the methods based on a preliminary identification of outliers discussed in section 3.7. Table B.19 displays the results of the likelihood ratio procedure.

	$\hat{\theta}_{M;a}$	$\hat{\theta}_{M;b}$	$\hat{\theta}_{LTS;a}$	$\hat{\theta}_{LTS;b}$
$\hat{a}_1$	0.7336	0.7401	0.7199	0.7389
$\hat{a}_2$	-0.2094	-0.2379	-0.2065	-0.2333
$\hat{a}_3$	-0.1014	-0.0841	-0.0954	-0.0876
$\hat{a}_4$	-0.1733	-0.1704	-0.1783	-0.1758
$\hat{a}_5$	0.0398	0.0335	0.0488	0.0370
$\hat{a}_6$	0.3796	0.3557	0.3646	0.3536
MSE($\hat{a}$)	0.3590	0.3796	0.5266	0.8625
var	0.9352 (0.1520)	0.9026 (0.1797)	0.9297 (0.1598)	0.8435 (0.2423)

Table B.17: Comparison of the M- and LTS-estimators (GICM procedure). The model is an intercorrelated panel of AR(6) processes ($n = 10$, $T = 100$) with $a = (1, 0.6, -0.2, 0.2, 0, 0.4)'$ and $\sigma^2 = \tau^2 = 0.5$. The true values of the variances are $\sigma_n^2 = 0.45$ and $\omega_n^2 = 0.55$.

	$\hat{\theta}_{M;dir}$			$\hat{\theta}_{LTS;dir}$		
	M_6	TS_6	AO_6	M_6	TS_6	AO_6
$\hat{a}_1$	0.9983	1.0208	0.5863	0.9967	1.0119	0.8702
$\hat{a}_2$	-0.6034	-0.5478	-0.0892	-0.5977	-0.5600	-0.3417
$\hat{a}_3$	0.2014	0.1355	-0.1942	0.1958	0.1542	-0.0663
$\hat{a}_4$	-0.2040	-0.1569	-0.1709	-0.1992	-0.1664	-0.0815
$\hat{a}_5$	0.0027	0.0343	-0.0027	0.0022	0.0198	-0.0113
$\hat{a}_6$	0.3912	0.3676	0.3802	0.3919	0.3799	0.3926
MSE($\hat{a}$)	0.5986	0.0368	0.0320	0.2672	0.2597	0.3174
var	1.0872 (0.0886)	1.2278 (0.0890)	4.6637 (1.0009)	1.1092 (0.1148)	1.2176 (0.1197)	1.9723 (0.2735)

Table B.18: Comparison of the M- and LTS-estimators (direct estimation procedure). The model is an intercorrelated panel of AR(6) processes ($n = 10$, $T = 100$) with $a = (1, 0.6, -0.2, 0.2, 0, 0.4)'$ and $\sigma^2 = \tau^2 = 0.5$. The true value of the variances is $\text{var} = 1$.

The parameters are estimated by minimising the conditional log-likelihood function. We see that the results do not differ much between intercorrelated panels and panels of independent time series. The only effect is that the estimators are more efficient in the latter case. In the uncontaminated model or if only arbitrary outliers are present, the null hypothesis of homogeneity is not rejected and thus the estimates do not change. If the outliers are independent white noise processes, the estimation is improved by the preliminary detection and elimination of outliers. If however the outlying time series are autoregressive time series with parameter $a = 0.9$, the estimation is drawn towards their parameter.

We now investigate the behaviour of the homogeneity test and of the outlier identification procedure in more detail. The empirical rejection rates (H_1) and the number n_{OL} of eliminated time series are given in table B.20. As the maximal proportion has been fixed at 20% beforehand, it may occur that the null hypothesis of homogeneity is still rejected after terminating the estimation procedure. Note that we have to test at an adjusted significance level α_n as the elimination procedure involves multiple testing (see remark 3.7.5). For $\alpha = 0.01$ and $n = 10$, the adjusted level is $\alpha = 0.0033$. This is reached in the uncontaminated model with independent processes. If the time series in the panel are intercorrelated, the rate of rejection grows. In the case of $n = 100$, $T = 50$, where only the number of the time series in the panel has been increased, it

model		*corr.*		*ind.*	
		$\hat{\theta}_{LR;1}$	$\hat{\theta}_{LR;2}$	$\hat{\theta}_{LR;1}$	$\hat{\theta}_{LR;2}$
M_1	$\hat{a}$	0.4929 (0.0646)	0.4929 (0.0646)	0.4974 (0.0375)	0.4975 (0.0375)
	var	0.9907 (0.1200)	0.9908 (0.1201)	0.9968 (0.0661)	0.9968 (0.0661)
TS_n	$\hat{a}$	0.4169 (0.0672)	0.4451 (0.0788)	0.4196 (0.0451)	0.4520 (0.0532)
	var	1.0373 (0.1063)	1.0222 (0.1111)	1.0374 (0.0731)	1.0231 (0.0731)
TS_1	$\hat{a}$	0.6875 (0.0664)	0.7129 (0.0822)	0.6882 (0.0614)	0.7195 (0.0739)
	var	1.0774 (0.1202)	1.0941 (0.1254)	1.0778 (0.0815)	1.0923 (0.0915)
AO_1	$\hat{a}$	0.2559 (0.0690)	0.2572 (0.0709)	0.2591 (0.0580)	0.2601 (0.0601)
	var	1.9496 (0.2739)	1.9469 (0.2758)	1.9577 (0.2610)	1.956 (0.2635)

Table B.19: Estimates before ($\hat{\theta}_{LR;1}$) and after ($\hat{\theta}_{LR;2}$) a preliminary outlier detection using the likelihood ratio test. *corr.:* panels of intercorrelated autoregressive processes with parameters $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$. *ind.:* panels of independent autoregressive processes with parameters $a = 0.5$ and $\sigma^2 = 1$. The size of the panels is $n = 10$ and $T = 50$, var=1.

model		*corr.*		*ind.*	
		$\hat{\theta}_{LR;1}$	$\hat{\theta}_{LR;2}$	$\hat{\theta}_{LR;1}$	$\hat{\theta}_{LR;2}$
M_1	H_1	0.0072	0.0006	0.0032	0.0002
	n_{OL}	0.0316	0.0314	0.0498	0.0492
AO_1	H_1	0.0448	0.0012	0.0564	0.0008
	n_{OL}	0.0508	0.0352	0.0662	0.0446
M_1; $n = 10, T = 100$	H_1	0.0072	6e-04	0.0022	0.0000
	n_{OL}	0.0290	0.0290	0.0434	0.0430
M_1; $n = 100, T = 50$	H_1	0.1320	0.0704	0.0006	0.0002
	n_{OL}	0.0748	0.0696	0.1726	0.1724
TS_n	H_1	0.5012	0.0004	0.5800	0.0014
	n_{OL}	1.0722	0.6374	1.0974	0.5972
	p_f	0.0047	–	0.0061	–
TS_1	H_1	0.8408	0.6782	0.8926	0.7200
	n_{OL}	1.5582	1.1646	1.6958	1.1842
	p_f	0.9642	–	0.9814	–
TS_1; $n = 10, T = 100$	H_1	0.9950	0.8708	0.9978	0.9102
	n_{OL}	3.4122	2.7836	3.4552	2.7658
	p_f	0.9642	–	0.9389	–
TS_1; $n = 100, T = 50$	H_1	1.0000	0.9874	1.0000	0.9998
	n_{OL}	7.9232	2.6562	9.3844	1.8052
	p_f	0.9890	–	0.9814	–

Table B.20: Performance and power of the likelihood ratio test (at significance level $\alpha = 0.01$) in various models. *corr.:* panels of intercorrelated autoregressive processes with parameters $a = 0.5$ and $\sigma^2 = \tau^2 = 0.5$. *ind.:* panels of independent autoregressive processes with parameters $a = 0.5$ and $\sigma^2 = 1$. Unless stated otherwise, the size of the panel is $n = 10$, $T = 50$. H_1: empirical rate of rejections; n_{OL}: average number of identified outliers; p_f: proportion of falsely identified outliers among the eliminated outliers.

even reaches more than 13%. This is reduced after eliminating time series from the panel. If outlying time series are present, the rejection rate also grows. However the test only discovers that the panel is not homogeneous. In the model TS_n the null hypothesis of homogeneity is rejected in only 50% of the cases. If outliers are detected, they are mostly correctly identified. In the case of autoregressive time series with parameter $a = 0.9$, the outliers are not identified correctly. Indeed the proportion of falsely identified outliers is overproportionally large, which means that the test statistic has been drawn to the outliers. This illustrates the masking effect outliers can have in a non-robust procedure.

B.3.2 Phase Space Representations

For the phase space procedure we compare the three different versions of the estimator addressed in section 3.7. $\hat{\theta}_{PS;rob}$ is obtained from the robust method, which is based upon the robust covariance estimator derived from the MCD. $\hat{\theta}_{PS}$ itself is the non-robust version using the sample covariance matrix. In both cases, the outlying time series are detected in a first step and all are deleted from the sample. The estimation is then based on the remaining time series, where we use the non-robust ICM parameter estimator $\hat{\theta}_{n,T}$. We assume that not more than 20% of the time series are outlying. $\hat{\theta}_{PS;rec}$ is obtained from a (non-robust) recursive procedure. In each step we only eliminate the time series with the smallest p-value from the panel until no further outliers are identified or until the upper bound of 20% outlying time series is reached. This procedure is included for comparison to the likelihood ratio method of the last section. As the latter method originally is defined for panels of uncorrelated time series, we also investigate this case. The results for the parameter estimators are displayed in table B.21. As significance level we have chosen $\alpha = 0.01$. We can see that the estimators behave in a similar way as the estimator $\hat{\theta}_{LR}$ of the last section. There is no large difference between the estimation in the correlated and the uncorrelated model. In the case of arbitrary outliers the estimator is less biased if the panel consists of independent time series, nevertheless the bias is unacceptably large. If the outlying time series are white noise processes or in the presence of arbitrary outliers the estimators are downward biased, even more so than $\hat{\theta}_{LR}$. However they all perform better than $\hat{\theta}_{LR}$ if the outlying time series are autoregressive processes with a large coefficient. Here $\hat{\theta}_{PS;rec}$ is more influenced by the outlying time series than the other estimators. Table B.22 shows the large sample behaviour of the estimators. There is essentially no difference between the estimation in the correlated and in the uncorrelated panel. In the non-contaminated model the parameters become even more accurate. As we use the non-robust ICM estimator $\hat{\theta}_{n,T}$ in the second step after the outlier detection, the properties of the estimators are largely dependent on those of $\hat{\theta}_{n,T}$. In particular, this explains the respective efficiencies (compare section A.1 in the Appendix A). In the contaminated case, $\hat{\theta}_{PS;rob}$ is slightly more biased if only the number of time series in the panel is increased. This is due to the fact that the test detects more easily the two outlying time series in a panel consisting of 10 time series than all 20 outliers in model TS_1 with panel size $n = 100$ and $T = 50$. The efficiency of $\hat{\theta}_{PS}$ is much lower, and in the case of $n = 100$ and $T = 50$ also the bias is increased. This illustrates that the non-robust estimator is indeed affected if outliers are present. $\hat{\theta}_{PS;rec}$ is not improved if only the length of the time series grows. However it competes favourably with the other estimators if $n = 100$.

As the behaviour of the estimators depends on the number of identified and eliminated outliers, these are given in table B.23. If the model contains outliers, furthermore the proportion of wrongly identified outliers is displayed. $\hat{\theta}_{PS}$, being a non-robust estimator, is more susceptible to masking effects. Therefore it yields the highest proportion of falsely detected outliers. Nevertheless the masking effects are much weaker than for $\hat{\theta}_{LR}$, where the proportion of falsely identified outliers lies over 90%. Altogether the estimator $\hat{\theta}_{PS;rob}$ performs best. The comparison with the independent case shows that the intercorrelation even helps to correctly identify the outliers. If the outlying time

model		*corr.*			*ind.*
		$\hat{\theta}_{PS}$	$\hat{\theta}_{PS;rob}$	$\hat{\theta}_{PS;rec}$	$\hat{\theta}_{PS;rob}$
M_1	$\hat{a}$	0.4957 (0.0394)	0.4959 (0.0405)	0.4948 (0.0397)	0.4920 (0.0416)
	$\hat{\sigma}^2$	0.4955 (0.0343)	0.4957 (0.0347)	0.4959 (0.0338)	0.9848 (0.0738)
	$\hat{\omega}_n^2$	0.5441 (0.1120)	0.5482 (0.1120)	0.5469 (0.1126)	0.1032 (0.0221)
TS_n	$\hat{a}$	0.3886 (0.0547)	0.3923 (0.0554)	0.3916 (0.0564)	0.4105 (0.0489)
	$\hat{\sigma}^2$	0.7196 (0.0635)	0.7145 (0.0685)	0.7149 (0.0671)	1.0236 (0.0810)
	$\hat{\omega}_n^2$	0.3831 (0.0792)	0.3881 (0.0814)	0.3845 (0.0770)	0.1075 (0.0230)
TS_1	$\hat{a}$	0.5501 (0.1044)	0.5229 (0.0624)	0.6187 (0.0635)	0.5163 (0.0549)
	$\hat{\sigma}^2$	0.5814 (0.1180)	0.5424 (0.0770)	0.6352 (0.0558)	0.9939 (0.0809)
	$\hat{\omega}_n^2$	0.5174 (0.1330)	0.5182 (0.1123)	0.4652 (0.0945)	0.1224 (0.0268)
AO_1	$\hat{a}$	0.2190 (0.0622)	0.2279 (0.0632)	0.2180 (0.0596)	0.2824 (0.0649)
	$\hat{\sigma}^2$	1.4059 (0.2606)	1.3636 (0.2704)	1.4074 (0.2487)	1.8073 (0.2780)
	$\hat{\omega}_n^2$	0.6019 (0.1319)	0.5921 (0.1287)	0.5935 (0.1292)	0.1971 (0.0456)

Table B.21: Estimates after a preliminary outlier detection using the phase space representation. The size of the panels is $n = 10$ and $T = 50$. Significance level: $\alpha = 0.01$. *corr.:* panels of intercorrelated autoregressive processes with parameters $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$ and $\omega_n^2 = 0.55$. *ind.:* panels of independent autoregressive processes with parameters $a = 0.5$, $\sigma^2 = 1$ and $\omega_n^2 = 0.1$.

model			*corr.*			*ind.*
			$\hat{\theta}_{PS}$	$\hat{\theta}_{PS;rob}$	$\hat{\theta}_{PS;rec}$	$\hat{\theta}_{PS;rob}$
M_1	$n = 10$, $T = 100$	$\hat{a}$	0.4979 (0.0280)	0.4979 (0.0279)	0.4976 (0.0276)	0.4962 (0.0286)
		$\hat{\sigma}^2$	0.4985 (0.0243)	0.4977 (0.0243)	0.4963 (0.0242)	0.9923 (0.0519)
		$\hat{\omega}_n^2$	0.5480 (0.0793)	0.5491 (0.0787)	0.5476 (0.0783)	0.1041 (0.0165)
	$n = 100$, $T = 50$	$\hat{a}$	0.4980 (0.0126)	0.4985 (0.0125)	0.4973 (0.0125)	0.4959 (0.0127)
		$\hat{\sigma}^2$	0.4980 (0.0103)	0.4981 (0.0104)	0.4973 (0.0107)	0.9896 (0.0218)
		$\hat{\omega}_n^2$	0.5036 (0.1016)	0.5040 (0.1034)	0.5021 (0.1020)	0.0102 (0.0021)
TS_1	$n = 10$, $T = 100$	$\hat{a}$	0.5309 (0.2050)	0.5057 (0.0401)	0.6334 (0.0494)	0.5035 (0.0359)
		$\hat{\sigma}^2$	0.5523 (0.2178)	0.5107 (0.0460)	0.6422 (0.0387)	1.0000 (0.0575)
		$\hat{\omega}_n^2$	0.5648 (0.2287)	0.5519 (0.0813)	0.4703 (0.0678)	0.1270 (0.0206)
	$n = 100$, $T = 50$	$\hat{a}$	0.5898 (0.0285)	0.5376 (0.0302)	0.5256 (0.0247)	0.5235 (0.0174)
		$\hat{\sigma}^2$	0.6036 (0.0309)	0.5486 (0.0334)	0.5356 (0.0281)	0.9982 (0.0231)
		$\hat{\omega}_n^2$	0.4263 (0.0789)	0.4540 (0.0841)	0.4659 (0.0871)	0.0120 (0.0024)

Table B.22: Large sample behaviour ($n = 10, T = 100$ and $n = 100, T = 50$) of the estimates obtained after a preliminary outlier detection using the phase space representation. Significance level: $\alpha = 0.01$. *corr.:* panels of intercorrelated autoregressive processes with parameters $a = 0.5$, $\sigma^2 = \tau^2 = 0.5$ and $\omega_n^2 = 0.55$. *ind.:* panels of independent autoregressive processes with parameters $a = 0.5$, $\sigma^2 = 1$ and $\omega_n^2 = 0.1$.

series are white noise processes, in each case very few outliers are detected. In the case of autoregressive processes outlying, the proportion of falsely identified outliers decreases. If $n = 10$, $\hat{\theta}_{PS;rec}$ has a low detection rate but those time series which are detected as outliers are indeed outliers with a high probability. In the case of $n = 100$, $T = 50$, the proportion of falsely detected outliers is as high as for $\hat{\theta}_{PS}$. Nevertheless the estimator performs much better (see table B.22). This is due to the larger number of outliers identified by the recursive procedure. Thus in absolute numbers more true outliers have been eliminated from the panel before applying the ICM estimator in the second step. Table B.23 moreover shows that the significance level $\alpha = 0.01$ of the test is not reached empirically in the smaller panels. If on the average 0.2 time series are identified as outliers in a panel consisting of 10, this corresponds to a rejection rate of 2%. However if $n = 100$, $\hat{\theta}_{PS;rob}$ has a rejection rate of 0.88%, and also $\hat{\theta}_{PS}$ is very close to the significance level. This is caused by two effects: σ_Y^2, the variance of the asymptotic distribution, can only be approximated (see theorem 3.7.9 and algorithm 3.7.10). Furthermore the underlying estimate of the covariance matrix used for calculating the test statistic depends on the size of the panel. If there are outlying time series in the panel, not necessarily all of these are detected. This is especially the case if the difference in the dynamic structure is not very strong. The power grows with the length of the time series T increasing: for TS_1 with the panel size of $n = 10$ it is 83.9% for $T = 50$ and 99.2 for $T = 100$. If the proportion of outliers stays the same but the number of time series in the panel grows, the probability that all of these are detected sinks. Thus the power is only 80.0% if $n = 100$ and $T = 50$ although here the overall sample size is largest.

B.4 Comparative Evaluation of the Simulation Results

A non-robust estimator as the ICM parameter estimator investigated in chapter 2 is not stable under contamination. Therefore we have to search for alternatives. The different properties of the proposed estimators are illustrated by the above simulation study. Sim-

model		*corr.*			*ind.*
		$\hat{\theta}_{PS}$	$\hat{\theta}_{PS;rob}$	$\hat{\theta}_{PS;rec}$	$\hat{\theta}_{PS;rob}$
M_1	n_{OL}	0.1902	0.2316	0.2078	0.4438
AO_1	n_{OL}	0.6679	0.8082	0.5534	0.8302
M_1; $n = 10, T = 100$	n_{OL}	0.1412	0.1982	0.1748	0.3986
M_1; $n = 100, T = 50$	n_{OL}	1.0618	0.8836	1.4360	2.1666
TS_n	n_{OL}	0.2310	0.3096	0.2674	0.4718
	p_f	0.8061	0.6835	0.7861	0.8733
TS_1	n_{OL}	1.9988	1.6786	0.9596	1.7658
	p_f	0.3091	0.0487	0.0025	0.1108
TS_1; $n = 10, T = 100$	n_{OL}	2.8899	1.9844	0.9972	2.0966
	p_f	0.4161	0.0427	0.0000	0.0974
TS_1; $n = 100, T = 50$	n_{OL}	13.919	16.0022	17.2240	16.0192
	p_f	0.1476	0.0128	0.1476	0.0383

Table B.23: Performance of the phase space outlier test (significance level $\alpha = 0.01$) in various models. *corr.:* panels of intercorrelated autoregressive processes with parameters $a = 0.5$ and $\sigma^2 = \tau^2 = 0.5$. *ind.:* panels of independent autoregressive processes with parameters $a = 0.5$ and $\sigma^2 = 1$. Unless stated otherwise, the size of the panel is $n = 10$, $T = 50$. n_{OL}: average number of identified outliers; p_f: proportion of falsely identified outliers among these.

ply robustifying the parameter estimator by replacing all non-robust parts with robust counterparts leads to the estimators of section B.1. The more basic robust estimators $\hat{\theta}_{rob}$ and $\hat{\theta}_{oa}$ are outperformed by the reweighted procedures. These do not entirely belong to the class of directly robustified estimators as they use robust methods only for detecting outlying data. Then a second, non-robust estimation step is performed without these data. Here $\hat{\theta}_{rw}$ is less biased in the case of no contamination or entire time series outlying than $\hat{\theta}_{rw_2}$. The latter estimator has been designed to be robust against arbitrary outliers (see section 3.3). However its performance on AO_1 is not convincing. With exception of $\hat{\theta}_{rw}$ all of these robust estimators do not perform well in the case of the AR(6) models. The behaviour of $\hat{\theta}_{rw}$ suggests that there were no outlying data identified. Therefore in this case also the non-robust parameter estimator is not influenced much by the outlying time series. As bootstrap methods are a non-robust approach, the bias of the robust estimators cannot be substantially reduced. The procedure performing best is the one where the most detailed assumptions on the sample distributions are made, namely the one based on sampling directly from normal distributions. Although this reduces the bias of $\hat{\theta}_{rob}$ to some extent, the main properties are retained as bootstrapping reflects these (see the discussion in section B.1.1).

The second approach investigated in section B.2 consists of robustifying the covariance matrices used in the estimation of $\hat{\theta}_{n,T}$. It turns out that the robust panel autocovariance estimator and thus $\hat{\theta}_Q$ is negatively biased. However the same is true for the estimator based on the covariance matrix obtained from the minimum covariance determinant method. Both estimators are relatively stable against the case of entire time series outlying, comparably to $\hat{\theta}_{oa}$. But they are very robust against arbitrary outliers, which is not the case for the other methods. Thus in particular the estimator $\hat{\theta}_Q$ is promising if the bias can be overcome. It may be possible to reduce the bias by adjusting the order statistic employed or by modifying the correction used in the computation (see section B.2.1). Still, $\hat{\theta}_Q$ only performs well if the autoregressive order is small. Since all entries of the covariance matrix are estimated separately, it is instable. This can be seen from the results for the AR(6) case. Thus here it is preferable to use the estimator $\hat{\theta}_{MCD}$.

For comparison we include estimators derived from standard robust procedures as described in section 3.6. In the panel situation they are however in general not stable in the presence of outliers. It is already stated in Rousseeuw and Leroy (1987) that M-estimators are not robust against additive outliers. There the authors suggest using the least median of squares estimator for the analysis of time series. However the simulations show that also the estimator $\hat{\theta}_{LTS}$, which is derived from the least trimmed squares estimator, which in turn is closely related to the least median of squares estimator, does not perform satisfyingly in the panel situation. In the case of the AR(6) processes the estimators nevertheless yield improved results. This however can be due to the specific combination of models chosen as the non-robust ICM parameter estimator performs comparably. Thus we do not consider these estimators further.

Finally we have analysed the behaviour of outlier detection methods in section B.3. The simulations illustrate that the non-robust likelihood ratio test is indeed much influenced by outliers and thus not recommendable. The test based on the phase space representation however performs in an acceptable way if no arbitrary outliers are present. Its behaviour is compatible to $\hat{\theta}_{oa}$ for the smaller panels. If the data set becomes larger, the bias of the variance estimators is smaller than that of $\hat{\theta}_{oa}$.

Summarising we recommend a reweighted estimator if there are no arbitrary outliers. Here $\hat{\theta}_{rw}$ and $\hat{\theta}_{PS;rob}$ perform in a very similar way. $\hat{\theta}_{rw}$ yields slightly better results, but it is based on a heuristic procedure whereas we know the asymptotic distribution of the test statistic for $\hat{\theta}_{PS;rob}$. In the case of arbitrary outliers it is better to use one of the estimators derived from the robust covariance estimators as the other ones are not reliable. If the autoregressive order is one, $\hat{\theta}_Q$ is preferable. For higher order autoregressive processes $\hat{\theta}_{MCD}$ should be used. In practice, the arbitrary outliers can be excluded from the data set in a preliminary analysis. One method which has been mentioned in this thesis is the procedure of Gather, Bauer and Fried (2002).

For completeness we include an overview of the computation times needed by the different estimators. These are displayed in table B.24. It can be seen from the computation times that the non-robust estimators and their robustified versions are very fast to compute. The procedure for the GICM estimators $\hat{\theta}_a$ and $\hat{\theta}_b$ is even faster as we here do not need to compute the estimators iteratively. The computation of the reweighted versions is slightly slower as one more iteration is necessary. Furthermore the computation of the M-estimator $\hat{\theta}_M$ (GICM procedure) and of the non-robust outlier detection procedures are as fast as for the above estimators, and the least trimmed squares procedure is only a little bit slower. The estimation of $\hat{\theta}_Q$ requires a sorting, and the fast

est.	M_1	M_6
$\hat{\theta}_{n,T}$	4.09	5.47
$\hat{\theta}_a$, $\hat{\theta}_b$	2.08	3.06
$\hat{\theta}_{rob}$	4.11	6.52
$\hat{\theta}_{oa}$	4.11	5.98
$\hat{\theta}_{rw}$	4.52	7.01
$\hat{\theta}_{rw_2}$	4.53	6.98
$\hat{\theta}_{RB}$	1085.29	–
$\hat{\theta}_{PB}$	1011.83	–
$\hat{\theta}_{NB}$	1020.44	–
$\hat{\theta}_Q$	6.75	13.26
$\hat{\theta}_{MCD}$	19.59	76.30
$\hat{\theta}_M$	4.04	4.20
$\hat{\theta}_{LTS}$	5.13	8.49
$\hat{\theta}_{LR}$	4.10	5.29
$\hat{\theta}_{PS}$	5.34	5.23
$\hat{\theta}_{PS;rob}$	18.80	70.14
$\hat{\theta}_{PS;rec}$	23.90	128.37

Table B.24: Approximate computation time of the parameter estimators discussed in this chapter. Given is the time (in 1/100 sec.) needed for 10 iterations in model M_1 and M_6 on a PC with a 2.8GHz processor. If possible, the estimation was performed using an ICM-type estimator.

algorithm proposed in Croux and Rousseeuw (1992) cannot be transferred easily to the panel case. Therefore the computation of $\hat{\theta}_Q$ needs around twice the time than that of the first estimators. The estimators $\hat{\theta}_{MCD}$, $\hat{\theta}_{PS;rob}$ and $\hat{\theta}_{PS;rec}$ are based on the minimum volume ellipsoid estimator which takes a longer computation time. As $\hat{\theta}_{PS;rec}$ is a recursive estimator, its computation needs even more time. The computation of the bootstrap procedures is around 250 times slower than the original estimator. Thus the bootstrap corrections are not feasible if the sample size becomes large. Moreover the estimation procedure becomes instable for the AR(6) case. Thus we were not able to obtain bootstrap estimators for the AR(6) time and therefore cannot state the computation time.

Appendix C

Proofs and Auxiliary Results

C.1 Derivatives

We have computed the first derivatives of $\mathcal{L}_{n,T}$ (the log-likelihood function of the ICM) in section 2.4.2 because we needed them in order to construct the minimisation algorithm. Theorem 2.5.16, which shows the convergence of $D_n\, M_{\mathcal{L}_{n,T}}(\theta)\, D_n$ to Γ_n, is proved by investigating the convergence properties of the second derivatives. In order to simplify the notation, we have omitted to state these explicitly in the proof of the theorem. For sake of completeness, we now present them and the derivatives of the pointwise limits of the conditional log-likelihood functions in this section. We also include the second derivatives of the conditional log-likelihood function $\mathcal{L}^{\circ}_{n,T}$ obtained from the individual effects in the GICM in proposition 2.4.6.

Furthermore we have moved the proof of lemma 2.5.7 to this section as it is elementary but rather lengthy.

Derivatives of the log-likelihood functions

C.1.1 LEMMA

Under the assumptions of the ICM (assumption 2.2.1), let $\theta = (a', \sigma^2, \tau^2)' \in \Theta$ and denote $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n}$. Then the second derivatives of the conditional log-likelihood function $\mathcal{L}_{n,T}$ derived in proposition 2.4.2 are

$$\frac{\partial^2}{\partial a_k \partial a_l} \mathcal{L}_{n,T}(\theta) = \frac{2}{\sigma^2} \frac{1}{n\,(T-p)} \sum_{i=1}^{n} \sum_{t=p+1}^{T} \mathring{X}^{(i)}_{t-k}\, \mathring{X}^{(i)}_{t-l} + \frac{2}{\omega_n^2} \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \bar{X}_{t-k}\, \bar{X}_{t-l}\,,$$

$$\frac{\partial^2}{\partial a_k \partial \sigma^2} \mathcal{L}_{n,T}(\theta) = -\frac{2}{\sigma^4} \frac{1}{n\,(T-p)} \sum_{i=1}^{n} \sum_{t=p+1}^{T} \left(a(\mathrm{L})\, \mathring{X}^{(i)}_{t}\right) \mathring{X}^{(i)}_{t-k} - \frac{2}{\omega_n^4} \frac{1}{n^2\,(T-p)} \sum_{t=p+1}^{T} \left(a(\mathrm{L})\, \bar{X}_{t}\right) \bar{X}_{t-k}\,,$$

$$
\begin{aligned}
\frac{\partial^2}{\partial a_k \partial \tau^2}\, \mathcal{L}_{n,T}(\theta) &= -\frac{2}{\omega_n^4}\,\frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \big(a(\mathrm{L})\,\bar{X}_t\big)\;\bar{X}_{t-k}\,, \\
\frac{\partial^2}{(\partial \sigma^2)^2}\, \mathcal{L}_{n,T}(\theta) &= -\frac{n-1}{n\,\sigma^4} + \frac{2}{\sigma^6}\,A_{n,T}(a) - \frac{1}{n^3\,\omega_n^4} + \frac{2}{n^3\,\omega_n^6}\,B_{n,T}(a)\,, \\
\frac{\partial^2}{(\partial \tau^2)^2}\, \mathcal{L}_{n,T}(\theta) &= -\frac{1}{n\,\omega_n^4} + \frac{2}{n\,\omega_n^6}\,B_{n,T}(a) \\
\text{and} \qquad \frac{\partial^2}{\partial \sigma^2 \partial \tau^2}\, \mathcal{L}_{n,T}(\theta) &= -\frac{1}{n^2\,\omega_n^4} + \frac{2}{n^2\,\omega_n^6}\,B_{n,T}(a)\,,
\end{aligned}
$$

where $A_{n,T}(a)$ and $B_{n,T}(a)$ are defined in remark 2.4.3.

PROOF:
We get the statements by direct calculation, taking into account that $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n}$. □

In theorem 2.5.4 we have derived the pointwise limit of $\mathcal{L}_{n,T}(\theta)$ for $n \to \infty$ as

$$
\mathcal{L}(\theta) = \frac{1}{\sigma^2} \sum_{k,l=0}^{p} a_k\, a_l\, c(k-l) + \log \sigma^2 + \log 2\pi\,,
$$

where $c(h) = \Psi(h)\,\sigma_0^2$, $h \in \mathbb{Z}$, is the true autocovariance function of the identically distributed unobservable processes $\{Z_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, of lemma 2.2.4. Its derivatives are given in the following lemma.

C.1.2 LEMMA
Let $\tilde{\theta} = (a_1, \ldots, a_p, \sigma^2)' \in \tilde{\Theta}$ and $\mathcal{L}(\tilde{\theta})$ be as in definition 2.5.3. Then

$$
\begin{aligned}
\frac{\partial}{\partial a_l}\, \mathcal{L}(\tilde{\theta}) &= \frac{2}{\sigma^2} \sum_{k=0}^{p} a_k\, c(k-l) \\
\text{and} \qquad \frac{\partial}{\partial \sigma^2}\, \mathcal{L}(\tilde{\theta}) &= -\frac{1}{\sigma^4} \sum_{k,l=0}^{p} a_k a_l\, c(k-l) + \frac{1}{\sigma^2}\,.
\end{aligned}
$$

Furthermore

$$
\begin{aligned}
\frac{\partial^2}{\partial a_k \partial a_l}\, \mathcal{L}(\tilde{\theta}) &= \frac{2}{\sigma^2}\, c(k-l)\,, \\
\frac{\partial^2}{\partial a_l \partial \sigma^2}\, \mathcal{L}(\tilde{\theta}) &= -\frac{2}{\sigma^4} \sum_{k=0}^{p} a_k\, c(k-l) \\
\text{and} \qquad \frac{\partial^2}{(\partial \sigma^2)^2}\, \mathcal{L}(\tilde{\theta}) &= \frac{2}{\sigma^6} \sum_{k,l=0}^{p} a_k a_l\, c(k-l) - \frac{1}{\sigma^4}\,.
\end{aligned}
$$

PROOF:
Again the statements are obtained by straightforward calculation. □

In the case of $T \to \infty$, n fixed, we need the analogous results for the pointwise limit of $\mathcal{L}_{n,T}(\theta)$, which is (see definition 2.5.3)

$$\mathcal{L}_n(\theta) = \mathbb{E}\,\mathcal{L}_{n,T}(\theta) = \frac{1}{\sigma^2}\left(\frac{n-1}{n}\right) c_\theta + \frac{1}{\omega_n^2}\,\frac{1}{n}\, d_\theta + \left(\frac{n-1}{n}\right)\log\sigma^2 + \frac{1}{n}\log\omega_n^2 + \frac{1}{n}\log n + \log(2\pi)\,,$$

where $c_\theta = \sum_{k,l=0}^p a_k\, a_l\, c(k-l)$ and $d_\theta = \sum_{k,l=0}^p a_k\, a_l\, \bar\gamma_n(k-l)$.

C.1.3 LEMMA

Let $\theta = (a_1,\ldots,a_p,\sigma^2,\tau^2)' \in \Theta$ and $\mathcal{L}_n(\theta)$ be as in definition 2.5.3. Then, denoting $\omega_n^2 = \tau^2 + \frac{\sigma^2}{n}$, we get

$$\frac{\partial}{\partial a_l}\mathcal{L}_n(\theta) = \frac{2\,(n-1)}{n\,\sigma^2}\sum_{k=0}^p a_k\, c(k-l) + \frac{2}{n\,\omega_n^2}\sum_{k=0}^p a_k\,\bar\gamma_n(k-l)\,,$$

$$\frac{\partial}{\partial\sigma^2}\mathcal{L}_n(\theta) = -\frac{n-1}{n\,\sigma^4}\sum_{k,l=0}^p a_k\, a_l\, c(k-l) + \frac{n-1}{n\,\sigma^2} - \frac{1}{n^2\,\omega_n^4}\sum_{k,l=0}^p a_k\, a_l \bar\gamma_n(k-l) + \frac{1}{n^2\,\omega_n^2}\,,$$

and
$$\frac{\partial}{\partial\tau^2}\mathcal{L}_n(\theta) = -\frac{1}{n\,\omega_n^4}\sum_{k,l=0}^p a_k a_l\,\bar\gamma_n(k-l) + \frac{1}{n\,\omega_n^2}\,.$$

The second derivatives are

$$\frac{\partial^2}{\partial a_k \partial a_l}\mathcal{L}_n(\theta) = \frac{2\,(n-1)}{n\,\sigma^2}\, c(k-l) + \frac{2}{n\,\omega_n^2}\,\bar\gamma(k-l)\,,$$

$$\frac{\partial^2}{\partial a_l \partial\sigma^2}\mathcal{L}_n(\theta) = -\frac{2\,(n-1)}{n\,\sigma^4}\sum_{k=0}^p a_k\, c(k-l) - \frac{2}{n^2\,\omega_n^4}\sum_{k=0}^p a_k\,\bar\gamma_n(k-l)\,,$$

$$\frac{\partial^2}{\partial a_l \partial\tau^2}\mathcal{L}_n(\theta) = -\frac{2}{n\,\omega_n^4}\sum_{k=0}^p a_k\,\bar\gamma_n(k-l)\,,$$

$$\frac{\partial^2}{\partial\sigma^2\partial\tau^2}\mathcal{L}_n(\theta) = \frac{2}{n^2\,\omega_n^6}\sum_{k,l=0}^p a_k\, a_l\,\bar\gamma_n(k-l) - \frac{1}{n^2\,\omega_n^4}\,,$$

$$\frac{\partial^2}{(\partial\sigma^2)^2}\mathcal{L}_n(\theta) = \frac{2\,(n-1)}{n\,\sigma^6}\sum_{k,l=0}^p a_k a_l\, c(k-l) - \frac{n-1}{n\,\sigma^4} + \frac{2}{n^3\,\omega_n^6}\sum_{k,l=0}^p a_k\, a_l\bar\gamma_n(k-l) - \frac{1}{n^3\,\omega_n^4}\,,$$

and
$$\frac{\partial^2}{(\partial\tau^2)^2}\mathcal{L}_n(\theta) = \frac{2}{n\,\omega_n^6}\sum_{k,l=0}^p a_k a_l\,\bar\gamma_n(k-l) - \frac{1}{n\,\omega_n^4}\,.$$

PROOF:
As before, the calculation is straightforward. □

In particular, this allows us to calculate the Hessian matrices at the true value θ_0.

C.1.4 COROLLARY
Let $\theta_0 = (a', \sigma_0^2, \tau_0^2)'$*, with* $a = (a_1, \ldots, a_p)'$*, be the true parameter in the ICM and let* $\tilde{\theta}_0 = (a', \sigma_0^2)'$*. Then the Hessian matrices of* $\mathcal{L}$ *and* $\mathcal{L}_n$ *at the true values are*

$$\Gamma = \nabla^2 \mathcal{L}(\tilde{\theta}_0) = \begin{pmatrix} (2\,\Psi(k-l))_{k,l=1,\ldots,p} & 0 \\ 0 & \frac{1}{\sigma_0^4} \end{pmatrix}$$

$$\Gamma_n = \nabla^2 \mathcal{L}_n(\theta_0) = \begin{pmatrix} (2\,\Psi(k-l))_{k,l=1,\ldots,p} & 0 & 0 \\ 0 & \frac{n-1}{n\,\sigma_0^4} + \frac{1}{n^3\,\omega_n^4} & \frac{1}{n^2\,\omega_n^4} \\ 0 & \frac{1}{n^2\,\omega_n^4} & \frac{1}{n\,\omega_n^4} \end{pmatrix} = \mathbb{E}\,\nabla^2 \mathcal{L}_{n,T}(\theta_0)\,,$$

where now $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$ *denotes the true parameter in the ICM.*

PROOF:
The derivatives of $\mathcal{L}(\tilde{\theta})$ and $\mathcal{L}_n(\theta)$ can be found in the preceding lemmas C.1.2 and C.1.3. The results are due to the fact that the true values fulfil $\sum_{k,l=0}^{p} a_k\, a_l \Psi(k-l) = 1$ if we denote $a_0 = -1$ (remark 1.1.5). The last equality $\Gamma_n = \mathbb{E}\,\nabla^2 \mathcal{L}_{n,T}(\theta_0)$ is a consequence of the mean-square convergence of the panel autocovariance estimator proved in lemma 1.2.4. □

Finally we include some considerations on the derivatives of the conditional log-likelihood function $\mathcal{L}^\circ_{n,T}$ in the GICM, which is based on the individual processes. It is given in proposition 2.4.6 as

$$\begin{aligned} \mathcal{L}^\circ_{n,T}(\theta_a) &= -\frac{2}{n\,(T-p)} \log \mathcal{L}(\mathring{\mathbf{X}}_{p+1}, \ldots, \mathring{\mathbf{X}}_T \mid \mathring{\mathbf{X}}_1, \ldots, \mathring{\mathbf{X}}_p) \\ &= \frac{n-1}{n}\, \log \tilde{\sigma}_n^2 + \frac{1}{\tilde{\sigma}_n^2}\, \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \left(a(\mathrm{L})\, \mathring{X}_t^{(i)} \right)^2 \\ &\quad + \frac{n-1}{n}\, \log(2\pi) - \frac{1}{n} \log n\,, \end{aligned}$$

where $\tilde{\sigma}_n^2 = \sigma_n^2 - \sigma_n^{ij}$.

C.1.5 LEMMA
In the setting of the GICM (assumption 2.3.1) let $\theta_a = (a_1, \ldots, a_p, \sigma^2)' \in \Theta_a$ *and* $\mathcal{L}^\circ_{n,T}$ *be as obtained in proposition 2.4.6. Then*

$$\frac{\partial}{\partial a_l}\, \mathcal{L}^\circ_{n,T}(\theta_a) = \frac{2}{\sigma^2\, n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \left(\sum_{k=0}^{p} a_k\, \mathring{X}_{t-k}^{(i)} \right) \mathring{X}_{t-l}^{(i)}$$

$$\textit{and} \quad \frac{\partial}{\partial \sigma^2}\, \mathcal{L}^\circ_{n,T}(\theta_a) = \frac{n-1}{n\,\sigma^2} - \frac{1}{\sigma^4\, n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \sum_{k,l=0}^{p} a_k a_l\, \mathring{X}_{t-k}^{(i)}\, \mathring{X}_{t-l}^{(i)}\,.$$

Furthermore

$$\frac{\partial^2}{\partial a_k \partial a_l} \mathcal{L}^\circ_{n,T}(\theta_a) = \frac{2}{\sigma^2\, n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{X}^{(i)}_{t-k}\, \mathring{X}^{(i)}_{t-l}\,,$$

$$\frac{\partial^2}{\partial a_l \partial \sigma^2} \mathcal{L}^\circ_{n,T}(\theta_a) = -\frac{2}{\sigma^4\, n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \sum_{k=0}^{p} a_k\, \mathring{X}^{(i)}_{t-k}\, \mathring{X}^{(i)}_{t-l}$$

and $$\frac{\partial^2}{(\partial \sigma^2)^2} \mathcal{L}^\circ_{n,T}(\theta_a) = \frac{2}{\sigma^6\, n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \sum_{k,l=0}^{p} a_k\, a_l\, \mathring{X}^{(i)}_{t-k}\, \mathring{X}^{(i)}_{t-l} - \frac{n-1}{n\,\sigma^4}\,.$$

Denoting the true parameter $\theta_0 = (a_{1,0},\dots,a_{p,0},\tilde\sigma_n^2)'$*, we obtain that*

$$\nabla^2 \mathcal{L}^\circ_{n,T}(\theta_0) - \Gamma^\circ_n = O_P\left(\frac{1}{\sqrt{n\,T}}\right)\,,$$

where $\Gamma^\circ_n = \mathbb{E}\left(\nabla^2 \mathcal{L}^\circ_{n,T}(\theta_0)\right) = \frac{n-1}{n}\begin{pmatrix} 2\,B & 0 \\ 0 & \frac{1}{\tilde\sigma_n^4}\end{pmatrix}$ *with* $B = (\Psi(k-l))_{k,l=1,\dots,p}$.

PROOF:
The derivatives are obtained by straightforward calculation. In the GICM, we have that $\mathring{Z}^{(i)}_t = \mathring{X}^{(i)}_t$ and $\mathbb{E}\left(\mathring{X}^{(i)}_t, \mathring{X}^{(j)}_t\right) = \left(\delta_{ij} - \frac{1}{n}\right)\tilde\sigma_n^2$ (remark 2.3.5). Thus the mean-square convergence of the panel autocovariance estimator proved in lemma 1.2.4 implies the result. □

Proof of lemma 2.5.7

Subsequently we show that θ_0 and $\tilde\theta_0$ are unique minima of $\mathcal{L}_n$ and $\mathcal{L}$ if the parameter spaces are small enough.

PROOF OF LEMMA 2.5.7:
We prove the lemma first for $\tilde\theta_0 = \operatorname{argmin}_{\tilde\theta\in\tilde\Theta}\mathcal{L}(\tilde\theta)$. By the same method, the statement can be shown for θ_0.
Recall that for $\tilde\theta = (\alpha_1,\dots,\alpha_p,\sigma^2) \in \tilde\Theta$ we have due to the assumptions of the ICM (assumption 2.2.1) that $\sigma^2 \geq c > 0$. The limit function $\mathcal{L}(\tilde\theta)$ has been given in definition 2.5.3 as $\mathcal{L}(\tilde\theta) = \frac{1}{\sigma^2}\, c_{\tilde\theta} + \log\sigma^2 + \log(2\pi)$ where $c_{\tilde\theta} = \sum_{k,l=0}^{p} \alpha_k\,\alpha_l\, c(k-l)$ with $\alpha_0 = -1$. Now choose $\tilde\Theta' \subseteq \tilde\Theta$ such that each $\tilde\theta \in \tilde\Theta'$ can be written as $\tilde\theta = \tilde\theta_0 + \nu$ with $\tilde\theta_0 = (a_1,\dots,a_p,\sigma_0^2)$ and $\nu = (\nu_1,\dots,\nu_p,\delta\,\sigma_0^2)\,,\ \delta > -1$. This condition ensures that $\sigma^2 > 0$. In order to simplify notations, denote $\nu_0 = 0$. We calculate the partial derivative of $\mathcal{L}(\tilde\theta)$ in direction of $x_{\tilde\theta} = (x_1,\dots,x_p,x_\sigma) = -\nu$.
For the true parameter $\tilde\theta_0$ and the covariance function $c(h)$ we know (remark 1.1.5) that $\sum_{k,l=0}^{p} a_k\, a_l\, c(k-l) = \sigma_0^2$. Therefore

$$\frac{\partial}{\partial x_{\tilde\theta}}\mathcal{L}(\tilde\theta) = \frac{2}{\sigma_0^2\,(1+\delta)} \sum_{k=1}^{p}\sum_{l=0}^{p} (a_l+\nu_l) c(k-l)\, x_k$$
$$- \frac{1}{\sigma_0^4\,(1+\delta)^2} \sum_{k,l=0}^{p} (a_k+\nu_k)(a_l+\nu_l) c(k-l)\, x_\sigma + \frac{1}{\sigma_0^2\,(1+\delta)}\, x_\sigma$$

$$
\begin{aligned}
&= \frac{2}{\sigma_0^2\,(1+\delta)} \sum_{k=1}^{p} \sum_{l=0}^{p} \nu_l\, c(k-l)\, x_k \;-\; \frac{1}{\sigma_0^4\,(1+\delta)^2} \sum_{k,l=0}^{p} a_k a_l\, c(k-l)\, x_\sigma \\
&\quad - \frac{1}{\sigma_0^4\,(1+\delta)^2} \sum_{k,l=0}^{p} \nu_k \nu_l\, c(k-l)\, x_\sigma + \frac{1}{\sigma_0^2\,(1+\delta)}\, x_\sigma \\
&= -\frac{2}{\sigma_0^2\,(1+\delta)} \sum_{k,l=0}^{p} \nu_k \nu_l\, c(k-l) + \frac{\delta}{(1+\delta)^2} \\
&\quad + \frac{\delta}{\sigma_0^2\,(1+\delta)^2} \sum_{k,l=0}^{p} \nu_k \nu_l\, c(k-l) - \frac{\delta}{(1+\delta)} \\
&= -\sum_{k,l=1}^{p} \nu_k \nu_l\, c(k-l)\, \frac{2+\delta}{\sigma_0^2\,(1+\delta)^2} - \frac{\delta^2}{(1+\delta)^2}\,.
\end{aligned}
$$

Being a covariance function, $(c(k-l))_{k,l=1,\dots,p}$ is positive semidefinite. Since $\delta > -1$, we thus get on a neighbourhood $\tilde{\Theta}' \subseteq \tilde{\Theta}$ of $\tilde{\theta}_0$ that

$$
\frac{\partial \mathcal{L}}{\partial x_{\tilde{\theta}}}(\tilde{\theta}) < 0 \qquad \text{for all } \tilde{\theta} \in \tilde{\Theta}'.
$$

In the second case, choose $\Theta' \subseteq \Theta$ such that for all $\theta = (\alpha_1, \dots, \alpha_p, \sigma^2, \tau^2) \in \Theta'$ we have $\theta = \theta_0 + \nu$ with $\theta_0 = (a_1, \dots, a_p, \sigma_0^2, \tau_0^2)$ and $\nu = (\nu_1, \dots, \nu_p, \delta_1 \sigma_0^2, \delta_2 \tau_0^2) = -x_\theta$, where $|\delta_i| < 1$ ($i = 1, 2$). Then the calculations are analogous. We only have to take into account that $\omega_\theta^2 = \tau^2 + \frac{\sigma^2}{n}$ and $\omega_n^2 = \tau_0^2 + \frac{\sigma_0^2}{n}$ lead to

$$
\omega_\nu = \omega_\theta^2 - \omega_n^2 = \delta_1 \frac{\sigma_0^2}{n} + \delta_2 \tau_0^2\,.
$$

If we denote $\min(\delta_1, \delta_2)$ by δ_- and $\max(\delta_1, \delta_2)$ by δ_+, this implies that

$$
\omega_\theta^2 \geq (1+\delta_-)\omega_n^2 \qquad \text{and thus} \qquad \omega_\nu \leq \delta_+ \omega_n^2 \leq \frac{\delta_+}{1+\delta_-} \omega_\theta^2\,.
$$

This guarantees that analogous calculations as above lead to

$$
\frac{\partial \mathcal{L}_n}{\partial x_\theta}(\theta) < 0 \qquad \text{for all } \theta \in \Theta',
$$

where Θ' is a neighbourhood of θ_0. Altogether this means that $\tilde{\theta}_0$ and θ_0 are unique minima if the parameter spaces are chosen small enough. □

C.2 Auxiliary Results for Section 2.5.4

C.2.1 Proof of Lemma 2.5.26

For better readability we have omitted this straightforward proof from section 2.5.4.

PROOF OF LEMMA 2.5.26
The innovations $\{\xi_t\}_{t\in\mathbb{Z}}$ are independently and identically distributed Gaussian random variables such that $\mathbb{E}\,\xi_t = 0$ and $\operatorname{var}\xi_t = \omega_n^2$. Therefore we get, because we know that $\operatorname{var}\bar{X}_t = \Psi(0)\,\omega_n^2$ (see lemma 2.2.4),

$$\mathbb{E}\Big(\frac{1}{\sqrt{n\,(T-p)}}\sum_{t=p+1}^{T}\frac{2}{\omega_n^2}\,\xi_t\,\bar{X}_{t-k}\Big)^2$$
$$= \frac{1}{n\,(T-p)}\sum_{s,t=p+1}^{T}\frac{4}{\omega_n^4}\,\mathbb{E}\left(\xi_s\,\xi_t\right)\,\mathbb{E}\left(\bar{X}_{s-k}\,\bar{X}_{t-k}\right) = \frac{4}{n}\,\Psi(0) = O\left(\frac{1}{n}\right).$$

As for Gaussian processes all cumulants of third and higher order are zero, we obtain from the formula of (Shiryayev 1984, p. 290), that $\mathbb{E}\left(\xi_s^2\,\xi_t^2\right) = \omega_n^2\,\left(1 + 2\,\delta_{st}\right)$, where δ_{st} is the Kronecker delta. This yields

$$\mathbb{E}\Big(\frac{1}{\sqrt{n\,(T-p)}}\sum_{t=p+1}^{T}\frac{1}{\omega_n^2}\left(-\frac{1}{\omega_n^2}\,\xi_t^2 + 1\right)\Big)^2$$
$$= \frac{1}{n\,(T-p)\,\omega_n^4}\sum_{s,t=p+1}^{T}\left(\frac{1}{\omega_n^4}\,\mathbb{E}\left(\xi_s^2\,\xi_t^2\right) - \frac{2}{\omega_n^2}\,\mathbb{E}\left(\xi_s^2\right) + 1\right)$$
$$= \frac{1}{n\,(T-p)\,\omega_n^4}\sum_{s,t=p+1}^{T}\left(1 + 2\,\delta_{st} - 2 + 1\right) = O\left(\frac{1}{n\,\omega_n^4}\right).$$

Furthermore we have in the ICM $\mathring{\varepsilon}_t^{(i)} = \varepsilon_t^{(i)} - \frac{1}{n}\sum_{j=1}^{n}\varepsilon_t^{(j)}$ and $\mathring{X}_t^{(i)} = \sum_{u=0}^{\infty}\psi_u\,\mathring{\varepsilon}_{t-u}^{(i)}$ for $i = 1,\ldots,n$ (remark 2.2.3), where the innovations $\varepsilon_t^{(i)}$ are independently and identically distributed for all $t \in \mathbb{Z}$ and $i = 1,\ldots,n$. Thus

$$\mathbb{E}\left(\frac{1}{\sqrt{n\,(T-p)}}\sum_{t=p+1}^{T}\sum_{i=1}^{n}\Big[-\frac{1}{\sigma_0^2}\left(\mathring{\varepsilon}_t^{(i)}\,\mathring{X}_{t-k}^{(i)} - \varepsilon_t^{(i)}\,Z_{t-k}^{(i)}\right)\Big]\right)^2$$
$$= \frac{1}{n\,(T-p)}\sum_{s,t=p+1}^{T}\sum_{i,j=1}^{n}\Psi(0)\,\delta_{st}\left(\left(\delta_{ij} - \frac{1}{n}\right)^2 - 2\left(\delta_{ij} - \frac{1}{n}\right)^2 + \delta_{ij}\right)\sigma_0^2$$
$$= \frac{4}{n}\,\Psi(0)\,\sigma_0^2.$$

Moreover, since the $\varepsilon_t^{(i)}$ are assumed to be Gaussian and

$$\sum_{i=1}^{n}\left(\mathring{\varepsilon}_t^{(i)\,2} - \varepsilon_t^{(i)\,2}\right) = -\frac{1}{n}\sum_{i,j=1}^{n}\varepsilon_t^{(i)}\,\varepsilon_t^{(j)},$$

we obtain

$$\mathbb{E}\left(\frac{1}{\sqrt{n\,(T-p)}}\sum_{t=p+1}^{T}\Big[-\frac{1}{\sigma_0^4}\sum_{i=1}^{n}\left(\mathring{\varepsilon}_t^{(i)\,2} - \varepsilon_t^{(i)\,2}\right) - \frac{1}{\sigma_0^2}\Big]\right)^2$$
$$= \frac{1}{n\,(T-p)}\sum_{s,t=p+1}^{T}\Big[\frac{1}{n^2\,\sigma_0^8}\,\mathbb{E}\left(\left(\sum_{i,j=1}^{n}\varepsilon_s^{(i)}\,\varepsilon_s^{(j)}\right)\left(\sum_{i,j=1}^{n}\varepsilon_t^{(i)}\,\varepsilon_t^{(j)}\right)\right)$$

$$- \frac{2}{n\,\sigma_0^6}\,\mathbb{E}\left(\sum_{i,j=1}^{n} \varepsilon_t^{(i)}\,\varepsilon_t^{(j)}\right) + \frac{1}{\sigma_0^4}\Big]$$

$$= \frac{1}{n\,(T-p)} \sum_{s,t=p+1}^{T} \frac{1}{\sigma_0^4}\,\left((1+2\,\delta_{st}) - 2 + 1\right) = \frac{2}{n\,\sigma_0^4} = O\left(\frac{1}{n}\right). \qquad \square$$

C.2.2 Properties of the Martingale Differences

Proposition 2.5.31 proves the convergence conditions on the martingale differences $D_{T,t,\lambda}$. Two steps have not been entirely covered there as the conclusions are straightforward but require more extensive calculations. These can be found here. First note the following.

C.2.1 REMARK
Summation over all possible indices yields

$$\begin{aligned}\sum_{i,j=1}^{n}\left(\delta_{ij}-\frac{1}{n}\right)^2 &= \sum_{i,j,k=1}^{n}\left(\delta_{ij}-\frac{1}{n}\right)\left(\delta_{jk}-\frac{1}{n}\right)\left(\delta_{ik}-\frac{1}{n}\right)\\ &= \sum_{i,j,k,l=1}^{n}\left(\delta_{ij}-\frac{1}{n}\right)\left(\delta_{jk}-\frac{1}{n}\right)\left(\delta_{kl}-\frac{1}{n}\right)\left(\delta_{il}-\frac{1}{n}\right)\\ &= n-1\,.\end{aligned}$$

We first prove the statement on the conditional variance.

C.2.2 PROPOSITION
The martingale differences constructed in lemma 2.5.30 fulfil

$$\sum_{t=p+1}^{T} \mathbb{E}(D_{T,t,\lambda}^2 \mid \mathcal{F}_{T,t-1}) - \lambda'\,\Sigma_n\,\lambda = o_P(1)\,,$$

with $\Sigma_n = 2\begin{pmatrix} (2\,\Psi(k-l))_{k,l=1,\dots,p} & 0 & 0 \\ 0 & \frac{n-1}{n\,\sigma_0^4} + \frac{1}{n^3\,\omega_n^4} & \frac{1}{n\,\sqrt{n}\,\omega_n^4} \\ 0 & \frac{1}{n^2\,\sqrt{n}\,\omega_n^4} & \frac{1}{\omega_n^4}\end{pmatrix}.$

PROOF:
By construction the martingale differences are $D_{T,t,\lambda} = \frac{1}{\sqrt{n_T\,(T-p)}} \sum_{t=p+1}^{\tau} \lambda'\,\mathbf{Z}_t^{(i)}$, where $\lambda \in \mathbb{R}^{p+2}$ and the variables $\mathbf{Z}_t^{(i)}$, $p+1 \le t \le T$, $i = 1,\dots,n_T$, are such that $D_n\,\nabla\mathcal{L}_{n_T,T}(\theta_0) = \frac{1}{n_T\,(T-p)} \sum_{t=p+1}^{T}\sum_{i=1}^{n_T} \mathbf{Z}_t^{(i)}$, where $\nabla\mathcal{L}_{n,T}(\theta_0)$ is given in remark 2.5.24 and $D_n = \begin{pmatrix} I_{p+1} & 0 \\ 0 & \sqrt{n}\end{pmatrix}$. This yields

$$\sum_{t=p+1}^{T} \mathbb{E}\Big(D_{T,t,\lambda}^2 \mid \mathcal{F}_{T,t-1}\Big)$$

$$
\begin{aligned}
&= \frac{1}{n_T\,(T-p)} \sum_{t=p+1}^{T} \\
&\qquad \Big[\sum_{k,l=1}^{p} \lambda_k\,\lambda_l \sum_{i,j=1}^{n_T} \left(\frac{4}{\sigma_0^2}\,\mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(j)} \left(\delta_{ij} - \frac{1}{n_T}\right) + \frac{4}{n_T^2\,\omega_n^2}\,\bar{X}_{t-k}\,\bar{X}_{t-l} \right) \\
&\qquad\quad + \lambda_{p+1}^2 \sum_{i,j=1}^{n_T} \left(\frac{1}{\sigma_0^4} \Big[\mathbb{E}\left(\mathring{\varepsilon}_t^{(i)\,2}\,\mathring{\varepsilon}_t^{(j)\,2}/\sigma_0^4 \right) - \frac{2\,(n_T-1)}{n_T}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i)\,2}/\sigma_0^2 \right) \right. \\
&\qquad\qquad\qquad\qquad \left. + \left(\frac{n_T-1}{n_T} \right)^2 \Big] + \frac{1}{n_T^4\,\omega_n^4}\,\mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1 \right)^2 \right) \\
&\qquad\quad + \lambda_{p+2}^2 \sum_{i,j=1}^{n_T} \frac{1}{n_T\,\omega_n^4}\,\mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1 \right)^2 \\
&\qquad\quad + 2\,\lambda_{p+1}\,\lambda_{p+2} \sum_{i,j=1}^{n_T} \frac{1}{n_T^2\,\sqrt{n_T}\,\omega_n^4}\,\mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1 \right)^2 \ \Big] \\
&= \frac{1}{n_T\,(T-p)} \sum_{t=p+1}^{T} \Big[\sum_{k,l=1}^{p} \lambda_k\,\lambda_l \\
&\qquad\qquad \times \left(\frac{4}{\sigma_0^2} \sum_{i,j=1}^{n_T} \mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(j)} \left(\delta_{ij} - \frac{1}{n_T}\right) + \frac{4}{\omega_n^2}\,\bar{X}_{t-k}\,\bar{X}_{t-l} \right) \Big] \\
&\quad + 2\,\lambda_{p+1}^2 \left(\frac{n_T-1}{n_T\,\sigma_0^4} + \frac{1}{n_T^3\,\omega_n^4} \right) + 2\,\lambda_{p+2}^2\,\frac{1}{\omega_n^4} + 4\,\lambda_{p+1}\,\lambda_{p+2}\,\frac{1}{n_T\,\sqrt{n_T}\,\omega_n^8}\,.
\end{aligned}
$$

The first two terms are mean-square convergent: as

$$
\begin{aligned}
\mathbb{E}\left(\sum_{i,j=1}^{n_T} \mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(j)} \left(\delta_{ij} - \frac{1}{n_T}\right) \right) &= \sum_{i,j=1}^{n_T} \left(\delta_{ij} - \frac{1}{n_T}\right)^2 \psi(k-l)\,\sigma_0^2 \\
&= (n_T - 1)\,\psi(k-l)\,\sigma_0^2\,,
\end{aligned}
$$

we obtain from the summation property of $\delta_{ij} - \frac{1}{n_T}$ (see the preceding remark) that

$$
\begin{aligned}
&\mathbb{E}\Big(\frac{1}{n_T\,(T-p)} \sum_{t=p+1}^{T} \sum_{i,j=1}^{n_T} \frac{4}{\sigma_0^2}\,\mathring{X}_{t-k}^{(i)}\,\mathring{X}_{t-l}^{(j)} \left(\delta_{ij} - \frac{1}{n_T}\right) - 4\,\frac{n_T-1}{n_T}\,\psi(k-l) \Big)^2 \\
&= \frac{16}{n_T^2\,(T-p)^2} \sum_{s,t=p+1}^{T} \\
&\qquad \sum_{i_1,i_2,i_3,i_4=1}^{n_T} \left(\delta_{i_1 i_2} - \frac{1}{n_T}\right) \left(\delta_{i_2 i_3} - \frac{1}{n_T}\right) \left(\delta_{i_3 i_4} - \frac{1}{n_T}\right) \left(\delta_{i_1 i_4} - \frac{1}{n_T}\right) \\
&\qquad\qquad \times \left(\psi(s-t-k+l)\,\psi(s-t-l+k) + \psi(s-t)^2 \right) \\
&= \frac{16}{n_T^2\,(T-p)^2} \sum_{s,t=p+1}^{T} (n_T - 1) \left(\psi(s-t-k+l)\,\psi(s-t-l+k) + \psi(s-t)^2 \right) \\
&= O\left(\frac{1}{n_T\,T} \right)\,.
\end{aligned}
$$

Analogously we get that

$$\begin{aligned}
&\mathbb{E}\Big(\frac{1}{n_T\,(T-p)}\sum_{t=p+1}^{T}\frac{4}{\omega_n^2}\,\bar{X}_{t-k}\,\bar{X}_{t-l}-\frac{4}{n_T}\,\psi(k-l)\Big)^2\\
&=\frac{16}{n_T^2\,(T-p)^2}\sum_{s,t=p+1}^{T}\left(\psi(s-t-k+l)\,\psi(s-t-l+k)+\psi(s-t)^2\right)\\
&=O\left(\frac{1}{n_T^2\,T}\right).
\end{aligned}$$

Thus we have altogether that $\sum_{t=p+1}^{T}\mathbb{E}(D_{T,t,\lambda}^2\mid\mathcal{F}_{T,t-1})\,-\,\lambda'\,\Sigma_n\,\lambda\,=\,o_P(1)$. □

In the proof of proposition 2.5.31 we use that the 4th moment of the martingale differences $D_{T,t,\lambda}$ is bounded. The proof of this statement involves rather lengthy calculations.

C.2.3 PROPOSITION
Let $\lambda\in\mathbb{R}^{p+2}$ *and*

$$D_{T,t,\lambda}=\frac{1}{\sqrt{n_T\,(T-p)}}\sum_{i=1}^{n_T}\lambda'\begin{pmatrix}\left(-\frac{2}{\sigma_0^2}\,\mathring{\varepsilon}_t^{(i)}\,\mathring{X}_{t-k}^{(i)}-\frac{2}{n_T\,\omega_n^2}\,\xi_t\,\bar{X}_{t-k}\right)_{k=1,\dots,p}\\ -\frac{1}{\sigma_0^4}\,\mathring{\varepsilon}_t^{(i)\,2}+\frac{n_T-1}{n_T\,\sigma_0^2}-\frac{1}{n_T^2\,\omega_n^4}\,\xi_t^2+\frac{1}{n_T^2\,\omega_n^2}\\ -\frac{1}{\sqrt{n_T}\,\omega_n^4}\,\xi_t^2+\frac{1}{\sqrt{n_T}\,\omega_n^2}\end{pmatrix}$$

as defined in lemma 2.5.30. Then

$$\mathbb{E}\left(D_{T,t,\lambda}^4\right)=O\left(\frac{1}{T^2}\right).$$

PROOF:
The proof is based on the fact that the variables $\mathring{\varepsilon}_s^{(i)}$ and ξ_t are independent for all $s,t\in\mathbb{Z}$, $i=1,\dots,n$, with $\mathbb{E}\left(\mathring{\varepsilon}_t^{(i)}\,\mathring{\varepsilon}_t^{(j)}\right)=\left(\delta_{ij}-\frac{1}{n}\right)\sigma_0^2$ and $\operatorname{var}\xi_t=\omega_n^2$. Due to the Gaussianity assumption, the higher moments can be calculated using second order cumulants. This means that we have to compute the values of the expectation for each partition of the variables into cycles of length two. Thus the calculations are lengthy but straightforward. The formula can e.g. be found in (Shiryayev 1984, p. 402). In order to facilitate the notation we here omit the index of n_T.
First recall (remark C.2.1) that

$$\begin{aligned}
\sum_{i,j=1}^{n}\left(\delta_{ij}-\frac{1}{n}\right)^2&=\sum_{i,j,k=1}^{n}\left(\delta_{ij}-\frac{1}{n}\right)\left(\delta_{jk}-\frac{1}{n}\right)\left(\delta_{ik}-\frac{1}{n}\right)\\
&=\sum_{i,j,k,l=1}^{n}\left(\delta_{ij}-\frac{1}{n}\right)\left(\delta_{jk}-\frac{1}{n}\right)\left(\delta_{kl}-\frac{1}{n}\right)\left(\delta_{il}-\frac{1}{n}\right)\\
&=n-1\,.
\end{aligned}$$

For the most complicated term we thus obtain that

$$\sum_{i_1,i_2,i_3,i_4=1}^{n} \mathbb{E}\Big(\mathring{\varepsilon}_t^{(i_1)\,2}\,\mathring{\varepsilon}_t^{(i_2)\,2}\,\mathring{\varepsilon}_t^{(i_3)\,2}\,\mathring{\varepsilon}_t^{(i_4)\,2}\Big)$$

$$= \Big(n^4\,\left(\frac{n-1}{n}\right)^4 + 12\,n^2\,\left(\frac{n-1}{n}\right)^2\,(n-1) + 32\,n\,\frac{n-1}{n}\,(n-1) + 12\,(n-1)^2 + 48\,(n-1)\Big)\,\sigma_0^8$$

$$= \big(n^4 + 8\,n^3 + 14\,n^2 - 8\,n - 15\big)\,\sigma_0^8\,.$$

Similarily, we get

$$\sum_{i_1,i_2,i_3=1}^{n} \mathbb{E}\Big(\mathring{\varepsilon}_t^{(i_1)\,2}\,\mathring{\varepsilon}_t^{(i_2)\,2}\,\mathring{\varepsilon}_t^{(i_3)\,2}\Big)$$

$$= \Big(n^3\,\left(\frac{n-1}{n}\right)^3 + 6\,n\,\frac{n-1}{n}\,(n-1) + 8(n-1)\Big)\,\sigma_0^6$$

$$= \Big(n^3 + 3\,n^2 - n - 3\Big)\,\sigma_0^6$$

and $$\sum_{i_1,i_2=1}^{n} \mathbb{E}\Big(\mathring{\varepsilon}_t^{(i_1)\,2}\,\mathring{\varepsilon}_t^{(i_2)\,2}\Big) = \Big(n^2\,\left(\frac{n-1}{n}\right)^2 + 2\,(n-1)\Big)\,\sigma_0^4 = \Big(n^2 - 1\Big)\,\sigma_0^4\,.$$

Since $\sum_{i=1}^{n}\left(\delta_{ik} - \frac{1}{n}\right) = 0$ for all $k = 1, \ldots, n$, we furthermore have that

$$\sum_{i,j,k=1}^{n} \mathbb{E}\left(\mathring{\varepsilon}_t^{(i)}\,\mathring{\varepsilon}_t^{(j)}\,\mathring{\varepsilon}_t^{(k)\,2}\right)\,\left(\delta_{ij} - \frac{1}{n}\right)$$

$$= \sum_{i,j,k=1}^{n} \Big[\frac{n-1}{n}\,\left(\delta_{ij} - \frac{1}{n}\right)^2 + 2\,\left(\delta_{ik} - \frac{1}{n}\right)\,\left(\delta_{jk} - \frac{1}{n}\right)\,\left(\delta_{ij} - \frac{1}{n}\right)\Big]$$

$$= \left(n\,\frac{n-1}{n}\,(n-1) + 2\,(n-1)\right)\,\sigma_0^6 = \big(n^2 - 1\big)\,\sigma_0^6$$

and $$\sum_{i,j,k,l=1}^{n} \mathbb{E}\left(\mathring{\varepsilon}_t^{(i)}\,\mathring{\varepsilon}_t^{(j)}\,\mathring{\varepsilon}_t^{(k)\,2}\,\mathring{\varepsilon}_t^{(l)\,2}\right)\,\left(\delta_{ij} - \frac{1}{n}\right)$$

$$= \sum_{i,j=1}^{n}\left(\delta_{ij} - \frac{1}{n}\right)^2 \sum_{k,l=1}^{n} \mathbb{E}\left(\mathring{\varepsilon}_t^{(k)\,2}\,\mathring{\varepsilon}_t^{(l)\,2}\right) + 4\sum_{i,j,k,l=1}^{n}\left(\delta_{ij} - \frac{1}{n}\right)\left(\delta_{ik} - \frac{1}{n}\right)\Big[\left(\delta_{jk} - \frac{1}{n}\right)\left(\frac{n-1}{n}\right) + 2\left(\delta_{jl} - \frac{1}{n}\right)\left(\delta_{kl} - \frac{1}{n}\right)\Big]$$

$$= \left((n^2 - 1)\,(n-1) + 4\,n\,\frac{n-1}{n}\,(n-1) + 8\,(n-1)\right)\,\sigma_0^6$$

$$= (n-1)\,(n^2 + 4\,n + 3)\,\sigma_0^6\,.$$

As

$$\mathbb{E}\left(-\mathring{\varepsilon}_t^{(i)\,2}/\sigma_0^2 + \frac{n-1}{n}\right) = \mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1\right) = 0\,,$$

$$\mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1\right)^2 = \mathbb{E}\,\xi_t^4/\omega_n^4 - 2\,\mathbb{E}\,\xi_t^2/\omega_n^2 + 1 = 2\,,$$

$$\mathbb{E}\,\xi_t^6/\omega_n^4 - 2\,\mathbb{E}\,\xi_t^4/\omega_n^2 + \mathbb{E}\,\xi_t^2 = \omega_n^2\,(15 - 2\times 3 + 1) = 10\,\omega_n^2$$

$$\text{and} \qquad \mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1\right)^4 = 105 - 4\times 15 + 6\times 3 - 4 + 1 = 60\,,$$

this yields

$$\begin{aligned}
\mathbb{E}\left(D_{T,t,\lambda}^4\right) = \frac{1}{n^2\,(T-p)^2} \sum_{i_1,i_2,i_3,i_4=1}^{n} \\
\Bigg(\sum_{k_1,k_2,l_1,l_2=1}^{p} \lambda_{k_1}\,\lambda_{k_2}\,\lambda_{l_1}\,\lambda_{l_2} \Bigg[\frac{16}{\sigma_0^8}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)}\,\mathring{\varepsilon}_t^{(i_2)}\,\mathring{\varepsilon}_t^{(i_3)}\,\mathring{\varepsilon}_t^{(i_4)}\right)\,\mathbb{E}\left(\mathring{X}_{t-k_1}^{(i_1)}\,\mathring{X}_{t-k_2}^{(i_2)}\,\mathring{X}_{t-l_1}^{(i_3)}\,\mathring{X}_{t-l_2}^{(i_4)}\right) \\
+ 6\,\frac{16}{n^2\,\sigma_0^4\,\omega_n^4}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)}\,\mathring{\varepsilon}_t^{(i_2)}\right)\,\mathbb{E}\left(\mathring{X}_{t-k_1}^{(i_1)}\,\mathring{X}_{t-k_2}^{(i_2)}\right)\,\mathbb{E}\left(\xi_t^2\right)\,\mathbb{E}\left(\bar{X}_{t-l_1}\,\bar{X}_{t-l_2}\right) \\
+ \frac{16}{n^4\,\omega_n^8}\,\mathbb{E}\left(\xi_t^4\right)\,\mathbb{E}\left(\bar{X}_{t-k_1}\,\bar{X}_{t-k_2}\,\bar{X}_{t-l_1}\,\bar{X}_{t-l_2}\right) \Bigg] \\
+ \lambda_{p+1}^4 \Bigg[\frac{1}{\sigma_0^8} \Bigg[\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)\,2}\,\mathring{\varepsilon}_t^{(i_2)\,2}\,\mathring{\varepsilon}_t^{(i_3)\,2}\,\mathring{\varepsilon}_t^{(i_4)\,2}\right)/\sigma_0^8 - 4\,\frac{n-1}{n}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)\,2}\,\mathring{\varepsilon}_t^{(i_2)\,2}\,\mathring{\varepsilon}_t^{(i_3)\,2}\right)/\sigma_0^6 \\
+ 6\left(\frac{n-1}{n}\right)^2\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)\,2}\,\mathring{\varepsilon}_t^{(i_2)\,2}\right)/\sigma_0^4 - 4\left(\frac{n-1}{n}\right)^3\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)\,2}\right)/\sigma_0^2 + \left(\frac{n-1}{n}\right)^4 \Bigg] \\
+ \frac{6}{n^2\,\sigma_0^4\,\omega_n^4} \Bigg[\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)\,2}\,\mathring{\varepsilon}_t^{(i_2)\,2}\right)/\sigma_0^4 - 2\,\frac{n-1}{n}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)\,2}\right)/\sigma_0^2 + \left(\frac{n-1}{n}\right)^2 \Bigg] \\
\times \left(\mathbb{E}\,\xi_t^4/\omega_n^4 - \mathbb{E}\,\xi_t^2/\omega_n^2 + 1\right) \\
+ \frac{1}{n^4\,\omega_n^8}\,\mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1\right)^4 \Bigg] \\
+ \lambda_{p+2}^4\,\frac{1}{n^2\,\omega_n^8}\,\mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1\right)^4 \\
+ 6 \sum_{k,l=1}^{p} \lambda_k\,\lambda_l\,\lambda_{p+1}^2 \Bigg[\frac{4}{\sigma_0^8} \Bigg[\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)}\,\mathring{\varepsilon}_t^{(i_2)}\,\mathring{\varepsilon}_t^{(i_3)\,2}\,\mathring{\varepsilon}_t^{(i_4)\,2}\right)/\sigma_0^4 - 2\,\frac{n-1}{n}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)}\,\mathring{\varepsilon}_t^{(i_2)}\,\mathring{\varepsilon}_t^{(i_3)\,2}\right)/\sigma_0^2 \\
+ \left(\frac{n-1}{n}\right)^2\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)}\,\mathring{\varepsilon}_t^{(i_2)}\right) \Bigg] \times \mathbb{E}\left(\mathring{X}_{t-k_1}^{(i_1)}\,\mathring{X}_{t-k_2}^{(i_2)}\right) \\
+ \frac{4}{n^2\,\sigma_0^4\,\omega_n^4}\,\mathbb{E}\left(\xi_t^2\right)\,\mathbb{E}\left(\bar{X}_{t-k_1}\,\bar{X}_{t-k_2}\right) \\
+ \frac{4}{n^4\,\sigma_0^4\,\omega_n^4}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_1)}\,\mathring{\varepsilon}_t^{(i_2)}\right)\,\mathbb{E}\left(\mathring{X}_{t-k_1}^{(i_1)}\,\mathring{X}_{t-k_2}^{(i_2)}\right) \times \mathbb{E}\left(-\xi_t^2/\omega_n^2 + 1\right)^2 \\
\times \Bigg[\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_3)\,2}\,\mathring{\varepsilon}_t^{(i_4)\,2}\right)/\sigma_0^4 - 2\,\frac{n-1}{n}\,\mathbb{E}\left(\mathring{\varepsilon}_t^{(i_3)\,2}\right)/\sigma_0^2 + \left(\frac{n-1}{n}\right)^2 \Bigg] \\
+ \frac{4}{n^6\,\omega_n^8} \left[\mathbb{E}\,\xi_t^6/\omega_n^4 - 2\,\mathbb{E}\,\xi_t^4/\omega_n^2 + \mathbb{E}\,\xi_t^2\right] \times \mathbb{E}\left(\bar{X}_{t-k_1}\,\bar{X}_{t-k_2}\right) \Bigg]
\end{aligned}$$

$$+ 6 \sum_{k,l=1}^{p} \lambda_k \, \lambda_l \, \lambda_{p+2}^2 \Big[\frac{4}{n \, \sigma_0^4 \, \omega_n^4} \, \mathbb{E} \left(\mathring{\varepsilon}_t^{(i_1)} \, \mathring{\varepsilon}_t^{(i_2)} \right) \, \mathbb{E} \left(\mathring{X}_{t-k_1}^{(i_1)} \, \mathring{X}_{t-k_2}^{(i_2)} \right) \, \mathbb{E} \left(-\xi_t^2 / \omega_n^2 + 1 \right)^2$$

$$+ \frac{4}{n^3 \, \omega_n^8} \left[\mathbb{E} \, \xi_t^6 / \omega_n^4 - 2 \, \mathbb{E} \, \xi_t^4 / \omega_n^2 + \mathbb{E} \, \xi_t^2 \right] \times \mathbb{E} \left(\bar{X}_{t-k_1} \, \bar{X}_{t-k_2} \right) \Big]$$

$$+ 6 \, \lambda_{p+1}^2 \, \lambda_{p+2}^2 \Big[\frac{1}{n \, \sigma_0^4 \, \omega_n^4} \Big[\mathbb{E} \left(\mathring{\varepsilon}_t^{(i_3)\,2} \, \mathring{\varepsilon}_t^{(i_4)\,2} \right) / \sigma_0^4 - 2 \, \frac{n-1}{n} \, \mathbb{E} \left(\mathring{\varepsilon}_t^{(i_3)\,2} \right) / \sigma_0^2 + \left(\frac{n-1}{n} \right)^2 \Big]$$

$$\times \, \mathbb{E} \left(-\xi_t^2 / \omega_n^2 + 1 \right)^2$$

$$+ \frac{1}{n^5 \, \omega_n^8} \, \mathbb{E} \left(-\xi_t^2 / \omega_n^2 + 1 \right)^4 \Big]$$

$$+ 4 \, \lambda_{p+1}^3 \, \lambda_{p+2} \Big[\frac{-3}{n^2 \sqrt{n} \, \sigma_0^4 \, \omega_n^4} \Big[\mathbb{E} \left(\mathring{\varepsilon}_t^{(i_1)\,2} \, \mathring{\varepsilon}_t^{(i_2)\,2} \right) / \sigma_0^4 - 2 \, \frac{n-1}{n} \, \mathbb{E} \left(\mathring{\varepsilon}_t^{(i_1)\,2} \right) / \sigma_0^2 + \left(\frac{n-1}{n} \right)^2 \Big]$$

$$\times \, \mathbb{E} \left(-\xi_t^2 / \omega_n^2 + 1 \right)^2$$

$$+ \frac{1}{n^6 \sqrt{n} \, \omega_n^8} \, \mathbb{E} \left(-\xi_t^2 / \omega_n^2 + 1 \right)^4 \Big]$$

$$+ 4 \, \lambda_{p+1} \, \lambda_{p+2}^3 \, \frac{1}{n^3 \sqrt{n} \, \omega_n^8} \, \mathbb{E} \left(-\xi_t^2 / \omega_n^2 + 1 \right)^4 \Big)$$

$$= \frac{1}{n^2 \, (T-p)^2} \Big(\sum_{k_1,k_2,l_1,l_2=1}^{p} \lambda_{k_1} \, \lambda_{k_2} \, \lambda_{l_1} \, \lambda_{l_2} \, 16$$

$$\Big[\left(\psi(k_1 - k_2) \, \psi(l_1 - l_2) + \psi(k_1 - l_1) \, \psi(k_2 - l_2) + \psi(k_1 - l_2) \, \psi(k_2 - l_1) \right)$$

$$\times \left(\left[(n-1)^2 + 2 \, (n-1) \right] + \frac{n^4}{n^4} \right)$$

$$+ \frac{6 \times 16}{n^2} \, \psi(k_1 - k_2) \, \psi(l_1 - l_2) \, n^2 \, (n-1) \Big]$$

$$+ \lambda_{p+1}^4 \Big[\frac{1}{\sigma_0^8} \Big(\left[n^4 + 8 \, n^3 + 14 \, n^2 - 8 \, n - 15 \right] - 4 \, \frac{n-1}{n} \, n \, (n^3 + 3 \, n^2 - n - 3)$$

$$+ 6 \left(\frac{n-1}{n} \right)^2 n^2 \, (n^2 - 1) - 4 \left(\frac{n-1}{n} \right)^4 n^4 + \left(\frac{n-1}{n} \right)^4 n^4 \Big)$$

$$+ \frac{6}{n^2 \, \sigma_0^4 \, \omega_n^4} \, n^2 \left([n^2 - 1] - \left(\frac{n-1}{n} \right)^2 n^2 \right) \times 2 + \frac{60 \, n^4}{n^4 \, \omega_n^8} \Big]$$

$$+ \lambda_{p+2}^4 \, \frac{60 \, n^4}{n^2 \, \omega_n^8}$$

$$+ 6 \sum_{k,l=1}^{p} \lambda_k \, \lambda_l \, \lambda_{p+1}^2 \, \psi(k-l) \, 4 \Big[\frac{1}{\sigma_0^4} \Big((n-1) \, (n^2 + 4 \, n + 3)$$

$$- 2 \, n \, \frac{n-1}{n} \, (n^2 - 1) + n^2 \left(\frac{n-1}{n} \right)^2 (n-1) \Big)$$

$$+ \frac{1}{n^2 \, \sigma_0^4} \, n^2 \left[(n^2 - 1) - n^2 \left(\frac{n-1}{n} \right)^2 \right] + \frac{1}{n^4 \, \omega_n^4} \, n^2 \, (n-1) \times 2 + \frac{n^4}{n^6 \, \omega_n^4} \, 10 \Big]$$

$$
\begin{aligned}
&+ 6 \sum_{k,l=1}^{p} \lambda_k \, \lambda_l \, \psi(k-l) \, \lambda_{p+2}^2 \, \big[\frac{4\,n^2}{n\,\omega_n^4}\,(n-1) \times 2 + \frac{4\,n^4}{n^3\,\omega_n^4}\,10\big] \\
&+ 6 \, \lambda_{p+1}^2 \, \lambda_{p+2}^2 \left[\frac{n^2}{n\,\sigma_n^4\,\omega_n^4}\left[(n^2-1)-(n-1)^2\right] \times 2 + \frac{60\,n^4}{n^5\,\omega_n^8}\right] \\
&+ 4 \, \lambda_{p+1}^3 \, \lambda_{p+2} \Big(\frac{-3\,n^2}{n^2\,\sqrt{n}\,\sigma_0^4\,\omega_n^4}\,((n^2-1)-(n-1)^2) \times 2 + \frac{60\,n^4}{n^6\,\sqrt{n}\,\omega_n^8}\Big) \\
&+ 4 \, \lambda_{p+1} \, \lambda_{p+2}^3 \frac{60\,n^4}{n^3\,\sqrt{n}\,\omega_n^8} \\
= {} & \frac{1}{n^2\,(T-p)^2}\Big[3\left(\sum_{k,l=1}^{p} \lambda_k \, \lambda_l \, \psi(k-l)\right)^2 \big(n^2 + 2\,(n-1)\big) \\
&+ 12 \, \lambda_{p+1}^4 \left(\frac{1}{\sigma_0^8}\,\big(n^2+2\,(n-1)\big) + \frac{1}{\sigma_0^4\,\omega_n^4}\,(n-1) + \frac{5}{\omega_n^8}\right) + 60 \, \lambda_{p+2}^4 \, \frac{1}{\omega_n^8}\, n^2 \\
&+ 48 \sum_{k,l=1}^{p} \lambda_k \, \lambda_l \, \lambda_{p+1}^2 \, \psi(k-l) \left(\frac{1}{\sigma_0^4}\,(n-1)\,(n+4) + \frac{1}{\omega_0^4}\,\frac{n+4}{n^2}\right) \\
&+ 48 \sum_{k,l=1}^{p} \lambda_k \, \lambda_l \, \lambda_{p+2}^2 \, \psi(k-l) \frac{1}{\omega_n^4}\, n\,(n+4) \\
&+ 12 \, \lambda_{p+1}^2 \, \lambda_{p+2}^2 \left(\frac{1}{\sigma_n^4\,\omega_n^4}\, n\,(n-1) + \frac{30}{\omega_n^8}\,\frac{1}{n}\right) \\
&+ 48 \, \lambda_{p+1}^3 \, \lambda_{p+2} \left(\frac{-1}{\sigma_n^4\,\omega_n^4}\,\frac{n-1}{\sqrt{n}} + \frac{5}{\omega_n^8}\,\frac{1}{n^2\,\sqrt{n}}\right) + 240 \, \lambda_{p+1} \, \lambda_{p+2}^3 \, \frac{1}{\omega_n^8}\,\sqrt{n}\ \Big] \\
= {} & O\left(\frac{1}{T^2}\right). \qquad \square
\end{aligned}
$$

C.2.4 REMARK

A second possibility for proving the above statement is to distinguish between the two cases $n_T \to \infty$ and n fixed. In the latter case it is clear that $\mathbb{E}\, D_{T,t,\lambda}^4$ is bounded. If $n \to \infty$, we can again employ the approach used for the case $n \to \infty$, T fixed: we define a new random vector by omitting the terms depending on $\{\bar{X}_t\}_{t\in\mathbb{Z}}$ from the first $p+1$ coordinates of the gradient vector. Defining martingale differences for the new vector in a similar way as it has been done in lemma 2.5.30, we can again prove proposition 2.5.31, but the length of the calculations is reduced. The considerations used in the case $n \to \infty$, T fixed, show that the gradient has asymptotically the same distribution as the new vector.

C.3 Proofs for Section 2.6

C.3.1 Rates of Convergence

In this section we derive the rates of convergence of several terms which later on are used for computing the rate of convergence of $\hat{a}$. The proofs are straightforward but have been excluded from section 2.6 in order to enhance readability.

PROOF OF LEMMA 2.6.3:
By assumption Θ is such that we have for all $\theta = (a', \sigma^2, \tau^2)' \in \Theta$ that $\sigma^2 \geq c > 0$. Thus $\frac{1}{\hat\sigma^2} \leq \frac{1}{c}$ and $\frac{1}{n\,\hat\omega_n^2} \leq \frac{1}{c}$. Therefore we get in any case that $\frac{1}{\hat\sigma^2} = O_P(1)$ and $\frac{1}{n\,\hat\omega_n^2} = O_P(1)$. If $\tau_0^2 > 0$, then $\omega_n^2 > 0$ and the consistency of $\hat\omega_n^2$ yields that already $\frac{1}{\hat\omega_n^2} = O_P(1)$. This can be summarised to $\frac{1}{n\,\hat\omega_n^2} = O_P\left(\frac{1}{n\,\omega_n^2}\right)$. The rates of convergence given in lemma 2.6.2 imply that $\frac{1}{\hat\sigma^2} - \frac{1}{\sigma_0^2} = \frac{\sigma_0^2 - \hat\sigma^2}{\hat\sigma^2\,\sigma_0^2} = O_P\left(\frac{1}{\sqrt{nT}}\right)$ and analogously $\frac{1}{n\,\hat\omega_n^2} - \frac{1}{n\,\omega_n^2} = O_P\left(\frac{1}{n\,\omega_n^2}\right) \times O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right) \times \frac{1}{\omega_n^2} = O_P\left(\frac{1}{n\,\omega_n^2\,\sqrt{T}}\right)$.
For proving the second assertion denote $\hat B_1 = \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{\mathbf{x}}_{t-1}^{(i)}\, \mathring{\mathbf{x}}_{t-1}^{(i)\prime}$ and $\hat B_2 = \frac{1}{T-p} \sum_{t=p+1}^{T} \bar{\mathbf{x}}_{t-1}\, \bar{\mathbf{x}}_{t-1}'$ (notation as in remark 2.6.1). Then

$$\hat B - B = \frac{1}{\sigma^2}\left(\hat B_1 - \frac{(n-1)\,\sigma^2}{n}\,B\right) + \left(\frac{1}{\hat\sigma^2} - \frac{1}{\sigma^2}\right)\hat B_1 + \frac{1}{n\,\omega_n^2}\left(\hat B_2 - \omega_n^2\,B\right) + \left(\frac{1}{n\,\hat\omega_n^2} - \frac{1}{n\,\omega_n^2}\right)\hat B_2\,.$$

Due to the mean-square convergence of the panel autocovariance estimator (compare e.g. remark 2.6.1), we see directly that the first term in this expression is of order $O_P\left(\frac{1}{\sqrt{n\,T}}\right)$ and that $\hat B_2 - \omega_n^2\,B = O_P\left(\frac{\omega_n^2}{\sqrt{T}}\right)$. Thus the third term is of order $O_P\left(\frac{1}{n\,\sqrt{T}}\right)$, independent of $\tau_0^2 > 0$ or $\tau_0^2 = 0$. Moreover the mean-square convergence yields that $\hat B_1 = O_P(1)$ and $\hat B_2 = O_P\left(\omega_n^2\right)$. Altogether we obtain that

$$\hat B - B = O_P\left(\frac{1}{\sqrt{n\,T}}\right)$$

due to the rates of convergence of the different parameter estimators. □

The second result states the rates of convergence of $\hat C_1$ and $\hat C_2$. We have omitted these calculations in the proof of theorem 2.6.5.

C.3.1 LEMMA
In the setting of theorem 2.6.5, we have

$$\frac{1}{\hat\sigma^2}\,\hat C_1 = O_P\left(\frac{1}{\sqrt{n\,T}}\right) \qquad \textit{and} \qquad \frac{1}{n\,\hat\omega_n^2}\,\hat C_2 = O_P\left(\frac{1}{n\,\sqrt{T}}\right)\,.$$

PROOF:
By assumption $\mathring{X}_s^{(i)}$ and $\mathring{\varepsilon}_t^{(i)}$ are independent for $s < t$. Furthermore lemma 2.2.4 states that

$$\operatorname{cov}\left(\mathring{X}_{t-k}^{(i)}, \mathring{X}_{t-l}^{(j)}\right) = \Psi(k-l)\operatorname{cov}\left(\mathring{\varepsilon}_t^{(i)}, \mathring{\varepsilon}_t^{(j)}\right) = \left(\delta_{ij} - \frac{1}{n}\right)\sigma_0^2\,\Psi(k-l)\,,$$

where $\sigma_0^2 = \operatorname{var}\varepsilon_t^{(i)}$. Denote the kth entry of $\hat C_1$ by $\hat C_{1,k}$ and of $\hat C_2$ by $\hat C_{2,k}$. Then we have for all $k = 1,\ldots,p$ that

$$\mathbb{E}\,\hat C_{1,k}^2 = \mathbb{E}\left(\frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{X}_{t-k}^{(i)}\, \mathring{\varepsilon}_t^{(i)}\right)^2$$

$$
\begin{aligned}
&= \frac{1}{n^2\,(T-p)^2} \sum_{s,t=p+1}^{T} \sum_{i,j=1}^{n} \mathbb{E}\left(\mathring{X}_{s-k}^{(i)}\, \mathring{\varepsilon}_s^{(i)}\, \mathring{X}_{t-k}^{(j)}\, \mathring{\varepsilon}_t^{(j)} \right) \\
&= \frac{1}{n^2\,(T-p)^2} \sum_{s,t=p+1}^{T} \delta_{st} \left(n \left(\frac{n-1}{n} \right)^2 + n\,(n-1)\,\frac{1}{n^2} \right) \sigma_0^4\, \Psi(0) \\
&= \frac{n-1}{n^2\,(T-p)}\, \sigma_0^4\, \Psi(0)\,.
\end{aligned}
$$

Analogously, we get from lemma 2.2.4 that for all $k = 1, \ldots, p$

$$
\mathbb{E}\, \hat{C}_{2,k}^2 = \frac{1}{(T-p)}\, \bar{\gamma}_n(0)\, \omega_n^2 = \frac{1}{(T-p)}\, \omega_n^4\, \Psi(0)\,.
$$

As we have shown in corollary 2.6.3 that $\frac{1}{\hat{\sigma}^2} = O_P(1)$ and $\frac{1}{n\,\hat{\omega}_n^2} = O_P\left(\frac{1}{n\,\omega_n^2}\right)$, we directly obtain that

$$
\frac{1}{\hat{\sigma}^2}\, \hat{C}_1 = O_P\left(\frac{1}{\sqrt{n\,T}} \right) \quad \text{and} \quad \frac{1}{n\,\hat{\omega}_n^2}\, \hat{C}_2 = O_P\left(\frac{1}{n\,\sqrt{T}} \right). \qquad \square
$$

C.3.2 Some Remarks on Cumulants

We need some results on cumulants for determining the mean-square rates of convergence of the bias terms $\hat{\beta}_1$ and $\hat{\beta}_2$ in the GICM. More precisely, we want to get bounds for cumulants of compound processes of the type $Y_{r,s}^{(i)}(t) = X_{t-r}^{(i)}\, X_{t-s}^{(i)}$, $t \in \mathbb{Z}$, where $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, are causal Gaussian autoregressive time series as described in section 1.1. For Gaussian processes all cumulants of order three and higher are zero. Therefore all higher order cumulants of the compound processes $Y_{r,s}^{(i)}(t)$, $t \in \mathbb{Z}$, $i = 1, \ldots, n$, $0 \le r, s \le p$, can be reduced to functions of second order cumulants. Thus all subsequent proofs are based on the following property:

C.3.2 THEOREM
Let $\{X_t^{(i)}\}_{t\in\mathbb{Z}}$, $i = 1, \ldots, n$, be a panel of stationary Gaussian autoregressive time series as in assumption 1.2.1 such that $\operatorname{cov}(X_s^{(i)}, X_t^{(j)}) = u_n \operatorname{cov}(X_s^{(i)}, X_t^{(i)})$ for $i \neq j$ with $u_n = O\left(\frac{1}{n}\right)$. Further let $0 \le r_k, s_k \le p$ for all $k = 1, \ldots, m$.
Then, identifying $m + 1 \equiv 1$, we get

$$
\sum_{i_1,\ldots,i_m=1}^{n} \sum_{t_1,\ldots,t_m=p+1}^{T} \prod_{k=1}^{m} \mathbb{E}\left(X_{t_k - r_k}^{(i_k)}\, X_{t_{k+1}-s_{k+1}}^{(i_{k+1})} \right) = O\left((n\,T)^{m-2} \right)
$$

for all $m \ge 3$.

PROOF:
By assumption, $\operatorname{cov}(X_s^{(i)}, X_t^{(i)}) = \gamma_n(s-t)$ and $\operatorname{cov}(X_s^{(i)}, X_t^{(j)}) = u_n\, \gamma_n(s-t)$ for $i \neq j$. Due to lemma 1.1.2, each $X_t^{(i)}$, $i = 1, \ldots, n$, admits a MA(∞) representation such that

$$
X_t^{(i)} = \sum_{u=0}^{\infty} \psi_u\, \varepsilon_{t-u}^{(i)}\,,
$$

where for each $i = 1, \ldots, n$ the $\{\varepsilon_t^{(i)}\}_{t\in\mathbb{Z}}$ are independently and identically distributed with $\mathbb{E}\,\varepsilon_t^{(i)} = 0$ and $\operatorname{var}\varepsilon_t^{(i)} = \sigma^2$. This means that $\gamma_n(h) = \sum_{u=0}^{\infty} \psi_u\,\psi_{u+|h|}\,\sigma^2$ for all $h \in \mathbb{Z}$. Because of proposition 1.1.3 we know that $|\psi_u| \le c\,\rho^u$ for all $u > 0$, where $c > 0$ and $0 < \rho < 1$.

For ease of notation let $\xi_k = r_k - s_{k+1}$ and $d_k = t_k - t_1$ for $k = 1, \ldots, m$. This implies that $-(p-1) \le \xi_k \le p-1$ and $-(T-p-1) \le d_k \le T-p-1$ for all $k = 1, \ldots, m$. We therefore get

$$
\begin{aligned}
&\sum_{i_1,\ldots,i_m=1}^{n} \sum_{t_1,\ldots,t_m=p+1}^{T} \prod_{k=1}^{m} \mathbb{E}\left(X_{t_k-r_k}^{(i_k)}\, X_{t_{k+1}-s_{k+1}}^{(i_{k+1})}\right) \\
&\quad= \sum_{i_1,\ldots,i_m=1}^{n} \sum_{t_1,\ldots,t_m=p+1}^{T} \prod_{k=1}^{m} \left(\delta_{i_k,i_{k+1}} - (1-\delta_{i_k,i_{k+1}})\,u_n\right) \\
&\qquad\qquad\qquad\qquad \times \gamma_n(t_k - r_k - t_{k+1} + s_{k+1}) \\
&\quad\le \sum_{i_1,\ldots,i_m=1}^{n} \sum_{t_1,\ldots,t_m=p+1}^{T} \sum_{u_1,\ldots,u_m=0}^{\infty} \prod_{k=1}^{m} c^2 \left(\delta_{i_k,i_{k+1}} - (1-\delta_{i_k,i_{k+1}})\,u_n\right) \\
&\qquad\qquad\qquad\qquad \times \rho^{u_k}\,\rho^{u_k+|t_k-r_k-t_{k+1}+s_{k+1}|}\,\sigma^2 \\
&\quad= \tilde{c}_n \sum_{t_1,\ldots,t_m=p+1}^{T} \prod_{k=1}^{m} \rho^{|t_k-t_{k+1}-\xi_k|} \\
&\quad\le \tilde{c}_n \sum_{t=p+1}^{T} \sum_{d_2,\ldots,d_m=-(T-p-1)}^{T-p-1} \rho^{|d_2+\xi_1|}\,\rho^{|d_m-\xi_m|} \prod_{k=2}^{m-1} \rho^{|d_k-d_{k+1}-\xi_k|} \\
&\quad\le \tilde{c}_n \cdot T \cdot (2T-2p-1)^{m-3} \sum_{d_2,d_m=-T}^{T} \rho^{|d_2|}\,\rho^{|d_m|} \\
&\quad= O\left(T^{m-2}\right),
\end{aligned}
$$

For the last inequality we have used that $|\rho| < 1$. Furthermore we have enlarged the sum in order to take the presence of the ξ_k, $k = 1, \ldots, m$, into account.

The constant $\tilde{c}_n$ is given by

$$
\tilde{c}_n = \sigma^{2m} \left(\frac{c^2}{1-\rho^2}\right)^m \sum_{i_1,\ldots,i_m=1}^{n} \prod_{k=1}^{m} \left(\delta_{i_k,i_{k+1}} - (1-\delta_{i_k,i_{k+1}})\,u_n\right) .
$$

If $u_n = O\left(\frac{1}{n}\right)$, then

$$
\sum_{i_1,\ldots,i_m=1}^{n} \prod_{k=1}^{m} \left(\delta_{i_k,i_{k+1}} - (1-\delta_{i_k,i_{k+1}})\,u_n\right) = O\left(n^{m-2}\right)
$$

as the product equals 1 if and only if $i_1 = \cdots = i_m$. Otherwise it is of order $O\left(u_n^2\right)$, because then at least two pairs in $\{(i_k, i_{k+1}), k = 1, \ldots, m\}$ must fulfil that $i_k \ne i_{k+1}$. This completes the proof. □

In the above proof, we just have used the fact that $\prod_{k=2}^{m-1} \rho^{|d_k-d_{k+1}+\xi_k|}$ is bounded by 1. Since the factors in this term are interconnected, an explicit calculation does not lead to a further reduction of the order. We also can see from the proof that we get the analogous result, of order $O(T^m)$, if we constrain ourselves to the fact that the MA(∞) coefficients are absolutely summable.

In order to simplify the notation, we deduce from the above result an analogous statement on cumulants.

C.3.3 LEMMA
In the setting of the above theorem let $Y^{(i_k)}_{r_k,s_k}(t) = X^{(i_k)}_{t-r_k}\, X^{(i_k)}_{t-s}$, $t \in \mathbb{Z}$.
Then for $m \geq 3$

$$\sum_{i_1,\dots,i_m=1}^{n} \sum_{t_1,\dots,t_m=p+1}^{T} \operatorname{cum}\left(Y^{(i_1)}_{r_1,s_1}(t_1),\dots,Y^{(i_m)}_{r_m,s_m}(t_m)\right) = O\left((n\,T)^{m-2}\right) .$$

PROOF:
For ease of notation, denote $X^{(i_l)}_{t_l-j}$ by $X_{l,j}$. Theorem 2.3 of Brillinger (1981) gives

$$\begin{aligned}\sum_{t_1,\dots,t_m=p+1}^{T} & \operatorname{cum}\left(Y^{(i_1)}_{r_1,s_1}(t_1),\dots,Y^{(i_m)}_{r_m,s_m}(t_m)\right) \\ &= \sum_{t_1,\dots,t_m=p+1}^{T} \sum_{\nu_1+\dots+\nu_q=\nu} \operatorname{cum}\left(X_{l,j};(l,j)\in\nu_1\right)\dots\operatorname{cum}\left(X_{l,j};(l,j)\in\nu_q\right) ,\end{aligned}$$

where the second summation runs over all indecomposable partitions of
$\nu = \{(l,j_l), l\in\{1,\dots,m\}, j_l\in\{r_l,s_l\}\}$.
As the processes $\{X^{(i)}_t\}_{t\in\mathbb{Z}},\ i = 1,\dots,n$, are Gaussian, all cumulants of order larger than two are zero (Shiryayev 1984, p. 291). Thus if we identify $m+1 \equiv 1$, the remaining partitions are of the form $\nu_k = \{(k,r_k),(k+1,s_{k+1}); k = 1,\dots,m\}$. This means that we sum over products of covariances which are of the form needed for the preceding lemma. □

Up to here, we have derived bounds for $m \geq 3$. The case $m = 2$ has to be treated separately.

C.3.4 LEMMA
Under the assumptions of theorem C.3.2,

$$\begin{aligned}\sum_{i_1,i_2=1}^{n} \sum_{t_1,t_2=p+1}^{T} & \operatorname{cum}\left(X^{(i_1)}_{t_1-r_1}\, X^{(i_1)}_{t_1-s_1}, X^{(i_2)}_{t_2-r_2}\, X^{(i_2)}_{t_2-s_2}\right) \\ &= \sum_{i_1,i_2=1}^{n} \sum_{t_1,t_2=p+1}^{T} \Big[\mathbb{E}\left(X^{(1)}_{t_1-r_1}\, X^{(2)}_{t_2-s_1}\right)\mathbb{E}\left(X^{(1)}_{t_1-r_2}\, X^{(2)}_{t_2-s_2}\right) \\ &\qquad\qquad + \mathbb{E}\left(X^{(1)}_{t_1-r_1}\, X^{(2)}_{t_2-s_2}\right)\mathbb{E}\left(X^{(1)}_{t_1-r_2}\, X^{(2)}_{t_2-s_1}\right)\Big] \\ &= O(nT) .\end{aligned}$$

PROOF:
Straightforward calculation gives

$$\sum_{i_1,i_2=1}^{n} \sum_{t_1,t_2=p+1}^{T} \mathbb{E}\left(X_{t_1-r_1}^{(1)}\, X_{t_2-s_1}^{(2)}\right)\, \mathbb{E}\left(X_{t_1-r_2}^{(1)}\, X_{t_2-s_2}^{(2)}\right)$$
$$\leq \left(n + u_n^2\, n\,(n-1)\right) c^2\, \sigma^4 \sum_{h=-(T-p-1)}^{T-p-1} \sum_{u,v=0}^{\infty} \rho^u\, \rho^{u+|h-r_1+s_1|}\, \rho^v\, \rho^{v+|h-r_2+s_2|}$$
$$\leq \left(n + u_n^2\, n\,(n-1)\right) c^2\, \sigma^4\, \frac{1}{(1-\rho^2)^2}\,(2\,T - 2\,p - 1) = O(nT)$$

if $u_n = O\left(\frac{1}{n}\right)$. As the first equality of the lemma's statement is due to theorem 2.3 of Brillinger (1981), this already implies the result. □

Now we are in the position to prove lemma 2.6.12.

PROOF OF LEMMA 2.6.12:
Denote the entries of B^{-1} by $b_{k,l}$, $k, l = 1, \dots, p$, and let $Y_{g,h}^{(i)}(t) = \mathring{X}_{t-g}^{(i)}\, \mathring{X}_{t-h}^{(i)}$. Furthermore recall (see the proof of theorem 2.6.5) that the kth entry of $\hat{C}_1$ fulfils

$$\hat{C}_{1,k} = \left(\hat{A}_1 - \hat{B}_1\, a\right)_k = -\sum_{l=0}^{p} a_l\, \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{n} \mathring{X}_{t-l}^{(i)}\, \mathring{X}_{t-k}^{(i)}$$

if we denote $a_0 = -1$. Therefore the mth entry of $\hat{\beta}_1$ is

$$\hat{\beta}_{1,m} = \sum_{g=1}^{p} b_{m,g} \sum_{h=1}^{p} \left[\frac{1}{n\,(T-p)} \sum_{s=p+1}^{T} \sum_{i=1}^{n} Y_{g,h}^{(i)}(s) - \mathring{\gamma}_n(g-h)\right]$$
$$\times \sum_{k=1}^{p} b_{h,k} \sum_{l=0}^{p} \frac{1}{n\,(T-p)} \sum_{t=p+1}^{T} \sum_{j=1}^{n} a_l\, Y_{k,l}^{(j)}(t)\,.$$

As the $\mathring{X}_t^{(i)}$, $t \in \mathbb{Z}$, $i = 1, \dots, n$, are Gaussian with $\operatorname{cov}(\mathring{X}_s^{(i)}, \mathring{X}_t^{(i)}) = \mathring{\gamma}_n(s-t)$ and $\operatorname{cov}(\mathring{X}_s^{(i)}, \mathring{X}_t^{(j)}) = u_n\, \mathring{\gamma}_n(s-t)$ for $i \neq j$, theorem 2.3 of Brillinger (1981) gives

$$\mathbb{E}\left(Y_{k_1,l_1}^{(j_1)}(t_1)\, Y_{k_2,l_2}^{(j_2)}(t_2)\right) = \mathring{\gamma}_n(k_1-l_1)\, \mathring{\gamma}_n(k_2-l_2)$$
$$+ \left(\delta_{j_1,j_2} - (1-\delta_{j_1,j_2})\, u_n^2\right) \Big[\mathring{\gamma}_n(t_1-t_2-k_1+k_2)\, \mathring{\gamma}_n(t_1-t_2-l_1+l_2)$$
$$+ \mathring{\gamma}_n(t_1-t_2-k_1+l_2)\, \mathring{\gamma}_n(t_1-t_2+k_2-l_1)\Big]\,.$$

Moreover we have by assumption that $k_1 > 0$, $k_2 > 0$ and thus also $k_1 + k_2 > 0$. As (see lemma 2.2.4) $\mathring{\gamma}_n(h) = \frac{n-1}{n}\, \Psi(h)\, \sigma_0^2$, we get using the relations given in proposition 1.1.6 that

$$\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2}\, \mathring{\gamma}_n(k_1-l_1)\, \mathring{\gamma}_n(k_2-l_2) = 0\,,$$

$$\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \sum_{t_1,t_2=p+1}^{T} \mathring{\gamma}_n(t_1-t_2-k_1+l_2)\,\mathring{\gamma}_n(t_1-t_2+k_2-l_1) = 0 \quad \text{and}$$

$$\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \sum_{t_1,t_2=p+1}^{T} \mathring{\gamma}_n(t_1-t_2-k_1+k_2)\,\mathring{\gamma}_n(t_1-t_2-l_1+l_2) = \frac{(T-p)(n-1)}{n}\,\sigma_0^2\,\mathring{\gamma}_n(k_1-k_2)\,.$$

Therefore

$$\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \sum_{t_1,t_2=p+1}^{T} \sum_{j_1,j_2=1}^{n} \mathbb{E}\left(Y_{k_1,l_1}^{(j_1)}(t_1)\, Y_{k_2,l_2}^{(j_2)}(t_2)\right) = \frac{(T-p)(n-1)}{n}\,\sigma_0^2\,\left(n+u_n^2\, n\,(n-1)\right)\mathring{\gamma}_n(k_1-k_2)\,.$$

Thus $\mathbb{E}\,\hat{\beta}_{1,m}^2$ reduces to

$$\mathbb{E}\,\hat{\beta}_{1,m}^2 = \sum_{g_1,g_2,h_1,h_2,k_1,k_2=1}^{p} b_{g,h,k}\Big[\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \frac{1}{n^4\,(T-p)^4} \times \sum_{s_1,s_2,t_1,t_2=p+1}^{T} \sum_{i_1,i_2,j_1,j_2=1}^{n} \mathbb{E}\left(Y_{g_1,h_1}^{(i_1)}(s_1)\, Y_{g_2,h_2}^{(i_2)}(s_2)\, Y_{k_1,l_1}^{(j_1)}(t_1)\, Y_{k_2,l_2}^{(j_2)}(t_2)\right)\Big] - \frac{(n-1)\,(1+u_n^2\,(n-1))}{n^2\,(T-p)}\,\sigma_0^2\,\mathring{\gamma}_n(g_1-h_1)\,\mathring{\gamma}_n(g_2-h_2)\,\mathring{\gamma}_n(k_1-k_2)\,.$$

where the constant is $b_{g,h,k} = b_{m,g_1}\, b_{m,g_2}\, b_{h_1,k_1}\, b_{h_2,k_2}$.
We now investigate the first term of this expression. As Leonov and Shiryayev (1959) have shown, the expectation of a product of random variables can be represented as a sum of cumulants of smaller or equal order. The formula also can be found in (Shiryayev 1984, p. 293). We have derived the order of the cumulants for $m \geq 2$ in theorem C.3.3 and lemma C.3.4. Thus for all $0 \leq r_i, s_i \leq p$, $i = 1, \ldots, 4$,

$$\frac{1}{n^4\,(T-p)^4} \sum_{i_1,i_2,i_3,i_4=1}^{n} \sum_{t_1,t_2,t_3,t_4=p+1}^{T} \operatorname{cum}\left(Y_{r_1,s_1}^{(i_1)}(t_1), Y_{r_2,s_2}^{(i_2)}(t_2), Y_{r_3,s_3}^{(i_3)}(t_3), Y_{r_4,s_4}^{(i_4)}(t_4)\right) = O\left(\frac{1}{n^2\,(T-p)^2}\right)$$

$$\frac{1}{n^3\,(T-p)^3} \sum_{i_1,i_2,i_3=1}^{n} \sum_{t_1,t_2,t_3=p+1}^{T} \operatorname{cum}\left(Y_{r_1,s_1}^{(i_1)}(t_1), Y_{r_2,s_2}^{(i_2)}(t_2), Y_{r_3,s_3}^{(i_3)}(t_3)\right) = O\left(\frac{1}{n^2\,(T-p)^2}\right)$$

and $$\frac{1}{n^4\,(T-p)^4}\Big(\sum_{i_1,i_2=1}^{n} \sum_{t_1,t_2=p+1}^{T} \mathring{\gamma}_n(t_1-r_1-(t_2-s_2))\,\mathring{\gamma}_n(t_2-r_2-(t_1-s_1))\Big)^2 = O\left(\frac{1}{n^2\,(T-p)^2}\right)\,.$$

Furthermore we have due to the preceding proposition that

$$\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \sum_{t_3,t_4=p+1}^{T} \mathring{\gamma}_n(t_3-k_1-t_4+k_2)\,\mathring{\gamma}_n(t_3-l_1-t_4+l_2) = 0$$

and

$$\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \sum_{t_3,t_4=p+1}^{T} \mathring{\gamma}_n(t_3-k_1-t_4+l_2)\,\mathring{\gamma}_n(t_3-l_1-t_4+k_2) = \frac{(n-1)\,(T-p)}{n}\,\sigma_0^2\,\mathring{\gamma}_n(k_1-k_2)\,.$$

Thus the remaining terms, which are based on the second order cumulants, become

$$\begin{aligned}\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \sum_{t_1,t_2,t_3,t_4=p+1}^{T} &\mathring{\gamma}_n(t_3-k_1-(t_4-l_2))\,\mathring{\gamma}_n(t_4-k_2-(t_3-l_1)) \\ &\times \mathring{\gamma}_n(g_1-h_1)\,\mathring{\gamma}_n(g_2-h_2) \\ &= \frac{(n-1)\,(T-p)^3}{n}\,\sigma_0^2\,\mathring{\gamma}_n(g_1-h_1)\,\mathring{\gamma}_n(g_2-h_2)\,\mathring{\gamma}_n(k_1-k_2)\end{aligned}$$

and

$$\sum_{l_1,l_2=0}^{p} a_{l_1}\, a_{l_2} \sum_{t_1,t_2,t_3,t_4=p+1}^{T} \mathring{\gamma}_n(g_1-h_1)\,\mathring{\gamma}_n(g_2-h_2)\,\mathring{\gamma}_n(k_1-l_1)\,\mathring{\gamma}_n(k_2-l_2) = 0\,.$$

Altogether we obtain that

$$\begin{aligned}\mathbb{E}\,\hat{\beta}_{1,m}^2 = &\sum_{g_1,g_2,h_1,h_2,k_1,k_2=1}^{p} b_{g,h,k}\,\frac{(n-1)\,\left(1+u_n^2\,(n-1)\right)}{n^2\,(T-p)}\,\sigma_0^2 \\ &\times \Big(\mathring{\gamma}_n(g_1-h_1)\,\mathring{\gamma}_n(g_2-h_2)\,\mathring{\gamma}_n(k_1-k_2) \\ &\quad - \mathring{\gamma}_n(g_1-h_1)\,\mathring{\gamma}_n(g_2-h_2)\,\mathring{\gamma}_n(k_1-k_2)\Big) + O\left(\frac{1}{n^2\,(T-p)^2}\right) \\ = &\, O\left(\frac{1}{n^2\,(T-p)^2}\right)\,.\end{aligned}$$

The proof for $\hat{\beta}_{2,m}$ is analogous. As it only depends on n via $\bar{X}_t = \frac{1}{n}\sum_{i=1}^{n} X_t^{(i)}$, we here get $\hat{\beta}_{2,m} = O\left(\frac{1}{(T-p)^2}\right)$. □

Bibliography

Arellano, M. (2003). *Panel Data Econometrics*. Advanced Texts in Econometrics, Oxford University Press, Oxford.

Baltagi, B. H. (2001). *Econometric Analysis of Panel Data*. 2nd edn, Wiley & Sons, Chichester.

Basawa, I., Billard, L. and Srinivasan, R. (1984). Large-sample tests of homogeneity for time series models. *Biometrika* **71**(1), 203–206.

Basawa, I. and Koul, H. (1979). Asymptotically minimax tests of composite hypotheses for non-ergodic type stochastic processes. *Stoch. Proc. Appl.* **14**, 41–54.

Bauwens, L., Lutrano, M. and Richard, J. F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

Becker, C. and Gather, U. (1999). The masking breakdown point of multivariate outlier identification rules. *Journal of the American Statistical Association* **94**(447), 947–955.

Brillinger, D. (1973). The analysis of time series collected in experimental design. In P. Krishnaiah (ed.), *Multivariate Analysis III*, Academic Press, New York, pp. 241 – 256.

Brillinger (1981). *Time Series*. McGraw-Hill, New York.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods*. Springer, NY.

Bustos, O. H. (1982). General M-estimates for contaminated pth-order autoregressive processes: Consistency and asymptotic normality. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **59**, 491–504.

Bustos, O. H. and Yohai, V. (1986). Robust estimates for ARMA models. *Journal of the American Statistical Association* **75**, 602–608.

Congdon, P. (2004). *Applied Bayesian Modelling*. Wiley Series in Probability and Statistics, Wiley, Chichester.

Cox, D. and Solomon, P. (1988). On testing for serial correlation in large numbers of small samples. *Biometrika* **75**, 145–148.

Croux, C. and Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis* **71**, 161–190.

Croux, C. and Rousseeuw, P. J. (1992). Time-efficient algorithms for two highly robust estimators of scale. In Y. Dodge and J. Whittaker (eds), *Computational Statistics, Volume I*, Physica-Verlag, Heidelberg, pp. 411–428.

Cruddas, A., Reid, N. and Cox, D. (1989). A time series illustration of approximate conditional likelihood. *Biometrika* **76**, 231–237.

Dahlhaus, R. (1999). Faktorisierung der bedingten Likelioodfunktion. Private communication.

Dahlhaus, R. (2000a). Graphical interaction models for multivariate time series. *Metrika* **51**(2), 157–172.

Dahlhaus, R. (2000b). A likelihood approximation for locally stationary processes. *The Annals of Statistics* **28**(6), 1762–1794.

Dahlhaus, R. and Giraitis, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis* **19**(6), 629–655.

Davidson, J. (1994). *Stochastic Limit Theory*. Advanced Texts in Econometrics, Oxford University Press, New York.

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

Diggle, P., Liang, K. and Zeger, S. (1994). *Analysis of Longitudinal Data*. Oxford Science Publications, Oxford Univ. Press, NY.

Drton, M. and Eichler, M. (2004). Maximum likelihood estimation in gaussian chain graph models under the alternative markov property. Preprint.

Eich, W., Hartmann, M., Müller, A., Heindl, B. and Fischer, H. (1998). Integrated psychological and physical group therapy (IGTF) of 259 patients with FMS: A one-year follow-up. *Arthritis. Rheum.* **41(suppl.)**, 259.

Feiler, S., Müller, K., Müller, A., Dahlhaus, R. and Eich, W. (2005). Using interaction graphs for analysing the therapy process. *Psychotherapy and Psychosomatics* **74**(2), 93–99.

Ferretti, N., Kelmansky, D. and Yohai, V. (1991). Estimators based on ranks for ARMA models. *Comm. Statist. Theory Methods* **20**, 3879–3907.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized factor model: Identification and estimation. *The Review of Economics and Statistics* **82**, 540–554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2001). The generalized dynamic factor model: Consistency and rates. Preprint.

Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics* **69**, 393–414.

Fu, B., Li, W.-K. and Fung, W.-K. (2002). Testing model adequacy for dynamic panel data with intercorrelation. *Biometrika* **89**(3), 591–601.

Gather, U., Bauer, M. and Fried, R. (2002). The identification of multiple outliers in online monitoring data. *Estadística* **54**(162–163), 289–338.

Gather, U., Imhoff, M. and Fried, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Stat. Med.* **21**(18), 2685–2701.

Gentleman, R. and Ihaka, R. (2004). R. `http://cran.at.r-project.org/`.

Haddad, J. N. (2000). On robust estimation in the first order autoregressive process. *Commun. Statist. – Theory Meth.* **29**(1), 45–54.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York, London.

Hallin, M. and Puri, M. L. (1988). Optimal rank-based procedures for time series analysis: Testing an ARMA-model against other ARMA-models. *The Annals of Statistics* **16**(1), 402–432.

Hampel, F. (1973). Robust estimation, a condensed partial survey. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **27**, 87–104.

Hjellvik, V. and Tjøstheim, D. (1999a). Modelling panels of intercorrelated autoregressive time series. *Biometrika* **86**(3), 573–590.

Hjellvik, V. and Tjøstheim, D. (1999b). Residual variance estimates and order determination in panels of intercorrelated autoregressive time series. In R. Engle and H. White (eds), *Cointegration, Causality and Forecasting, Festschrift in Honour of Clive Granger*, Oxford University Press, Oxford, pp. 385–409.

Hjellvik, V., Chen, R. and Tjøstheim, D. (2004). Nonparametric estimation and testing in panels of intercorrelated time series. *Journal of Time Series Analysis* **25**(6), 831–872.

Holtz-Eakin, D., Newey, W. and Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica* **75**, 1371–1395.

Hsiao, C. (1986). *Analysis of Panel Data*. Econometric Society Monographs, Cambridge University Press, Cambridge.

Huber, P. J. (1981). *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics, Wiley & Sons, New York.

Huber, P. J. (1996). *Robust Statistical Procedures*. siam, Philadelphia.

Jaeckel, L. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43**, 1449–1458.

Karioti, V. and Caroni, C. (2002). Detecting outlying time series in sets of short time series. *Computational Statistics & Data Analysis* **39**, 351–364.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* **68**, 53–78.

Koul, H. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with application to autoregression. *The Annals of Statistics* **22**, 540–562.

Koul, H. and Saleh, A. (1993). R-estimation for the parameters of autoregressive [AR(p)] models. *The Annals of Statistics* **21**, 534–551.

Kreiss, J.-P. (1997). Asymptotical properties of residual bootstrap for autoregressions. *Preprint*, Technische Universität Braunschweig, Braunschweig.

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

Künsch, H. R. (1995). Locally stationary AR-processes. Private communication.

Leonov, V. and Shiryayev, A. N. (1959). On a method of calculation of semi-invariants. *Theory of Probability and its Applications* **IV**(3), 319–329.

Lo, C.-L. and Li, W.-K. (1990). Two new approaches to robust estimation in time series. *J. Statist. Comput. Simul.* **34**, 209–255.

Lütkepohl, H. (1991). *Introduction to Multivariate Time Series*. Springer, Berlin.

Lütkepohl, H. (1996). *Handbook of Matrices*. Wiley & Sons, Chichester.

Maronna, R. A. (1974). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics* **4**(1), 51–67.

Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis* **21**(6), 663–684.

Maddala, G. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica* **39**(2), 341–358.

Maronna, R. A., Stahel, W. A. and Yohai, V. J. (1992). Bias-robust estimators of multivariate scatter based on projections. *Journal of Multivariate Analysis* **42**, 141–161.

Martin, R. and Yohai, V. (1985). Robustness in time series and estimating ARMA models. In D. B. E.J. Hannan and P. Krishnaiah (eds), *Handbook of Statistics*, Vol. 5, Elsevier, pp. 119–155.

Martin, R. and Yohai, V. (1991). Bias robust estimation of autoregression parameters. In W. Stahel and S. Weisberg (eds), *Directions in Robust Statistics and Diagnostics Part I*, Vol. 33, Springer, New York, pp. 233–246.

Mátyás, L. and Sevestre, P. (eds) (1992). *The Econometrics of Panel Data*. Kluwer Academic, Dordrecht.

Müller, A., Hartmann, M., Müller, K. and Eich, W. (2003). Validation of the arthritis self-efficacy shortform scale in german fibromyalgia patients. *Europ. J. Pain* **7**(2), 163–171.

Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross-sections. *Econometrica* **39**(6), 359–387.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Stat. Prob. Lett.* **1**(4), 327–333.

Rousseeuw, P. J. and Croux, C. (1993). An alternative to the median absolute deviation. *Journal of the American Statistical Association* **88**(424), 1273–1283.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley Series in Probability and Mathematical Statistics, Wiley & Sons, New York.

Rousseeuw, P. (1984). Least median of squares regression. *Journal of the American Statistical Association* **79**, 871–880.

Rousseeuw, P. and Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212–223.

Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle and R. Martin (eds), *Robust and Nonlinear Time Series Analysis*, number 26 in *Lecture Notes in Statistics*, Springer, New York, pp. 256–272.

Sen, P. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association* **63**, 1379–1389.

Serfling, R. J. (1984). Generalized L-, M- and R-statistics. *The Annals of Statistics* **12**(1), 76–86.

Sethuraman, S. and Basawa, I. (1994). Parameter estimation in a stationary autoregressive process with correlated multiple observations. *Journal of Statistical Planning and Inference* **39**, 137–164.

Shaman, P. and Stine, R. A. (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association* **83**(403), 842–848.

Shiryayev, A. N. (1984). *Probability*. Springer, NY.

Steinbeck, J. (2000). *Of Mice and Men*. Penguin Books, London.

Terpstra, J. T., McKean, J. W. and Naranjo, J. D. (2001). Weighted Wilcoxon estimates for autoregression. *Aust. N.Z.J. Stat* **43**(4), 399–419.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis (parts 1-3). *Ned. Akad. Wet. Proc. Ser. A* **53**, 386–392,521–525,1397–1412.

Tjøstheim, D. and Paulsen, J. (1983). Bias of some commonly-used time series estimates. *Biometrika* **70**(2), 389–399. Correction, Biometrika, 71, 656.

Tyler, D. (1987). A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics* **15**, 234–251.

Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* **1 and 2**(50), 129–134.

Wilson, K., Watson, S. and Currie, S. (1998). Daily diary and ambulatory activity monitoring of sleep in patients with insomnia associated with chronic musculoskeletal pain. *Pain* **75**(1), 75–84.

Wolfe, F., Smythe, H., Yunus, M., Bennett, R., Bombardier, C. and et al., D. G. (1990). The american college of rheumatology 1990 criteria for the classification of fibromyalgia. report of the multicenter criteria committee. *Arthritis. Rheum.* **33**, 160–172.

Yao, Q., Tong, H., Finkenstadt, B. and Stenseth, N. (2000). Common structure in panels of short ecological time-series. *J. R. Statist. Soc. B* **267**, 2459–2467.