Robust Estimation for GARCH Models and VARMA Models

Lancaster University

This dissertation is submitted for the degree of Doctor of Philosophy

Hang Liu

February 21, 2021

Abstract

This thesis contributes to the theory and methodology of robust estimation for two time series models: the generalized autoregressive conditional heteroscedasticity (GARCH) model and the vector autoregressive moving-average (VARMA) model.

More specifically, the first part (Chapter 3) of this thesis considers a class of M-estimators of the parameters of the GARCH models which are asymptotically normal under mild assumptions on the moments of the underlying error distribution. Since heavy-tailed error distributions without higher order moments are common in the GARCH modeling of many real financial data, it becomes worthwhile to use such estimators for the time series inference instead of the quasi maximum likelihood estimator. We discuss the weighted bootstrap approximations of the distributions of M-estimators. Through extensive simulations and data analysis, we demonstrate the robustness of the M-estimators under heavy-tailed error distributions and the accuracy of the bootstrap approximation. In addition to the GARCH(1, 1) model, we obtain extensive computation and simulation results which are useful in the context of higher order models such as GARCH(2, 1) and GARCH(1, 2) but have not yet received sufficient attention in the literature. We use M-estimators for the analysis of three real financial time series fitted with GARCH(1, 1) or GARCH(2, 1) models.

In the second part (Chapter 4) of this thesis, we propose a novel class of estimators of the GARCH parameters based on ranks, called R-estimators, with the property that they are asymptotic normal under the existence of a more than second moment of the errors and are highly efficient. We also consider the weighted

bootstrap approximation of the finite sample distributions of the R-estimators. We propose fast algorithms for computing the R-estimators and their bootstrap replicates. Both real data analysis and simulations show the superior performance of the proposed estimators under the normal and heavy-tailed distributions. Our extensive simulations also reveal excellent coverage rates of the weighted bootstrap approximations. In addition, we discuss empirical and simulation results of the R-estimators for the higher order GARCH models such as the GARCH$(2, 1)$ and asymmetric models such as the GJR model.

In the third part (Chapter 5 and Chapter 6) of this thesis, we propose a new class of R-estimators for semiparametric VARMA models in which the innovation density plays the role of the nuisance parameter. Our estimators are based on the novel concepts of multivariate center-outward ranks and signs. We show that these concepts, combined with Le Cam's asymptotic theory of statistical experiments, yield a class of semiparametric estimation procedures, which are efficient (at a given reference density), root-n consistent, and asymptotically normal under a broad class of (possibly non elliptical) actual innovation densities. No kernel density estimation is required to implement our procedures. A Monte Carlo comparative study of our R-estimators and other routinely-applied competitors demonstrates the benefits of the novel methodology, in large and small sample.

Dedication

This work is dedicated to my mum and dad with love and thanks

Declaration

This thesis has not been submitted in support of an application for another degree at this or any other university. It is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated. Many of the ideas in this thesis were the product of discussion with my supervisor Dr. Kanchan Mukherjee and Professor Marc Hallin and Professor Davide La Vecchia. The following publications are derived from this thesis:

(1) Hallin, M., La Vecchia, D. and Liu, H. (2020). Center-outward R-estimation for semiparametric VARMA models. *Journal of the American Statistical Association*, 1–14.

(2) Liu, H. and Mukherjee, K. (2019). R-estimators in GARCH models; asymptotics, applications and bootstrapping. *arXiv preprint arXiv:1912.07592.*

(3) Liu, H. and Mukherjee, K. (2020). M-estimation in GARCH models without higher order moments. *arXiv preprint arXiv:2001.10782.*

Hang Liu

Lancaster University, UK

Acknowledgements

First, I would like to thank my supervisor Dr Kanchan Mukherjee for his patient guidance and constant support since my MSc study, as well as giving me much liberty to my research. I would also like to express my great gratitude to Professor Marc Hallin and Davide La Vecchia for offering me the precious opportunity to work with them.

It is quite difficult for me, being an international student, to find a funding institution. Therefore, I gratefully acknowledges the *ESRC North West Social Science Doctoral Training Partnership* (NWSSDTP Grant Number ES/P000665/1) for their generous funding.

I thank my mother-in-law for helping look after my son. I really enjoy the time that I spent with my wife and my son, and I appreciate my wife for what she did for this family and my son for bringing so much happiness to us. Words cannot express my thanks to my parents for that they always respect my decisions and their constant financial and emotional support. I feel quite guilty for that I was not there while my parents really need my help during my PhD study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Why a robust estimator? The reasons are twofold. First, for a statistical model, an estimator of high quality (e.g. large efficiency) is more desirable. Second, by using a robust estimator, heavy-tails of the innovation or a certain degree of contamination of observations would not lead to a disaster (e.g. large deviation from the true parameter).

This thesis is devoted to the development of robust estimators for two time series models: the generalized autoregressive conditional heteroscedasticity (GARCH) model, which is univariate, and the vector autoregressive moving-average (VARMA) model, which is multivariate. For the GARCH model, we investigate a class of M-estimators proposed by Mukherjee (2008), and we propose a class of rank-based estimators (R-estimators hereafter). We also consider a type of bootstrap for both the M-estimators and R-estimators based on a sequence of exchangeable weights to approximate their finite sample distributions. For the VARMA model, we propose a class of *center-outward R-estimators*.

Although both are rank-based robust estimators, there are some methodological differences between the R-estimators for the GARCH model and for the VARMA model: (i) The former is based on classical univariate ranks and signs, while the latter, due to lack of canonical ordering in dimension $d \geq 2$, takes advantage of novel concepts of *center-outward ranks and signs* recently proposed by Hallin (2017), which hinge on measure-transportation theory. (ii) Strictly speak-

ing, the former is not genuinely rank-based as the objective function involves both the ranks and the observations; on the contrary, the objective function of the latter is measurable with respect to the σ-field generated by the ranks and signs; hence it is distribution-free. (iii) The techniques we use to establish the asymptotic normalities of these R-estimators are completely different: the former is based on some results for empirical processes; the later uses Le Cam's asymptotic theory of statistical experiments and a Hájek asymptotic representation result for center-outward rank statistics.

1.1 M- and R-estimation for the GARCH model

Financial time series, such as returns of stock indices or exchange rates, often exhibit some important "stylized facts" such as heavy-tailed distribution, volatility clustering, and so on; see Taylor (2005) for details. To capture these features, Engle (1982) proposed the autoregressive conditional heteroscedasticity (ARCH) model for the volatility process, in which the conditional variance is a linear function of the lagged squared returns. However, one disadvantage of the ARCH model is that it often needs many parameters to fit a financial data adequately. To deal with this, Bollerslev (1986) proposed the generalized ARCH (GARCH) model, in which the conditional variance is the linear function of both the lagged squared returns and the lagged squared volatilities. Since then, the GARCH model has enjoyed wide popularity among both researchers and financial practitioners, and it has become an important model in financial risk management. For example, see Francq and Zakoïan (2010) and Christoffersen (2012).

In terms of parameter estimation in the GARCH model, the most commonly-applied method is the quasi-maximum likelihood estimator (QMLE), which is obtained by maximizing the normal likelihood function. The asymptotic normality of the QMLE was established by Weiss (1986) for the ARCH model and by Lee and Hansen (1994) and Lumsdaine (1996) for the GARCH$(1,1)$ model. For the general GARCH(p,q) model, it was proved by Berkes et al. (2003). They assumed the

existence of the fourth moment of the innovation term in the ARCH or GARCH model. However, financial time series were frequently reported to have heavy-tailed distributions and this assumption may not hold. Hence, it is necessary to develop robust estimation procedures.

A popular type of robust estimation procedure is M-estimation, which was introduced by Huber (1964) for location models. M-estimator, with an appropriate defined score, is usually defined through minimizing a function associated with the score or as a solution of an equation; see van der Vaart (1998, Chapter 5). It is a generalization of some classical estimators such as the least squares estimator (LSE), least absolute deviations (LAD) estimator, MLE, and so on. The advantage of the M-estimator, compared to the QMLE, is that the consistency (usually $\sqrt{n}$-consistency) holds under milder moment assumption; see Huber (1964), Huber (1973) in the context of location and regression models and Zhou et al. (2018) in the context of high-dimensional regression model.

Another type of robust estimation procedure is R-estimation, which is not only robust against heavy-tails but also achieves the *semiparametric efficiency bound* when the chosen reference density coincides with the underlying distribution; see, e.g., Hallin and La Vecchia (2017, Section 3.2) for details. R-estimation has been proposed first in the context of location (Hodges and Lehmann 1956) and regression (Jurečková 1971, Jaeckel 1972) models with independent observations. R-estimation later on was extended to autoregressive time series models (Koul and Saleh 1993, Koul and Ossiander 1994, Terpstra et al. 2001, Mukherjee and Bai 2002, Andrews 2008, 2012). Extensions to the estimation of non-linear time series such as AR-GARCH, discretely observed diffusions with jumps, or autoregressive conditional duration models were considered by Mukherjee (2007), Andreou and Werker (2015), and Hallin and La Vecchia (2017, 2019).

This thesis considers both M-estimation (Chapter 3) and R-estimation (Chapter 4) for the GARCH model. The former one was proposed by Mukherjee (2008), for which the asymptotic normality only requires existence of a fractional moment

of the innovation term in the GARCH model. In this thesis, a new computation-efficient algorithm for the M-estimation is employed. In addition to some classic M-scores, such as the QMLE, Least Absolute Deviation (LAD) and Huber's scores, this thesis also considers some alternatives including μ-score and Cauchy score which were not investigated in the literature. Instead of considering just the GARCH$(1, 1)$ model, the M-estimation is applied to some higher order GARCH models, which has attracted attention in recent literature; see, for example, Francq and Zakoïan (2009). Extensive simulation shows the M-estimators are more robust than the QMLE under heavy-tailed distributions. Also, it is shown that the robustness holds when the underlying data generating process is the GARCH(1, 1) model but is misspecified to the GARCH(2, 1) model (this type of misspecification is essentially the case where the parameter is at the boundary of the parameter space).

The class of the R-estimators for the GARCH model proposed in this thesis is defined through a one-step procedure, which is based on an asymptotic linearity result of a *rank-based central sequence*. It is shown that under some mild conditions, the R-estimators, similar to the M-estimators, converge to normal distributions at rate $n^{1/2}$. Extensive simulation is carried out to compare the QMLE with different types of the R-estimators. It shows that the QMLE is outperformed by the R-estimators under heavy-tailed error distributions in terms of both bias and mean square error (MSE). Except for the robustness against heavy-tails, the R-estimators are also efficient. For example, simulation results show that for a finite sample size, the R-estimator based on the normal score function achieves almost the same efficiency as the QMLE under the normal error distribution, and it is more efficient than the QMLE for heavy-tailed distributions.

This thesis also provides empirical study on the M-estimators and R-estimators through analysing various financial time series from stock indices and exchange rates. It is shown that the M-estimates and R-estimates are quite different from the QMLE for some financial data. Analysis of the residuals shows that the difference

may be caused by infinite fourth moment of the innovation term, which leads to the failure of the QMLE.

This thesis also investigates bootstrap method for both the M-estimators and R-estimators. Instead of resampling residuals (see Christoffersen and Gonclaves 2005, Hall and Yao 2003) or blocks of the likelihood function (see Corradi and Iglesias 2008) in the GARCH model, which are two commonly-employed bootstrap techniques in the literature, this thesis considers resampling weights, which are independent of the data. The main advantages of the weighted bootstrap are its simplicity and computational efficiency: only the weights need to be generated for each bootstrap replicate; the computation algorithm for computing the bootstrap estimates is at the same time software-friendly as the algorithm for computing the robust estimates. Through extensive simulation, this thesis reveals good performance of the weighted bootstrap approximation for both the M-estimators and R-estimators under finite sample size settings.

1.2 Center-outward R-estimation for the VARMA model

The VARMA model is normally used for studying the correlation structure between multiple time series. It has applications in various fields including economics, biology and so on; see, e.g., Tsay (2014, Section 3.15), Fujita et al. (2007) for details. Similar to the GARCH model, the QMLE is routinely-used for estimating the VARMA parameter, and it is $\sqrt{n}$-consistent and asymptotically normal under finite fourth moment assumption. Therefore, despite its popularity, the QMLE for the VARMA model is sensitive to heavy-tails (in fact, it is also sensitive to asymmetry; see examples in Chapter 5), and its asymptotic and finite-sample performance can be quite poor under non-Gaussian ones.

In principle, the ultimate theoretical remedy to this is the semiparametric estimation method described in the monograph by Bickel et al. (1993), which

yields uniformly, locally and asymptotically, semiparametrically efficient estimators. However, semiparametric estimation procedures are not easily implemented, since they rely on kernel-based estimation of the actual innovation density (hence the choice of a kernel, the selection of a bandwidth) and the use of sample splitting techniques. All these niceties thus require relatively large samples and are hard to put into practice in a multivariate context: the higher the dimension, the more delicate multivariate kernel density estimation and the larger the required sample size.

A natural question is thus: "Can R-estimation palliate the drawbacks of the QMLE and the Bickel et al. techique in dimension $d \geq 2$ the way it does in dimension $d = 1$"? This question, however, immediately comes up against another one: "What are ranks and signs, hence what is R-estimation, in dimension $d \geq 2$"? Indeed, starting with dimension $d = 2$, the real space $\mathbb{R}^d$ is no longer canonically ordered. Several notions of multivariate ranks and signs have been proposed in the statistical literature. Among them, the componentwise ranks (Puri and Sen 1971), the spatial ranks (Oja 2010), the depth-based ranks (Liu 1992; Liu and Singh 1993), and the Mahalanobis ranks and signs (Hallin and Paindaveine (2002a)). Those ranks and signs all have their own merits but also some drawbacks, which make them unsuitable for our needs (essentially, they are not distribution-free, or not maximally so); we refer to the introduction of Hallin et al. (2020a) for details. The Mahalanobis ranks and signs have been successfully considered for testing purposes in the time series context (Hallin and Paindaveine 2002b, 2004). However, no results on estimation are available, and their distribution-freeness property is limited to elliptical densities—a very strong symmetry assumption which we are not willing to make in this thesis.

Based on measure-transportation results, a data-driven ordering yielding a concept of ranks and signs for multivariate observations has been proposed recently by Chernozhukov et al. (2017), Hallin (2017), and Hallin et al. (2020a). Those *center-outward ranks and signs* (see Chapter 5 for details) enjoy all the properties

that make traditional univariate ranks a successful tool of inference. In particular, they are distribution-free (see Hallin et al. (2020a) for details), thus preserve the validity of rank-based procedures irrespective of the possible misspecification of the innovation density. Moreover, they are invariant with respect to shift and global scale factors and equivariant under orthogonal transformations; see Hallin et al. (2020b). Essentially, the center-outward ranks and signs extend the validity of Mahalanobis ranks and signs-based methods to arbitrary absolutely continuous distributions in $\mathbb{R}^d$.

Based on the center-outward ranks and signs, this thesis (Chapter 5) proposes a new class of R-estimators for semiparametric VARMA models, with the innovation density playing the role of nuisance parameter. Combined with Le Cam's theory (see, e.g., Le Cam and Yang 2000, Chapter 6 and van der Vaart 1998, Chapter 7) and the local asymptotic normality results in Garel and Hallin (1995) and Hallin and Paindaveine (2004), this thesis derives the relevant asymptotic theory of the R-estimators, which are root-n consistent and asymptotically normal under a broad class of innovation densities including, e.g., multi-modal mixtures of Gaussians. For the sake of applicability, an algorithm which explains how to implement our estimators is given. Extensive simulation study shows significant superiority of R-estimators over the QMLE for non-elliptical innovations.

Contributions. The major contributions of this thesis are as follows. First, for the GARCH model, extensive empirical analysis of various types of M-estimators and the weighted bootstrap approximation are carried out; a novel class of R-estimators are proposed, and the asymptotic results are derived. Second, for the VARMA model, a class of R-estimators based on the center-outward ranks and signs are proposed, and the asymptotic theory is developed. To the best of our knowledge, this thesis is the first successful attempt to apply the center-outward R-estimation to multivariate time series. It extends the application of multivariate R-estimation to a wider horizon—from elliptic distributions to non-elliptic distributions. Third, for easy of implementation, this thesis provides algorithms

for computing the M-estimators and R-estimators for the GARCH model and R-estimators for the VARMA model (the codes for the latter are available from the authors' `GitHub` page `https://github.com/HangLiu10/RestVARMA`).

1.3 Structure of the thesis

This thesis is organised as follows: Chapter 2 contains some preliminaries. The GARCH model and some relevant topics are discussed. Different types of the bootstrap methods and robust estimators are described. The VARMA model is introduced.

Chapter 3 explores the M-estimators and the weighted bootstrap in the GARCH model. Real data analysis and simulation study under different error distributions are carried out to compare the QMLE with the M-estimators. Different bootstrap schemes are compared through simulation. Also, the M-estimators are computed and studied for higher order GARCH models in addition to the traditional GARCH$(1, 1)$ model.

Chapter 4 proposes a class of the R-estimators and their weighted bootstrap in the GARCH model. Asymptotic normality are proved based on some results on empirical processes. Different types of the R-estimators are compared with the QMLE through both real data analysis and simulation. Bootstrap schemes are compared under different sample sizes. Applications to an asymmetric GARCH model is included.

Chapter 5 proposes a class of the center-outward R-estimators for semiparametric VARMA models. The asymptotic normality result is derived by using Le Cam's theory and the local asymptotic normality. Simulation is carried out under both elliptic and non-elliptic distributions to compare the R-estimators with the QMLE. A real data example is included.

Chapter 6, as the supplementary material for Chapter 5, collects all proofs, computational aspects, and further numerical results of Chapter 5.

Chapter 7 concludes the thesis and discusses the topics that of future interest.

Chapter 2

Preliminaries and literature review

2.1 GARCH models

Financial time series, such as returns of stock indices or exchange rates, often exhibit three important"stylized facts"; see empirical examples in Taylor (2005), Christoffersen (2012), Bauwens et al. (2012), and so on. First, the distribution of returns is nonnormal and has heavy-tails. Second, there is almost no correlation between returns. Moreover, there is positive correlation between absolute returns, and likewise for squared returns. The third fact is reflected by volatility clustering, that is, high volatilities are often followed high volatilities and vice versa. Considering the fact of volatility clustering, it is inappropriate to assume constant variance for financial time series. Therefore, the classical Autoregressive Moving Average (ARMA) model that assumes constant variance of error term is not suitable. Instead, one should adopt a model with variance autocorrelated and changing throughout time.

Engle (1982) proposed the autoregressive conditional heteroscedasticity (ARCH) model for the volatility process, which is consistent with the stylized facts. Specifically, let $\{X_t; t \in \mathbb{Z}\}$ denote a sequence of a financial time series that forms

the observations, and let $\mathcal{F}_{t-1}$ stand for the σ-field generated by $X_{t-1}, X_{t-2}, \cdots$. Denote by $v_t = \sigma_t^2$ the conditional variance at time t given $\mathcal{F}_{t-1}$, where $\sigma_t := (\mathrm{E}(X_t^2|\mathcal{F}_{t-1}))^{1/2}$ is usually called the conditional volatility. Then the ARCH(p) model is defined by two equations

$$X_t = \sigma_t \epsilon_t, \tag{2.1.1}$$

and

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2, \quad p \in \mathbb{Z}^+, \tag{2.1.2}$$

where $\omega_0 > 0$,$\alpha_{0i} \geq 0, \forall i$ are assumed to ensure that $\{\sigma_t^2; t \in \mathbb{Z}\}$ are strictly positive (throughout, we use suffix 0 to denote the true model parameters), and $\{\epsilon_t; t \in \mathbb{Z}\}$ are unobservable independent and identical distributed (i.i.d.) errors with mean zero.

It has been shown in the literature that for a financial time series, one often needs to use a high order ARCH model to fit the data adequately. We can check the adequacy of the fitted ARCH model by testing the independence of the residuals using, e.g. ,the Ljung–Box statistics. See Tsay (2010, Section 3.4.3) for further details. For example, for the monthly excess returns of the S&P 500 index analyzed in Tsay (2010), an ARCH(9) model is needed. To circumvent this, Bollerslev (1986) proposed the generalized ARCH (GARCH) model, in which the conditional variance is the linear function of both the lagged squared returns and the lagged squared volatilities. The GARCH(p, q) model is defined by (2.1.1) and

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2, \ t \in \mathbb{Z}, \tag{2.1.3}$$

with $\omega_0 > 0$, $\alpha_{0i}, \beta_{0j} \geq 0$, $\forall\, i, j$. Throughout, we will let

$$\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, ..., \alpha_{0p}, \beta_{01}, ..., \beta_{0q})'$$

denote the true parameter and $\boldsymbol{\Theta}$ denote the parameter space.

The recursive property of the GARCH model allows it to be represented as an ARCH(∞) model. Thus, as shown in the literature, a low order GARCH model, such as GARCH$(1,1)$, is usually able to fit financial time series adequately.

For a GARCH(p,q) model, when the innovation term ϵ_t is assumed to have unit variance, which is a standard assumption adopted by many researcher, (2.1.1) and (2.1.3) imply that the unconditional variance $\mathrm{E}(\sigma_t^2)$ satisfies

$$\mathrm{E}(\sigma_t^2) = \omega_0/(1 - \sum_{i=1}^{p} \alpha_{0i} - \sum_{j=1}^{q} \beta_{0j}).$$

Therefore, Bollerslev (1986) showed that the condition

$$\sum_{i=1}^{p} \alpha_{0i} + \sum_{j=1}^{q} \beta_{0j} < 1 \tag{2.1.4}$$

needs to be met for the GARCH model to be second order stationary. However, Bougerol and Picard (1992a; 1992b) argued that many empirical financial time series do not satisfy (2.1.4) while they are still strictly stationary (an example that (2.1.4) does not hold is the integrated GARCH (IGARCH) proposed by Engle and Bollerslev (1986), which assumes that the left hand side of (2.1.4) equals 1). Hence, Bougerol and Picard (1992a; 1992b) proposed a necessary and sufficient condition for a GARCH(p,q) model to be stationary. To state the condition, we introduce the following notations. Let

$$\boldsymbol{\tau}_n = (\beta_1 + \alpha_1\epsilon_n^2, \beta_2, ..., \beta_{q-1}) \in \mathbb{R}^{q-1},$$

$$\boldsymbol{\xi}_n = (\epsilon_n^2, 0, ..., 0) \in \mathbb{R}^{q-1}$$

and

$$\boldsymbol{\alpha} = (\alpha_2, ..., \alpha_{p-1}) \in \mathbb{R}^{p-2},$$

where $p, q \geq 2$ can always be achieved by letting some α_i, β_j equal to 0. Define a

$(p+q-1)\times(p+q-1)$ matrix $\boldsymbol{A}_n$, written in block form, by

$$\boldsymbol{A}_n = \begin{bmatrix} \boldsymbol{\tau}_n & \beta_q & \boldsymbol{\alpha} & \alpha_p \\ \boldsymbol{I}_{q-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\xi}_n & 0 & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{p-1} & \mathbf{0} \end{bmatrix},$$

where $\boldsymbol{I}_{q-1}, \boldsymbol{I}_{p-1}$ are the identity matrices of size $q-1$ and $p-1$, respectively. Also, define the matrix norm of any $d \times d$ matrix $\boldsymbol{M}$ by

$$||\boldsymbol{M}|| = \sup\{||\boldsymbol{M}\boldsymbol{x}||/||\boldsymbol{x}||; \boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{x} \neq \mathbf{0}\}.$$

The top Lyapunov exponent associate to a sequence $\{\boldsymbol{A}_n; n \in \mathbb{Z}\}$ is defined by

$$\gamma = \inf_{0 \leq n < \infty} \left\{ E\left(\frac{1}{n+1} \log ||\boldsymbol{A}_0 \boldsymbol{A}_1 ... \boldsymbol{A}_n|| \right) \right\}.$$

Then a GARCH(p,q) model has a unique stationary solution if and only if

$$\gamma < 0. \tag{2.1.5}$$

For a GARCH$(1,1)$ model, this condition reduces to $E[\log(\beta_1 + \alpha_1 \epsilon_1^2)] < 0$, which is the same as the result obtained in Nelson (1990). Under the stationary condition (2.1.5), the asymptotic normality of the quasi-maximum likelihood estimator (QMLE) was proved by Berkes et al. (2003) under some mild conditions. It is also under this condition, the robust estimators in this thesis are proposed. Throughout, this thesis assumes (2.1.5) holds so that the proposed robust estimators for the GARCH model can achieve some nice properties, such as $\sqrt{n}$-consistence and the asymptotic normality.

Since it has been proposed, the GARCH model has enjoyed wide popularity among both practitioners and researcher, and there is a huge literature exploring its applications, among which an important one is financial risk management. In

particular, it can be used to estimate value at risk (VaR). As a popular measurement of risk, the VaR was proposed by the investment bank J.P. Morgan. Given the information up to $t-1$ and a constant p $(0 < p < 1)$, the VaR at time t is defined as the conditional p-th quantile of the return at time t. The estimation of the VaR using the GARCH model was discussed by Angelidis et al. (2004), Christoffersen and Goncalves (2005), Mancini and Trojani (2006), So and Yu (2006), Orhan and Koksal (2012), among others. Iqbal and Mukherjee (2010) and Liu (2016) employed the likelihood ratio tests proposed by Kupiec (1995) and Christoffersen (1998), which evaluate the VaR estimate according to the coverage rate and independence, and concluded that the GARCH model based estimation is reasonable. Another application of the GARCH model is the option pricing in finance. For details, see Engle and Mustafa (1992), Duan (1995), Heston and Nandi (2000), Duan and Simonato (2001), among others.

Except for the GARCH model, there are also other autoregressive conditional heteroscedastic models proposed for various purposes. As reported frequently in the literature, apart from the stylized facts mentioned above, financial time series also exhibits other properties. Therefore, some researcher argued that models that are more complicated than GARCH are needed to accommodate these properties. Examples of some popular models are as follows.

It was shown in the literature that many financial datasets exhibit asymmetry property, that is, negative and positive news tend to have different impacts on the volatility. In particular, negative news usually have greater influence than positive news. The volatilities in the time period pertaining to negative values of the time series are usually larger in magnitude than those corresponding to the time period of positive values. This is often refereed to as leverage effect of financial time series. To allow for this asymmetry effect, Glosten, Jagannathan and Runkle (1993) proposed a threshold GARCH (TGARCH) model (also called the GJR model in the literature). The TGARCH(p, q) model is defined by (2.1.1)

and

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \left[\alpha_{0i} + \gamma_{0i} I(X_{t-i} < 0)\right] X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2, \tag{2.1.6}$$

where $\omega_0 > 0, \alpha_{0i}, \gamma_{0i}, \beta_{0j} \geq 0, \forall i, j$. The greater impact of the negative returns is reflected by γ_{0i}. When $\gamma_{0i} = 0$ for all $i = 1, ..., p$, the model reduces to the GARCH(p, q) model. Note that Zakoïan (1994) also proposed a TGARCH(p, q) model, which, instead of using the quadratic form in (2.1.6), has the form

$$\sigma_t = \omega_0 + \sum_{i=1}^{p} \left[\alpha_{0i} X_{t-i}^{+} - \gamma_{0i} X_{t-i}^{-}\right] + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}$$

with $X_{t-i}^{+} = \max\{X_{t-i}, 0\}$ and $X_{t-i}^{-} = \min\{X_{t-i}, 0\}$. Another GARCH-type model dealing with the asymmetry effect is the exponential GARCH (EGARCH) model proposed by Nelson (1991). The EGARCH(p, q) model is defined by (2.1.1) and an ARMA parameterization:

$$\log(\sigma_t^2) = \omega_0 + \frac{1 + \sum_{j=1}^{q} \beta_{0j} L^j}{1 - \sum_{i=1}^{p} \alpha_{0i} L^i} g(\epsilon_{t-1}),$$

where L is the lag operator, $g(\epsilon_t)$ is a linear combination of both ϵ_t and $|\epsilon_t|$ with two parameters θ and γ that takes the form

$$g(\epsilon_t) = \theta \epsilon_t + \gamma[|\epsilon_t| - \mathrm{E}(|\epsilon_t|)].$$

Apart from the asymmetric impact, long-memory dependence between the squared returns was also reported for some financial data. Therefore, Baillie (1996) suggested a fractional integrated GARCH (FIGARCH) model that uses a fractional differencing operator $(1 - L)^d$, with $0 \leq d \leq 1$, to allow for a slow hyperbolic rate of decay of the lagged squared returns in the conditional variance function.

For more details regarding models in the GARCH-family, see monographs by Taylor (2005), Francq and Zakoïan (2010), Tsay (2010), Bauwens et al. (2012), among others.

2.2 Bootstrap methods

Due to its simplicity, the bootstrap proposed by Efron (1979) has enjoyed wide popularity among statisticians. It provides a convenient tool to study the sample distribution of some pre-specified random variable based on the given data. In particular, the motivation behind his methodology is, as described in Efron (1982), for a random variable $R(\boldsymbol{X}, F)$ of interest, where $\boldsymbol{X} = \{X_1, X_2, ..., X_n\}$ is a sequence of i.i.d. random variables from an unknown probability distribution F, we wish to estimate some aspect of R's distribution, for instance, its expectation $\mathrm{E}_F R$ or the probability $\mathrm{P}(R < 2)$.

The bootstrap method provides an extremely simple solution to this problem:

Step 1. Construct the empirical distribution function $\hat{F}$ by assigning probability mass $1/n$ to each observed values of $X_i, 1 \leq i \leq n$.

Step 2. Draw a"bootstrap sample" from $\hat{F}$, i.e.,

$$X_1^*, X_2^*, ..., X_n^* \overset{i.i.d.}{\sim} \hat{F}.$$

Step 3. Approximate the distribution of $R(\boldsymbol{X}, F)$, which is usually unknown, by the bootstrap distribution of

$$R^* = R(\boldsymbol{X}^*, \hat{F}).$$

As a concrete example, the bootstrap is always used to estimate bias of a statistic $\theta(F)$, i.e., $\text{Bias} = \mathrm{E}_F\theta(\hat{F}) - \theta(F)$. Then one can take $R(\boldsymbol{X}, F) = \theta(\hat{F}) - \theta(F)$ and get

$$R^* = \theta(\hat{F}^*) - \theta(\hat{F}).$$

Accordingly, the bootstrap estimate of bias is $\widehat{\text{Bias}} = \mathrm{E}_* R^*$. By drawing the bootstrap sample B times, where B is usually a large number, $\widehat{\text{Bias}}$ can be approximated by

$$1/B \sum_{b=1}^{B} \hat{\theta}^{*b} - \hat{\theta}.$$

When one is interested in estimating variance of $\theta(F)$, i.e., Var $= \mathrm{E}_F[\theta(\hat{F}) - \theta(F)]^2$. Then the bootstrap estimate of variance is the approximation by

$$1/(B-1)\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \hat{\theta}\right)^2.$$

The bootstrap method described above that uses the empirical distribution function is nonparametric. Another case to be distinguished, as described in the monograph by Davison and Hinkley (1997), is when a particular parametric model is given for the distribution of $\boldsymbol{X} = \{X_1, X_2, ..., X_n\}$. In this situation, one can first estimate the parameter ψ in the model by $\hat{\psi}$, which is often the maximum likelihood estimate since the distribution is known. Then by using its substitution in the model to give the fitted model, with the distribution function $\hat{F} = F_{\hat{\psi}}$, one can draw the bootstrap sample from the fitted model.

When it comes to dependent data, which is quite common for financial time series, one cannot simply use the bootstrap methods mentioned above as they will break the dependence structure between the random variables. Therefore, some alternative bootstrap methods need to be considered. A good reference on this is the monograph by Lahiri (2010), which describes various aspects of the bootstrap for dependent data, including methodology, second-order properties, and so on. Some methods in the book are as follows.

Suppose a sequence of random variables $\{X_n; n \geq 1\}$ satisfying

$$X_n = h(X_{n-1}, ..., X_{n-p}; \boldsymbol{\beta}) + \epsilon_n, \tag{2.2.1}$$

where $\{\epsilon_n; n \geq 1\}$ is a sequence of i.i.d. random variables from a probability distribution F with mean zero. Let $\hat{\boldsymbol{\beta}}_n$ denote an estimator, e.g., the maximum likelihood estimator. Define the raw residuals

$$\hat{\epsilon}_i = X_i - h(X_{i-1}, ..., X_{i-p}; \hat{\boldsymbol{\beta}}_n), \quad p < i \leq n.$$

By centring the raw residuals $\hat{\epsilon}_i$'s, one obtains the centred residuals

$$\tilde{\epsilon}_i = \hat{\epsilon}_i - \bar{\epsilon}_n, \quad p+1 \leq i \leq n,$$

where

$$\bar{\epsilon}_n = 1/(n-p) \sum_{i=1}^{n-p} \hat{\epsilon}_{i+p}.$$

Then one can draw bootstrap sample $\{\epsilon_i^*; p+1 \leq i \leq n\}$ from $\{\tilde{\epsilon}_i; p+1 \leq i \leq n\}$ with replacement. Thus using the model (2.2.1), the bootstrap observations can be constructed by letting

$$X_i^* = X_i \quad \text{for} \quad 1 \leq i \leq p$$

and

$$X_i^* = h(X_{i-1}^*, ..., X_{i-p}^*; \hat{\boldsymbol{\beta}}_n) + \epsilon_i^* \quad \text{for} \quad p+1 \leq i \leq n.$$

This type of the bootstrap is often referred to as the residual bootstrap in the literature. As pointed out by Lahiri (2010), its performance depends heavily on the structure of the sequence $\{X_n; n \geq 1\}$ and the prior specification of the model. In particular, it works well when the sequence $\{X_n; n \geq 1\}$ is stationary, whereas it is very sensitive to the parameter value when $\{X_n; n \geq 1\}$ is nonstationary. Therefore, alternative methods were developed to overcome this disadvantage.

Kunsch (1989) proposed a moving block bootstrap (MBB) method, which, unlike the residual bootstrap, does not require us to fit a parametric or semiparametric model but still works for short-range dependence data. The idea is that by resampling blocks of observations instead of a single observation at a time, one can preserve the dependence structure of the data. In particular, let $\mathcal{B}_i = \{X_i, X_{i+1}, ..., X_{i+l-1}; 1 \leq i \leq N\}$ denote a block of length l, where $N = n - l + 1$. Then by resampling with replacement from $\{\mathcal{B}_1, ..., \mathcal{B}_N\}$, one can obtain a sequence of bootstrap blocks $\{\mathcal{B}_1^*, ..., \mathcal{B}_k^*; k \geq 1\}$. Let $\{X_{(i-1)l+1}^*, X_{(i-1)l+2}^*, ..., X_{il}^*; 1 \leq i \leq k\}$ denote the elements in $\mathcal{B}_i^*$. Then the MBB sample $\{X_1^*, X_2^*, ..., X_m^*\}$ is of size

$m = kl$. Thus, repeating the step 3 of the Efron's bootstrap, one can obtain the bootstrap distribution of R^* as an approximation to the distribution of $R(\boldsymbol{X}, F)$. Under appropriate conditions, the consistency is achieved if $l = l(n) \to \infty$ and $l/n \to 0$. However, as pointed out by Lahiri (2010), the MBB resampling scheme suffers from an undesirable boundary effect since it assigns less weights to the observations toward the beginning and the end than to the middle part. To overcome this disadvantage, he described a generalized block bootstrap by introducing a new sequence of random variables, which is essentially the periodic extension of the original dataset $\boldsymbol{X}$.

Apart from the bootstrap methods mentioned above, there are also many others in the literature. A popular one is the weighted bootstrap (WBS) (or generalized bootstrap) of Chatterjee and Bose (2005). Suppose $\{\boldsymbol{\phi}_{ni}(X_{ni}, \boldsymbol{\beta}); 1 \leq i \leq n, n \geq 1\}$ is a triangular sequence of functions taking values in $\mathbb{R}^p$. Assume that $\mathrm{E}\boldsymbol{\phi}_{ni}(X_{ni}, \boldsymbol{\beta}) = \mathbf{0}$ for some unique parameter $\boldsymbol{\beta}_0$ and an estimator $\hat{\boldsymbol{\beta}}_n$ is defined as the solution of the equation

$$\sum_{i=1}^{n} \boldsymbol{\phi}(X_i, \boldsymbol{\beta}) = \mathbf{0}.$$

An example of this type of estimator is M-estimator. The WBS estimator $\hat{\boldsymbol{\beta}}_B^*$ of Chatterjee and Bose (2005) is defined as the solution of

$$\sum_{i=1}^{n} w_{ni} \boldsymbol{\phi}(X_i, \boldsymbol{\beta}) = \mathbf{0},$$

where $\{w_{ni}; 1 \leq i \leq n, n \geq 1\}$ is a triangular sequence of random variables that are exchangeable and independent of $\{X_{ni}\}$. It turns out that a host of bootstrap schemes such as the Efron's bootstrap, subsampling, Bayesian bootstrap are special cases of the WBS. Chatterjee and Bose (2005) assumed that the weights satisfy the basic conditions

$$\mathrm{E}(w_{n1}) = 1, 0 < \mathrm{Var}(w_{n1}) = o(n), \mathrm{Corr}(w_{n1}, w_{n2}) = O(1/n),$$

and they proved the consistency of the WBS under these conditions.

When it comes to the bootstrap for the GARCH(p, q) model, the residual bootstrap is widely-employed since the residuals $\{\epsilon_t; t \in \mathbb{Z}\}$ are usually assumed to be i.i.d.. Some examples include Christoffersen and Gonclaves (2005) and Mancini and Trojani (2006). Jeong (2017) established the second-order asymptotic refinement of the residual bootstrap for the GARCH$(1, 1)$ model. For the GARCH(p, q) model with sample size n, the residual bootstrap is illustrated as follows.

Step 1. Estimate the parameter and fit the model to obtain the estimates of the residuals $\{\hat{\epsilon}_t; 1 \leq t \leq n\}$, i.e.,

$$\hat{\epsilon}_t = X_t/\hat{\sigma}_t$$

with

$$\hat{\sigma}_t^2 = \hat{\omega}_0 + \sum_{i=1}^{p} \hat{\alpha}_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \hat{\beta}_{0j} \hat{\sigma}_{t-j}^2.$$

Step 2. Draw bootstrap residuals $\{\epsilon_t^*; 1 \leq t \leq n\}$ from $\{\hat{\epsilon}_t; 1 \leq t \leq n\}$ according to the step 1 and 2 of the Efron's bootstrap. Note that in the literature, people also use the centred bootstrap residuals $\{\tilde{\epsilon}_t^*; 1 \leq t \leq n\}$, with

$$\tilde{\epsilon}_t^* = \epsilon_t^* - 1/n \sum_{i=1}^{n} \epsilon_t^*,$$

to ensure that the zero mean assumption is met. See Christoffersen and Gonclaves (2005) for details.

Step 3. Construct the sequence $\{X_t^*; 1 \leq t \leq n\}$ using the bootstrap residuals $\{\epsilon_t^*; 1 \leq t \leq n\}$.

Step 4. Obtain the bootstrap estimate using $\{X_t^*; 1 \leq t \leq n\}$.

As pointed out by Hall and Yao (2003) and Linton et al. (2010), when the fourth moment of the innovation is not finite, the above residual bootstrap fails for the QMLE as its asymptotic distribution is nonnormal. Hence, they suggested a percentile-t subsampling bootstrap, which works for both normal and nonnormal

cases. In particular, the bootstrap is carried out by following steps:

Step 1. Conduct the step 1 of the residual bootstrap for the GARCH model to obtain $\{\hat{\epsilon}_t; 1 \leq t \leq n\}$.

Step 2. Standardize $\{\hat{\epsilon}_t; 1 \leq t \leq n\}$ as follows so that the residual has mean zero and unit variance, i.e.,

$$\tilde{\epsilon}_t = \frac{\hat{\epsilon}_t - n^{-1}\sum_{t=1}^{n}\hat{\epsilon}_t}{\sqrt{n^{-1}\sum_{t=1}^{n}\hat{\epsilon}_t^2 - (n^{-1}\sum_{t=1}^{n}\hat{\epsilon}_t)^2}}.$$

Step 3. Resampling $\{\epsilon_t^*; 1 \leq t \leq n\}$ from $\{\tilde{\epsilon}_t; 1 \leq t \leq n\}$ with replacement.

Step 4. Construct $\{X_t^*; 1 \leq t \leq n\}$ using the bootstrap residuals $\{\epsilon_t^*; 1 \leq t \leq n\}$.

Step 4. Let $m < n$, compute the bootstrap estimate $\hat{\boldsymbol{\theta}}_m^*$ using the subsample $\{X_1^*, ..., X_m^*\}$.

Assuming $m = m(n) \to \infty$ and $m/n \to 0$, Hall and Yao (2003) showed that the percentile-t subsampling bootstrap is consistent.

Apart from the residual bootstrap, the block bootstrap is also employed in the literature for the GARCH model due to that $\{X_t; 1 \leq t \leq n\}$ are dependent. As described in Corradi and Iglesias (2008), it is based on resampling blocks of likelihood function. In particular, let

$$l_t(\boldsymbol{\theta}) = -\frac{1}{2}\log v_t(\boldsymbol{\theta}) - \frac{1}{2}\frac{X_t^2}{v_t(\boldsymbol{\theta})}$$

denote the log-likelihood function when the innovation is assumed with standard normal distribution. The block bootstrap for the GARCH model is by conducting the MBB procedure of Kunsch (1989) for $\{l_t(\boldsymbol{\theta}); 1 \leq t \leq n\}$, one obtains the MBB sample of the log-likelihood function $\{l_t^*(\boldsymbol{\theta}); 1 \leq t \leq n\}$. Then the block bootstrap estimator $\hat{\boldsymbol{\theta}}_n^*$ is

$$\hat{\boldsymbol{\theta}}_n^* = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\ \frac{1}{n}\sum_{t=1}^{n} l_t^*(\boldsymbol{\theta}).$$

A disadvantage of the block bootstrap for the GARCH model, as shown in Corradi and Iglesias (2008), is that for a good approximation, high moment of the

innovation must be finite. However, this does not hold for many financial data.

Due to its simplicity and computational friendly features, the WBS of Chatterjee and Bose (2005) is also employed for the GARCH model in the literature. Bose and Mukherjee (2009) considered the WBS version of a weighted linear estimator for the ARCH model and proved the bootstrap distribution is consistent. Iqbal and Mukherjee (2010) and Liu (2016) considered the WBS for the M-estimator of the GARCH$(1, 1)$ model and showed it provides good approximation through simulation. Mukherjee (2020) extended their work to the GARCH(p, q) model and proved the consistence of the WBS.

2.3 Robust estimation

In a parameter estimation procedure, statistical assumptions are sometimes made for mathematical convenience. For example, in a location or linear regression model, the error term is usually assumed with normal distribution. Thus the commonly-applied least squares estimator (LSE) is essentially equivalent to the maximum likelihood estimator (MLE) and it is therefore asymptotic optimal (in terms of efficiency). However, in real world, these assumptions may not hold and small deviation could end up with loss of efficiency or even failure of the statistic. To motivate the usefulness of the robust estimators, we mention below an example from Mukherjee and Wang (2014, Table 1.1), where the data was heavily contaminated. Using the LSE, which is obtained by minimizing the sum of the square residuals, they obtained an estimate of the coefficient that was greatly influenced by the contamination. However, when they adopted other estimation procedures, the contamination tended to have less influence. Due to their insensitivity, these estimators are often called robust estimators in the literature.

In Huber and Ronchetti (2009), robustness was defined as being insensitive to small deviations from the assumptions. They pointed out that a robust procedure should achieve three goals:

(i). Efficiency: It should have a reasonably good (optimal or nearly optimal)

efficiency at the assumed model.

(ii). Stability: It should have good performance when there are small deviations from the assumptions.

(iii). Breakdown: Somewhat larger deviations from the assumptions should not cause a catastrophe.

Aiming to achieve these goals, various types of robust procedures have been proposed in the literature, among which includes three important types of robust estimators: M-estimator, R-estimator and L-estimator. These estimators were first proposed for the location model; see Hettmansperger and McKean (2010) for definition of the location model.

The M-estimator of a location parameter was proposed by Huber (1964), where M stands for maximum likelihood type. It is a generalization of some classical estimators such as the LSE, least absolute deviations (LAD) estimator, MLE, and so on. In particular, suppose a sequence $\{X_i; 1 \leq i \leq n\}$ satisfies the location model

$$X_i = \theta + \epsilon_i, \tag{2.3.1}$$

where $\{\epsilon_i; 1 \leq i \leq n\}$ is an i.i.d. sequence with mean zero and θ is the location parameter of interest. Let ρ denote a function and ψ denote its derivative. Then the M-estimator of the location parameter is defined as

$$\hat{\theta}_n = \arg\min \sum_{i=1}^{n} \rho(X_i - \epsilon_i)$$

or the solution of the equation

$$\sum_{i=1}^{n} \psi(X_i - \epsilon_i) = 0.$$

Some examples of the M-estimator of the location parameter are as follows.

(i). The LSE: Letting $\rho(x) = x^2$, this yields the sample mean as the estimator. The estimator is sensitive to outliers.

(ii). The LAD: Here $\rho(x) = |x|$. This gives the sample median as the estimator

and is less sensitive to outliers than the LSE.

(iii). The Huber's estimator: Here $\rho(x) = \frac{1}{2}x^2 I(|x| \leq k) + \{k|x| - \frac{1}{2}k^2\} I(|x| > k)$, where $k > 0$ is a constant and $I(\cdot)$ is the indicator function. Similar to the LAD estimator, the Huber's estimator is less sensitive to outliers than the LSE.

As mentioned in Jurečková and Sen (1996), the M-estimator is generally locally robust. That is, it only allows small deviations from the model assumption. In a global sense, when the underlying distribution is unknown and belongs to a class of distributions, e.g. symmetric distribution, a rank based procedure or R-estimator is robust. As summarized in Hallin (2017), the reasons for using R-estimators are twofold:

(DF) (distribution-freeness): In an i.i.d. noise models (with θ denoting the parameter of the model), the vector of (θ-residuals) ranks is distribution-free over the (nonparametric) distribution family of non-vanishing densities f over $\mathbb{R}$. Specifically, the vector $\mathbf{R}^{(n)} := \{R_1^{(n)}, ..., R_n^{(n)}\}$ of ranks is uniform over the $n!$ permutations of $\{1, ..., n\}$, hence distribution-free.

(HW) (semiparametric efficiency perservation): The semiparametric efficiency bound can be reached via rank-based procedures in i.i.d. noise models; see Hallin and Werker (2003) for more details.

The R-estimator of a location parameter was proposed by Hodges and Lehmann (1963) for both one-sample and two-sample location models. It was motivated by rank test statistic such as the Wilcoxon or normal scores statistic. In particular, for the two-sample location model, the general rank test statistic, as described in Huber and Ronchetti (2009) and Hettmansperger and McKean (2010), is as follows.

Consider two i.i.d. samples $\{X_1, ..., X_m\}$ and $\{Y_1, ..., Y_n\}$ with distributions $F(x)$ and $F(x - \Delta)$ respectively. Let R_i denote the rank of X_i in the combined sample of size $m + n$. Let $\varphi : (0, 1) \to \mathbb{R}$ be a square integrable, nondecreasing function. Then the rank test for the null hypothesis $H_0 : \Delta = 0$ against the

alternative $H_A : \Delta > 0$ is defined as

$$S_{m,n} = \frac{1}{m} \sum_{i=1}^{m} a(R_i), \tag{2.3.2}$$

where the score function $a(i)$ is generated by φ as follows

$$a(i) = \varphi\left(\frac{i}{m+n+1}\right).$$

Under assumption that

$$\sum_{i=1}^{m+n} a(i) = 0$$

or

$$\int_0^1 \varphi(t)dt = 0,$$

the expected value of (2.3.2) is zero under H_0. Thus, the estimator of shift $\hat{\Delta}$ is obtained by solving the equation

$$S_{m,n} = 0. \tag{2.3.3}$$

Some popular rank scores are as follows:

(i). Sign rank score: Here $\varphi(u) = \text{sign}(u - \frac{1}{2})$.

(ii). Wilcoxon rank score: Here $\varphi(u) = u - \frac{1}{2}$. This yields the Wilcoxon rank test for the two-sample location model.

(iii). van der Waerden (vdW) rank score. Here $\varphi(u) = \Phi^{-1}(u)$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution.

Note that there are two cases to be distinguished, i.e., $mn = 2k$ and $mn = 2k+1$ for some positive integer k. Specifically, Let $\{W^{(1)}, ..., W^{(mn)}\}$ denote the ordered differences $\{Y_j - X_i; 1 \leq i \leq m, 1 \leq j \leq n\}$. When $mn = 2k + 1$, (2.3.3) has unique solution

$$\hat{\Delta} = W^{(k+1)}.$$

However, when $mn = 2k$, (2.3.3) does not have unique solution. Hence, Hodges

and Lehmann (1963) defined the estimate of shift as

$$\hat{\Delta} = \frac{1}{2}(\Delta^* + \Delta^{**}),$$

where Δ^* and Δ^{**} are obtained by letting

$$\Delta^* = \sup\{\Delta : S_{m,n} > 0\} \quad \text{and} \quad \Delta^{**} = \inf\{\Delta : S_{m,n} < 0\}.$$

In both cases, these three rank scores yields

$$\hat{\Delta} = \text{med}\{Y_j - X_i\}$$

with $1 \leq i \leq m$ and $1 \leq j \leq n$.

For the R-estimator of the location, the i.i.d. one-sample case is equivalent to the two-sample case with i.i.d. samples $\{X_1, ..., X_m\}$ and $\{2\Delta - X_1, ..., 2\Delta - X_m\}$. Thus, with the sign/Wilcoxon/vdW score, the rank statistic (2.3.3) yields the R-estimator

$$\hat{\Delta} = \text{med}\{\frac{1}{2}(X_i + X_j)\}$$

with $1 \leq i, j \leq m$.

In terms of the L-estimator, it is defined as a linear combination of order statistics; see Jurečková and Sen (1996), Huber and Ronchetti (2009) for more details of its definition. In particular, for the location model (2.3.1), the L-estimator $\hat{\theta}_n$ satisfies

$$\hat{\theta}_n = \sum_{i=1}^{n} c_{ni} X_{(i)}, \tag{2.3.4}$$

where the coefficients $\{c_{ni}; 1 \leq i \leq n\}$ are known and $\{X_{(i)}; 1 \leq i \leq n\}$ is the ordered sequence of the sample. Due to the simplicity of equation (2.3.4), the L-estimator is computationally appealing. Two classical L-estimators of the location defined in Bickel (1965) are as follows. They are robust since the outliers are either trimmed off or replaced.

(i). The α-trimmed mean. For a constant $\alpha \in [0, \frac{1}{2})$,

$$\hat{\theta}_n = (n - 2\lfloor n\alpha \rfloor)^{-1} \sum_{i=\lfloor n\alpha \rfloor + 1}^{n-\lfloor n\alpha \rfloor} X_{(i)}.$$

(ii). The α-Winsorized mean. Here

$$\hat{\theta}_n = n^{-1} \left[\lfloor n\alpha \rfloor X_{(\lfloor n\alpha \rfloor)} + \sum_{i=\lfloor n\alpha \rfloor + 1}^{n-\lfloor n\alpha \rfloor} X_{(i)} + \lfloor n\alpha \rfloor X_{(n-\lfloor n\alpha \rfloor + 1)} \right].$$

The robust estimators mentioned above for the location model have been extended to linear regression and some nonlinear models.

In terms of the estimation procedure for the GARCH model, the most commonly-employed one is the quasi-maximum likelihood estimator (QMLE), which is obtained by maximizing the normal likelihood function. In particular, it assumes that $\{\epsilon_t; 1 \leq t \leq n\}$ are i.i.d. normal distributed with mean zero. Hence, the QMLE $\hat{\boldsymbol{\theta}}_n$ is obtained by minimizing the negative log-likelihood function, i.e.,

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\min} \frac{1}{n} \sum_{t=1}^{n} \left[\log v_t(\boldsymbol{\theta}) + \frac{X_t^2}{v_t(\boldsymbol{\theta})} \right]$$

or equivalently solving the equation

$$\frac{1}{n} \sum_{t=1}^{n} \left[1 - \frac{X_t^2}{v_t(\boldsymbol{\theta})} \right] \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{v_t(\boldsymbol{\theta})} = \mathbf{0}.$$

The asymptotic normality of the QMLE was established by Weiss (1986) for the ARCH model and by Lee and Hansen (1994) and Lumsdaine (1996) for the GARCH$(1, 1)$ model. For the general GARCH(p, q) model, it was proved by Berkes et al. (2003). They assumed that $\mathrm{E}\epsilon_t^4 < \infty$ to prove that the QMLE converges at rate $\sqrt{n}$ to a normal distribution. However, financial time series were frequently reported with heavy-tails so this assumption may be violated. Therefore, various types of robust estimators were proposed in the literature. For the ARCH model, Horvath and Liese (2004) introduced a family of L_p-estimators based on

weighted M-estimators. They showed that for their L_1-estimator, with a suitable choice of the weight function, the asymptotic normality holds without any moment assumption when ϵ_t^2 admits a density. For the GARCH model, Peng and Yao (2003) considered three types of the LAD estimators, among which the one based on logarithmic transformation is unbiased and asymptotically normal when $\mathrm{E}\epsilon_t^2 < \infty$. Muler and Yohai (2008) defined two types of M-estimators through the minimization of some functions. Mukherjee (2008) proposed a class of M-estimators as the solution of an equation. It turns out that the QMLE, LAD and Huber's estimators all belong to this class of the M-estimators. He proved that for some M-estimators, the asymptotic normality holds only under the existence of fractional moment of the error distribution. For the R-estimator of the GARCH model, Andrews (2012) considered a procedure based on the logarithm transformation of the squared observations and conditional volatilities. The R-estimator was obtained through minimization of a rank-based residual dispersion function and it is similar to the one proposed by Jaeckel (1972) for the linear regression model.

2.4 The VARMA model

2.4.1 The ARMA model

Time series are usually autocorrelated, that is, for a time series $\{X_t; t \in \mathbb{Z}\}$, X_t is a function of its lagged series $X_{t-1}, X_{t-2}, \cdots$. A simple and popular model for accommodating this autocorrelation is the autoregressive (AR) model (with order p), which takes a linear form as follows

$$(1 - \sum_{i=1}^{p} a_i L^i) X_t = \epsilon_t, \quad t \in \mathbb{Z},$$

where L denotes the lag operator and $\{\epsilon_t; t \in \mathbb{Z}\}$ is the innovation term, which is usually assumed to be i.i.d.. The associated characteristic equation of this model

is

$$1 - \sum_{i=1}^{p} a_i x^i = 0. \tag{2.4.1}$$

If all solutions of this equation are greater than 1 in modulus, then the AR(p) model is stationary; see Tsay (2010).

A disadvantage of the AR(p) model is that a high-order model may be needed to fit a time series adequately well. One way to deal with this is to assume X_t is a linear function of lagged innovation terms, which leads to the moving-average (MA) model (with order q) that takes the form

$$X_t = (1 + \sum_{j=1}^{q} b_j L^j)\epsilon_t, \quad t \in \mathbb{Z}.$$

The associated characteristic equation of this model is

$$1 + \sum_{j=1}^{q} b_j x^j = 0. \tag{2.4.2}$$

For the MA(q) model to be invertible, it is required that all solutions of the above equation are greater than 1 in modulus.

A more flexible model describing the dependence structure of time series than the AR and MA models is obtained by combining them, which leads to the autoregressive moving-average (ARMA) model as follows

$$(1 - \sum_{i=1}^{p} a_i L^i)X_t = (1 + \sum_{j=1}^{q} b_j L^j)\epsilon_t, \quad t \in \mathbb{Z}.$$

An advantage of the ARMA model is that it often takes less parameters than the AR and MA models to fit a time series adequately well. It is often assumed that all solutions of both (2.4.1) and (2.4.2) are greater than 1 in modulus so that the ARMA(p, q) model is both stationary and invertible. Also, we assume that there are no common factors between the AR polynomial $1 - \sum_{i=1}^{p} a_i L^i$ and the MA polynoimal $1 + \sum_{j=1}^{q} b_j L^j$; otherwise the order (p, q) of the model can be reduced; see Tsay (2010).

2.4.2 The VARMA model

Let $\{\boldsymbol{X}_t; t \in \mathbb{Z}\}$ denote a d-dimensional time series and $\boldsymbol{I}_d$ denote a $d \times d$ identity matrix. The ARMA model can be generalized to the vector ARMA (VARMA) model that takes the form

$$\left(\boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{A}_i L^i\right) \boldsymbol{X}_t = \left(\boldsymbol{I}_k + \sum_{j=1}^{q} \boldsymbol{B}_j L^j\right) \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z},$$

where $\boldsymbol{A}_1, ..., \boldsymbol{A}_p, \boldsymbol{B}_1, ..., \boldsymbol{B}_q$ are $d \times d$ matrices and $\{\boldsymbol{\epsilon}_t; t \in \mathbb{Z}\}$ is the innovation term. The following assumption is needed for the model to be stationary and invertible:

(i) All solutions of the determinantal equations

$$\det\left(\sum_{i=0}^{p} \boldsymbol{A}_i z^i\right) = 0 \text{ and } \det\left(\sum_{i=0}^{q} \boldsymbol{B}_i z^i\right) = 0, \; z \in \mathbb{C}$$

lie outside the unit ball in $\mathbb{C}$;

(ii) $|\boldsymbol{A}_p| \neq 0 \neq |\boldsymbol{B}_q|$;

(iii) $\boldsymbol{I}_d$ is the greatest common left divisor of $\sum_{i=0}^{p} \boldsymbol{A}_i z^i$ and $\sum_{i=0}^{q} \boldsymbol{B}_i z^i$.

Chapter 3

M-estimation and its bootstrapped version in GARCH models

3.1 Introduction

Recall that a series $\{X_t; t \in \mathbb{Z}\}$ is said to follow a GARCH(p, q) model if

$$X_t = \sigma_t \epsilon_t, \tag{3.1.1}$$

where $\{\epsilon_t; t \in \mathbb{Z}\}$ are unobservable i.i.d. errors with mean zero and

$$\sigma_t = (\omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2)^{1/2}, \ t \in \mathbb{Z}, \tag{3.1.2}$$

with $\omega_0, \ \alpha_{0i}, \ \beta_{0j} > 0, \ \forall\, i, j$. Mukherjee (2008) proposed a class of M-estimators for estimating the GARCH parameter

$$\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})' \tag{3.1.3}$$

based on observations $\{X_t; 1 \leq t \leq n\}$. The M-estimators are asymptotically normal under some moment assumptions on the error distribution and are more robust than the commonly-used quasi maximum likelihood estimator (QMLE). Mukherjee (2020) considered a class of weighted bootstrap methods to approximate the distributions of these estimators and established the asymptotic validity of such bootstrap. In this chapter, we apply an iteratively re-weighted algorithm to compute the M-estimates and the corresponding bootstrap estimates with specific attention to Huber's, μ- and Cauchy-estimates which were not considered in the literature in details. The iteratively re-weighted algorithm turns out to be particularly useful in computing bootstrap replicates since it avoids the re-computation of some core quantities for new bootstrap samples.

The class of M-estimators includes the QMLE. The asymptotic normality and the asymptotic validity of bootstrapping the QMLE were derived under the finite fourth moment assumption on the error distribution. However, there are other M-estimators such as the μ-estimator and Cauchy-estimator which are asymptotic normal under mild assumption on the finiteness of lower order moments. Since heavy-tailed error distributions without higher order moments are common in the GARCH modeling of many real financial time series, it becomes worthwhile to use these estimators for such series but unfortunately they have not been investigated in the literature. One of the contributions of this chapter is to reveal precisely the importance of such alternative M-estimators to analyze financial data instead of using the QMLE.

In an earlier work, Muler and Yohai (2008) analyzed the Electric Fuel Corporation (EFCX) time series and fitted a GARCH(1, 1) model. Using exploratory analysis, they detected presence of outliers and considered estimation of parameters based on robust methods. It turned out that estimates based on different methods vary widely and so it is difficult to assess which method should be relied on in similar situations. In this chapter, we use M-estimates with mild assumptions on error moments to analyze the EFCX series.

Francq and Zakoïan (2009) underscored the importance of using higher order GARCH models such as GARCH(2, 1) for some real financial time series but the computation and simulation results for such models are not available widely in the literature. We investigate the role of M-estimators for the GARCH(2, 1) model through extensive simulations and real data analysis. We also provide simulation results and analysis for the GARCH(1, 2) model.

The chapter is organized as follows. Sections 3.2 and 3.3 set the background. In particular, we discuss the class of M-estimators and give examples in Section 3.2. Section 3.3 contains bootstrap formulation and the statement on the asymptotic validity of the bootstrap. Section 3.4 discusses computational aspects of M-estimators and its bootstrapped replicates. Section 3.5 reports simulation results for various M-estimators. Section 3.6 compares bootstrap approximation with the asymptotic normal approximation to distributions of M-estimators through simulation. Section 3.7 analyzes three real financial time series.

3.2 M-estimators of the GARCH parameters

Throughout this chapter, for a function g, we use $\dot{g}$ to denote its derivative whenever it exists. Also, $\text{sign}(x) := I(x > 0) - I(x < 0)$. For $x > 0$, $\log^+(x) := I(x > 1)\log(x)$. Moreover, ϵ will denote a generic r.v. having same distribution as errors $\{\epsilon_t\}$ of (3.1.1).

Let $\psi : \mathbb{R} \to \mathbb{R}$ be an odd function which is differentiable at all but finite number of points. Let $\mathcal{D} \subset \mathbb{R}$ denote the set of points where ψ is differentiable and let $\bar{\mathcal{D}}$ denote its complement. Let $H(x) := x\psi(x)$, $x \in \mathbb{R}$ so that H is symmetric. The function H is called the *score function* of the M-estimation in the scale model. Some of the following examples were considered in Mukherjee (2008).

Example 1. QMLE score: Let $\psi(x) = x$. Then $H(x) = x^2$.

Example 2. LAD score: Let $\psi(x) = \text{sign}\,(x)$. Then $\bar{\mathcal{D}} = \{0\}$ and $H(x) = |x|$.

Example 3. Huber's k score: Let $\psi(x) = xI(|x| \le k) + k\,\text{sign}\,(x)I(|x| > k)$, where $k > 0$ is a known constant. Then $\bar{\mathcal{D}} = \{-k, k\}$ and $H(x) = x^2 I(|x| \le$

$k) + k|x|I(|x| > k)$.

Example 4. Score function for the maximum likelihood estimation (MLE): Let $\psi(x) = -\dot{f}(x)/f(x)$, where f is the true density of ϵ, assumed to be known. Then $H(x) = x\{-\dot{f}(x)/f(x)\}$.

Example 5. μ score: Let $\psi(x) = \mu\,\text{sign}(x)/(1+|x|)$, where $\mu > 1$ is a known constant. Then $\bar{\mathcal{D}} = \{0\}$ and $H(x) = \mu|x|/(1+|x|)$ is bounded.

Example 6. Cauchy score: Let $\psi(x) = 2x/(1+x^2)$. Then $H(x) = 2x^2/(1+x^2)$ is bounded.

Example 7. Score function for the exponential pseudo-maximum likelihood estimation: Let $\psi(x) = \delta_1|x|^{\delta_2-1}\text{sign}(x)$, where $\delta_1 > 0$ and $1 < \delta_2 \leq 2$ are known constants. Here $\bar{\mathcal{D}} = \{0\}$ and $H(x) = \delta_1|x|^{\delta_2}$.

Assume that for some $\kappa_1 \geq 2$ and $\kappa_2 > 0$,

$$\text{E}(|\epsilon|^{\kappa_1}) < \infty \text{ and } \lim_{t\to 0} \text{P}(\epsilon^2 < t)/t^{\kappa_2} = 0. \tag{3.2.1}$$

Then σ_t^2 of (3.1.2) has the following unique almost sure representation:

$$\sigma_t^2 = c_0 + \sum_{j=1}^{\infty} c_j X_{t-j}^2, \ \ t \in \mathbb{Z}, \tag{3.2.2}$$

where $\{c_j; j \geq 0\}$ are defined in (2.9)-(2.16) of Berkes et al. (2003).

Let $\boldsymbol{\Theta}$ be a compact subset of $(0,\infty)^{1+p} \times (0,1)^q$. A typical element in $\boldsymbol{\Theta}$ is denoted by $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$. Define the variance function on $\boldsymbol{\Theta}$ by

$$v_t(\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} c_j(\boldsymbol{\theta}) X_{t-j}^2, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, t \in \mathbb{Z}, \tag{3.2.3}$$

where the coefficients $\{c_j(\boldsymbol{\theta}); j \geq 0\}$ are given in Berkes et al. (2003) (Section 3, and display (3.1)) with the property

$$c_j(\boldsymbol{\theta}_0) = c_j, \ \forall j \geq 0. \tag{3.2.4}$$

Hence the variance functions satisfy $v_t(\boldsymbol{\theta}_0) = \sigma_t^2$, $t \in \mathbb{Z}$. Using (3.2.4), (3.1.1) can be rewritten as

$$X_t = \{v_t(\boldsymbol{\theta}_0)\}^{1/2}\epsilon_t, \;\; 1 \le t \le n. \tag{3.2.5}$$

Consider observable approximation $\{\hat{v}_t(\boldsymbol{\theta})\}$ of the process $\{v_t(\boldsymbol{\theta})\}$ of (3.2.3) defined by

$$\hat{v}_t(\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + I(2 \le t)\sum_{j=1}^{t-1} c_j(\boldsymbol{\theta})X_{t-j}^2, \;\; \boldsymbol{\theta} \in \boldsymbol{\Theta}, \;\; 1 \le t \le n. \tag{3.2.6}$$

Then an M-estimator $\hat{\boldsymbol{\theta}}_n$ is defined as the solution of $\widehat{\boldsymbol{M}}_{n,H}(\boldsymbol{\theta}) = \mathbf{0}$, where

$$\widehat{\boldsymbol{M}}_{n,H}(\boldsymbol{\theta}) := \sum_{t=1}^{n}\left\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\right\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}. \tag{3.2.7}$$

Next we describe the iterative relation of $\{c_j(\boldsymbol{\theta})\}$ that is used to write computer program for their numerical evaluation. The computation is discussed in Section 3.4.

Example 1. GARCH$(1,1)$ model: With $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$,

$$c_0(\omega, \alpha, \beta) = \omega/(1-\beta), \; c_j(\omega, \alpha, \beta) = \alpha\beta^{j-1}, \;\; j \ge 1.$$

Example 2. GARCH$(2,1)$ model: With $\boldsymbol{\theta} = (\omega, \alpha_1, \alpha_2, \beta)'$,

$$c_0(\boldsymbol{\theta}) = \omega/(1-\beta), c_1(\boldsymbol{\theta}) = \alpha_1, c_2(\boldsymbol{\theta}) = \alpha_2 + \beta c_1(\boldsymbol{\theta}) = \alpha_2 + \beta\alpha_1$$

and

$$c_j(\boldsymbol{\theta}) = \beta c_{j-1}(\boldsymbol{\theta}), \;\; j \ge 3.$$

Example 3. GARCH$(1,2)$ model: With $\boldsymbol{\theta} = (\omega, \alpha, \beta_1, \beta_2)'$,

$$c_0(\boldsymbol{\theta}) = \omega/(1-\beta_1-\beta_2), c_1(\boldsymbol{\theta}) = \alpha, c_2(\boldsymbol{\theta}) = \beta_1 c_1(\boldsymbol{\theta}) = \beta_1\alpha,$$

and

$$c_j(\boldsymbol{\theta}) = \beta_1 c_{j-1}(\boldsymbol{\theta}) + \beta_2 c_{j-2}(\boldsymbol{\theta}), \;\; j \ge 3.$$

Example 4. GARCH(2, 2) model: With $\boldsymbol{\theta} = (\omega, \alpha_1, \alpha_2, \beta_1, \beta_2)'$,

$$c_0(\boldsymbol{\theta}) = \omega/(1 - \beta_1 - \beta_2), c_1(\boldsymbol{\theta}) = \alpha_1, c_2(\boldsymbol{\theta}) = \alpha_2 + \beta_1\alpha_1$$

and

$$c_j(\boldsymbol{\theta}) = \beta_1 c_{j-1}(\boldsymbol{\theta}) + \beta_2 c_{j-2}(\boldsymbol{\theta}),\ j \geq 3.$$

3.2.1 Asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$

The asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ is derived under the following assumptions.

Model assumptions: The parameter space $\boldsymbol{\Theta}$ is a compact set and its interior $\boldsymbol{\Theta}_0$ contains both $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_{0H}$ of (3.1.3) and (3.2.10), respectively. Moreover, (3.2.1), (3.2.3) and (3.2.5) hold and $\{X_t\}$ is stationary and ergodic.

Conditions on the score function:

Identifiability condition: Corresponding to the score function H, there exists a unique number $c_H > 0$ satisfying

$$\mathrm{E}[H(\epsilon/c_H^{1/2})] = 1. \tag{3.2.8}$$

Moment conditions:

$$\mathrm{E}[H(\epsilon/c_H^{1/2})]^2 < \infty \text{ and } 0 < \mathrm{E}[(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})] < \infty. \tag{3.2.9}$$

Also various *Smoothness conditions* on H as in Mukherjee (2008) are assumed which are satisfied in all examples of H considered above. Define the score function factor

$$\sigma^2(H) := 4\ \mathrm{Var}\{H(\epsilon/c_H^{1/2})\}/\{\mathrm{E}[(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})]\}^2,$$

the matrix

$$\boldsymbol{G} := \mathrm{E}\{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0H})\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0H})/v_1^2(\boldsymbol{\theta}_{0H})\}$$

and the transformed parameter

$$\boldsymbol{\theta}_{0H} = (c_H\omega_0, c_H\alpha_{01}, \ldots, c_H\alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})'. \tag{3.2.10}$$

Theorem 3.2.1. *(Mukherjee 2008) Suppose that the model assumptions, identifiability condition, moment conditions and smoothness conditions hold. Then*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0H}) \to \mathcal{N}(0, \sigma^2(H)\boldsymbol{G}^{-1}). \tag{3.2.11}$$

Note that c_H used in above formulas are given by (i) $c_H = \mathrm{E}(\epsilon^2)$ for the QMLE, (ii) $c_H = (\mathrm{E}|\epsilon|)^2$ for the LAD while for the Huber, μ-estimator, Cauchy and other scores, c_H does not have closed-form expression. For such score functions, c_H is calculated using (3.2.8) as follows. We fix a large positive integer I and generate $\{\epsilon_i; 1 \le i \le I\}$ from the error distribution considered for the simulation. Then, using the bisection method on $c > 0$, we solve the equation

$$(1/I)\sum_{i=1}^{I}\left\{H\left(\epsilon_i/c^{1/2}\right)\right\} - 1 = 0.$$

Values of c_H computed in this way were provided in Mukherjee (2008, page 1541) for some error distributions and score functions. In Table 3.1 we provide c_H for few more error distributions and score functions such as Huber's k-score and μ-estimator with $k = 1.5$ and $\mu = 3$ which are used in simulations and data analysis of later sections. In the sequel, Student's t-distributions with ν degrees of freedom is abbreviated as $t(\nu)$. The double exponential distribution refers to the Laplace distribution, and it is abbreviated as DE. Note that we select $t(2.2)$ instead of $t(2)$ since the latter does not have finite second moment.

3.3 Bootstrapping M-estimators

In this section, we describe the bootstrap formulation of Mukherjee (2020) for M-estimators in GARCH models. Let $\{w_{nt}; 1 \le t \le n, n \ge 1\}$ be a triangular

Table 3.1: Values of c_H for M-estimators (Huber, μ-, Cauchy) under various error distributions.

	Huber's	μ-estimator	Cauchy
Normal	0.825	1.692	0.377
DE	0.677	1.045	0.207
Logistic	0.781	1.487	0.316
$t(3)$	0.533	0.850	0.172
$t(2.2)$	0.204	0.274	0.053

array of r.v.'s such that for each $n \geq 1$, $\{w_{nt}; 1 \leq t \leq n\}$ are exchangeable and independent of the data $\{X_t; t \geq 1\}$ and errors $\{\epsilon_t; t \geq 1\}$. Also, $\forall t \geq 1$, $w_{nt} \geq 0$ and $\mathrm{E}(w_{nt}) = 1$.

Based on these weights, bootstrap estimate $\hat{\boldsymbol{\theta}}_{*n}$ is defined as the solution of $\widehat{\boldsymbol{M}}^*_{n,H}(\boldsymbol{\theta}) = \mathbf{0}$, where

$$\widehat{\boldsymbol{M}}^*_{n,H}(\boldsymbol{\theta}) := \sum_{t=1}^{n} w_{nt} \Big\{ 1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\} \Big\} \{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}. \tag{3.3.1}$$

Examples. From many different choices of bootstrap weights, we consider the following three schemes in Chatterjee and Bose (2005) and Mukherjee (2020) for comparison.

(i) Scheme M. The sequence of weights $\{w_{n1}, \ldots, w_{nn}\}$ has a multinomial $(n, 1/n, \ldots, 1/n)$ distribution, which is essentially the classical paired bootstrap.

(ii) Scheme E. When $w_{nt} = (nE_t)/\sum_{i=1}^n E_i$, where $\{E_t\}$ are i.i.d. exponential r.v. with mean 1. Under scheme E, $\hat{\boldsymbol{\theta}}_{*n}$ is a weighted M-estimator with weights proportional to E_t, $1 \leq i \leq n$.

(iii) Scheme U. When $w_{nt} = (nU_t)/\sum_{i=1}^n U_i$, where $\{U_t\}$ are i.i.d. uniform r.v. on $(0.5, 1.5)$. Under scheme U, $\hat{\boldsymbol{\theta}}_{*n}$ is a weighted M-estimator with weights proportional to U_t, $1 \leq i \leq n$.

A host of other bootstrap methods in the literature are special cases of the above bootstrap formulation. Such general formulation of weighted bootstrap offers a unified way of studying several bootstrap schemes simultaneously. See, for example, Chatterjee and Bose (2005) for details in other contexts.

We assume that the weights satisfy the following basic conditions (Conditions BW of Chatterjee and Bose (2005)) where $\sigma_n^2 = \mathrm{Var}(w_{ni})$ and $k_3 > 0$ is a constant.

$$\mathrm{E}(w_{n1}) = 1,\ 0 < k_3 < \sigma_n^2 = o(n) \text{ and } \mathrm{Corr}\,(w_{n1}, w_{n2}) = O(1/n). \tag{3.3.2}$$

Under (3.3.2) and some additional smoothness and moment conditions in Mukherjee (2020), weighted bootstrap is asymptotic valid.

Theorem 3.3.1. *(Mukherjee 2020) For almost all data, as $n \to \infty$,*

$$\sigma_n^{-1} n^{1/2} (\hat{\boldsymbol{\theta}}_{*n} - \hat{\boldsymbol{\theta}}_n) \to \mathcal{N}(0, \sigma^2(H) \boldsymbol{G}^{-1}). \tag{3.3.3}$$

We remark that since $0 < 1/\sigma_n < 1/\sqrt{k_3}$, the rate of convergence of the bootstrap estimator is the same as that of the original estimator. The standard deviation of the weights $\{\sigma_n\}$ at the denominator of the scaling reflects the contribution of the corresponding weights.

The distributional result of (3.3.3) is useful for constructing the confidence interval of the GARCH parameters as follows. Let B denote the number of bootstrap replicates, γ_0 denote a generic parameter (one of ω_0, α_{0i} or β_{0j}) and let $\hat{\gamma}_n$ and $\hat{\gamma}_{*nb}$ denote its M-estimator and b-th bootstrap estimator ($1 \le b \le B$), respectively. Let γ_{0H} be one of $c_H\omega_0$, $c_H\alpha_{0i}$ or β_{0j}, as appropriate, which has a known value for a simulation experiment. Using the approximation of $\sqrt{n}(\hat{\gamma}_n - \gamma_{0H})$ by $\sigma_n^{-1} n^{1/2} (\hat{\gamma}_{*n} - \hat{\gamma}_n)$, the bootstrap confidence interval of γ_{0H} is of the form

$$[\hat{\gamma}_n - n^{-1/2}\{\sigma_n^{-1} n^{1/2} (\hat{\gamma}_{*n,\alpha/2} - \hat{\gamma}_n)\}, \hat{\gamma}_n + n^{-1/2}\{\sigma_n^{-1} n^{1/2} (\hat{\gamma}_{*n,1-\alpha/2} - \hat{\gamma}_n)\}] \tag{3.3.4}$$

where $\hat{\gamma}_{*n,\alpha/2}$ is the $\alpha/2$-th quantile of the numbers $\{\hat{\gamma}_{*nb}, 1 \le b \le B\}$. Consequently, the bootstrap coverage probability is computed by the proportion of the above set of B confidence intervals containing γ_{0H}.

Similarly, using (3.2.11) of Theorem 3.3.1, we can obtain the confidence interval of γ_{0H} based on the asymptotic normality of $\hat{\gamma}_n$, and this will be called the *normal*

confidence interval. Specifically, in view of Proposition 3.1 of Mukherjee (2008) on the estimation of the variance-covariance matrix $\sigma^2(H)\boldsymbol{G}^{-1}$, we can obtain the asymptotic confidence interval of γ_{0H} as

$$[\hat{\gamma}_n - n^{-1/2}\hat{d}z_{1-\alpha/2}, \hat{\gamma}_n + n^{-1/2}\hat{d}z_{1-\alpha/2}], \tag{3.3.5}$$

where $(\hat{d})^2$ is the estimated variance of $\hat{\gamma}_n$ obtained from the appropriate diagonal entry of the estimator of $\sigma^2(H)\boldsymbol{G}^{-1}$ and $z_{1-\alpha/2}$ is the $1-\alpha/2$-th quantile of the standard normal distribution.

In the following Section 3.6, we will compare the accuracy of the confidence intervals constructed by the bootstrap and asymptotic approximations.

3.4 Algorithm

We discuss the implementation of an iteratively re-weighted algorithm proposed in Mukherjee (2020) for computing M-estimates. In particular, we highlight μ-estimate and Cauchy-estimate of the GARCH parameters in this chapter since their asymptotic distributions are derived under mild moment assumptions but they were not consider in the literature before. We also consider the bootstrap estimators based on the corresponding score functions.

3.4.1 Computation of M-estimates

For the convenience of writing, let $\alpha(c) = \mathrm{E}[H(c\epsilon)]$ for $c > 0$. Using a Taylor expansion of $\widehat{\boldsymbol{M}}_{n,H}$, we obtain the following recursive equation for computing the updated estimate $\tilde{\boldsymbol{\theta}}$ of $\hat{\boldsymbol{\theta}}_n$ from the current estimate $\boldsymbol{\theta}$ of $\widehat{\boldsymbol{M}}_{n,H}(\boldsymbol{\theta}) = \boldsymbol{0}$:

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \{\dot{\alpha}(1)/2\}^{-1}\Big[\sum_{t=1}^{n} \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'/\hat{v}_t^2(\boldsymbol{\theta})\Big]^{-1} \sum_{t=1}^{n} \Big\{H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\} - 1\Big\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}, \tag{3.4.1}$$

where $\dot{\alpha}(1) = \mathrm{E}\{\epsilon\dot{H}(\epsilon)\}$ under smoothness conditions on H. Since the GARCH residuals $\{X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n)\}$ estimate only $\{\epsilon_t/c_H^{1/2}\}$, in general, we cannot estimate

$\dot{\alpha}(1)$ from the data. Therefore, we use ad hoc techniques such as simulating $\{\tilde{\epsilon}_t; 1 \leq t \leq n\}$ from $\mathcal{N}(0,1)$ or standardized DE distribution and then use $n^{-1}\sum_{t=1}^{n}\tilde{\epsilon}\dot{H}(\tilde{\epsilon})$ to carry out the iteration. Note that if the iteration in (3.4.1) converges then $\tilde{\boldsymbol{\theta}} \approx \boldsymbol{\theta}$. Therefore in this case from (3.4.1), $\widehat{\boldsymbol{M}}_{n,H}(\boldsymbol{\theta}) \approx \mathbf{0}$ and hence $\tilde{\boldsymbol{\theta}}$ is the desired $\hat{\boldsymbol{\theta}}_n$. Based on our extensive simulation study and real data analysis, the algorithm is robust enough to converge to the same value of $\hat{\boldsymbol{\theta}}_n$ irrespective of different values of the unknown factor $\dot{\alpha}(1)$ used in computation.

In the following examples, we discuss (3.4.1) when specialized to the M-estimators computed in this chapter.

QMLE: Here $H(x) = x^2$ and $\alpha(c) = c^2\mathrm{E}(\epsilon^2)$. Hence $\dot{\alpha}(1)/2 = \mathrm{E}(\epsilon^2)$ and

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \Big\{\mathrm{E}(\epsilon^2)\Big\}^{-1}\Big\{\sum_{t=1}^{n}\Big[\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'/\hat{v}_t^2(\boldsymbol{\theta})\Big]\Big\}^{-1}\sum_{t=1}^{n}\Big\{[X_t^2/\hat{v}_t(\boldsymbol{\theta})] - 1\Big\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}.$$

With

$$W_t = 1/\hat{v}_t^2(\boldsymbol{\theta}), x_t = \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}), y_t = X_t^2 - \hat{v}_t(\boldsymbol{\theta}),$$

$\tilde{\boldsymbol{\theta}}$ can be computed iteratively as

$$\tilde{\boldsymbol{\theta}}_{(r+1)} = \tilde{\boldsymbol{\theta}}_{(r)} + \Big\{\mathrm{E}(\epsilon^2)\Big\}^{-1}\left\{\sum_t W_t x_t x_t'\right\}^{-1}\left\{\sum_t W_t x_t y_t\right\}.$$

Note that when $\mathrm{E}(\epsilon^2) = 1$, this is same as the formula obtained through the BHHH algorithm proposed by Berndt et al. (1974).

LAD: Here $H(x) = |x|$ and $\alpha(c) = c\mathrm{E}|\epsilon|$. Hence $\dot{\alpha}(1) = \mathrm{E}|\epsilon|$ and

$$\begin{aligned}
\tilde{\boldsymbol{\theta}} &= \boldsymbol{\theta} + \{2/\mathrm{E}|\epsilon|\}\Big\{\sum_{t=1}^{n}\Big[\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'/\hat{v}_t^2(\boldsymbol{\theta})\Big]\Big\}^{-1}\sum_{t=1}^{n}\Big\{|X_t|/\hat{v}_t^{1/2}(\boldsymbol{\theta}) - 1\Big\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\} \\
&= \boldsymbol{\theta} + \{2/\mathrm{E}|\epsilon|\}\Big\{\sum_{t=1}^{n}\Big[\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'/\hat{v}_t^2(\boldsymbol{\theta})\Big]\Big\}^{-1}\sum_{t=1}^{n}\Big\{|X_t| - \hat{v}_t^{1/2}(\boldsymbol{\theta})\Big\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t^{3/2}(\boldsymbol{\theta})\} \\
&= \boldsymbol{\theta} + \{2/\mathrm{E}|\epsilon|\}\Big\{\sum_{t=1}^{n}\Big[\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'/\hat{v}_t^2(\boldsymbol{\theta})\Big]\Big\}^{-1}\sum_{t=1}^{n}\Big\{\hat{v}_t^{1/2}(\boldsymbol{\theta})(|X_t| - \hat{v}_t^{1/2}(\boldsymbol{\theta}))\Big\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t^2(\boldsymbol{\theta})\}.
\end{aligned}$$

With

$$W_t = 1/\hat{v}_t^2(\boldsymbol{\theta}), x_t = \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}), y_t = \hat{v}_t^{1/2}(\boldsymbol{\theta})(|X_t| - \hat{v}_t^{1/2}(\boldsymbol{\theta})),$$

$\tilde{\boldsymbol{\theta}}$ can be computed iteratively as

$$\tilde{\boldsymbol{\theta}}_{(r+1)} = \tilde{\boldsymbol{\theta}}_{(r)} + \{2/\mathrm{E}|\epsilon|\} \left\{ \sum_t W_t x_t x_t' \right\}^{-1} \left\{ \sum_t W_t x_t y_t \right\}.$$

Huber: Here $H(x) = x^2 I(|x| \leq k) + k|x| I(|x| > k)$ and

$$\alpha(c) = \mathrm{E}\left[(c\epsilon)^2 I(|c\epsilon| \leq k) + k|c\epsilon| I(|c\epsilon| > k) \right].$$

Hence

$$\dot{\alpha}(1) = \mathrm{E}\left[2\epsilon^2 I(|\epsilon| \leq k) + k|\epsilon| I(|\epsilon| > k) \right]$$

and

$$\begin{aligned} \tilde{\boldsymbol{\theta}} &= \boldsymbol{\theta} - \left\{ \dot{\alpha}(1)/2 \right\}^{-1} \left\{ \sum_{t=1}^n \left[\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}) \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'}{\hat{v}_t^2(\boldsymbol{\theta})} \right] \right\}^{-1} \\ &\quad \times \sum_{t=1}^n \left[1 - \frac{X_t^2}{\hat{v}_t(\boldsymbol{\theta})} I\left(\frac{|X_t|}{\hat{v}_t^{1/2}(\boldsymbol{\theta})} \leq k \right) - k \frac{|X_t|}{\hat{v}_t^{1/2}(\boldsymbol{\theta})} I\left(\frac{|X_t|}{\hat{v}_t^{1/2}(\boldsymbol{\theta})} > k \right) \right] \left\{ \frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})}{\hat{v}_t(\boldsymbol{\theta})} \right\}. \end{aligned}$$

With

$$W_t = 1/\hat{v}_t^2(\boldsymbol{\theta}), x_t = \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})$$

and

$$y_t = X_t^2 I\left(|X_t|/\hat{v}_t^{1/2}(\boldsymbol{\theta}) \leq k \right) + k|X_t| \hat{v}_t^{1/2}(\boldsymbol{\theta}) I\left(|X_t|/\hat{v}_t^{1/2}(\boldsymbol{\theta}) > k \right) - \hat{v}_t(\boldsymbol{\theta}),$$

$\tilde{\boldsymbol{\theta}}$ can be computed iteratively as

$$\tilde{\boldsymbol{\theta}}_{(r+1)} = \tilde{\boldsymbol{\theta}}_{(r)} + \left\{ \dot{\alpha}(1)/2 \right\}^{-1} \left\{ \sum_t W_t x_t x_t' \right\}^{-1} \left\{ \sum_t W_t x_t y_t \right\}.$$

μ-estimator: Here $H(x) = \mu|x|/(1 + |x|)$ and $\alpha(c) = \mu - \mu \mathrm{E}\left[1/(1 + |c\epsilon|)\right]$.

Hence

$$\dot{\alpha}(1) = \mu \mathrm{E}\left[|\epsilon|/(1+|\epsilon|)^2\right]$$

and

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \left\{\frac{\mu}{2}\mathrm{E}\left[\frac{|\epsilon|}{(1+|\epsilon|)^2}\right]\right\}^{-1}\left\{\sum_{t=1}^{n}\left[\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'}{\hat{v}_t^2(\boldsymbol{\theta})}\right]\right\}^{-1}\sum_{t=1}^{n}\left[\frac{\mu|X_t|}{\hat{v}_t^{1/2}(\boldsymbol{\theta})+|X_t|}-1\right]\left\{\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})}{\hat{v}_t(\boldsymbol{\theta})}\right\}.$$

With

$$W_t = 1/\hat{v}_t^2(\boldsymbol{\theta}), x_t = \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}), y_t = \frac{\mu|X_t|\hat{v}_t(\boldsymbol{\theta})}{\hat{v}_t^{1/2}(\boldsymbol{\theta})+|X_t|} - \hat{v}_t(\boldsymbol{\theta}),$$

$\tilde{\boldsymbol{\theta}}$ can be computed iteratively as

$$\tilde{\boldsymbol{\theta}}_{(r+1)} = \tilde{\boldsymbol{\theta}}_{(r)} + \left\{\frac{\mu}{2}\mathrm{E}\left[\frac{|\epsilon|}{(1+|\epsilon|)^2}\right]\right\}^{-1}\left\{\sum_t W_t x_t x_t'\right\}^{-1}\left\{\sum_t W_t x_t y_t\right\}.$$

Cauchy-estimator: Here $H(x) = 2x^2/(1+x^2)$ and $\alpha(c) = \mathrm{E}\left[2c^2\epsilon^2/(1+c^2\epsilon^2)\right]$.

Hence

$$\dot{\alpha}(1) = \mathrm{E}\left[4\epsilon^2/(1+\epsilon^2)^2\right]$$

and

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \left\{2\mathrm{E}\left[\frac{\epsilon^2}{(1+\epsilon^2)^2}\right]\right\}^{-1}\left\{\sum_{t=1}^{n}\left[\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'}{\hat{v}_t^2(\boldsymbol{\theta})}\right]\right\}^{-1}\sum_{t=1}^{n}\left[1-\frac{2X_t^2}{\hat{v}_t(\boldsymbol{\theta})+X_t^2}\right]\left\{\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})}{\hat{v}_t(\boldsymbol{\theta})}\right\}.$$

With

$$W_t = 1/\hat{v}_t^2(\boldsymbol{\theta}), x_t = \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}), y_t = \frac{2X_t^2\hat{v}_t(\boldsymbol{\theta})}{\hat{v}_t(\boldsymbol{\theta})+X_t^2} - \hat{v}_t(\boldsymbol{\theta}),$$

$\tilde{\boldsymbol{\theta}}$ can be computed iteratively as

$$\tilde{\boldsymbol{\theta}}_{(r+1)} = \tilde{\boldsymbol{\theta}}_{(r)} + \left\{2\mathrm{E}\left[\frac{\epsilon^2}{(1+\epsilon^2)^2}\right]\right\}^{-1}\left\{\sum_t W_t x_t x_t'\right\}^{-1}\left\{\sum_t W_t x_t y_t\right\}.$$

3.4.2 Computation of bootstrap M-estimates

Here the relevant function is $\widehat{\boldsymbol{M}}^{*}_{n,H}(\boldsymbol{\theta})$ defined in (3.3.1) and the bootstrap estimate $\hat{\boldsymbol{\theta}}_{*n}$, according to Mukherjee (2020), can be computed using the updating equation

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_{*} &= \boldsymbol{\theta} - \{2/\dot{\alpha}(1)\}\Big\{\sum_{t=1}^{n} w_{nt}\Big[\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})'/\hat{v}_t^2(\boldsymbol{\theta})\Big]\Big\}^{-1} \\ &\quad \times \sum_{t=1}^{n} w_{nt}\Big\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\Big\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}. \qquad (3.4.2)\end{aligned}$$

Notice also that weighted bootstrap is particularly computation-friendly and is easy to program in R. In particular, one can store

$$\Big\{1 - H\{X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})\}\Big\}\{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})/\hat{v}_t(\boldsymbol{\theta})\}$$

while computing M-estimates once and for all. After that, one simply needs to generate weights and compute the weighted sum while solving the above equation through iteration. Each time, the initial bootstrap estimator is taken to be the M-estimator $\hat{\boldsymbol{\theta}}_n$.

3.5 Simulating the distributions of M-estimators

To compare performance of various M-estimators via bias and MSE, we simulate n observations from GARCH(p, q) models with specific choice of parameters and error distributions and compute M-estimates based on various score functions. This procedure is replicated R-times to enable the estimation of bias and MSE. For illustration with $p = 1 = q$, let $\hat{\boldsymbol{\theta}}_n = (\hat{\omega}_r, \hat{\alpha}_r, \hat{\beta}_r)'$ be the M-estimator of $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \beta_{01})'$ based on a specified score function H at the r-th replication, $1 \le r \le R$. Notice that $(\hat{\omega}_r, \hat{\alpha}_r, \hat{\beta}_r)$ estimates $(c_H\omega_0, c_H\alpha_0, \beta_0)$, where c_H depends on both the score function and the underlying error distribution but is known in a simulation scenario. Therefore, to compare the performance for a specified error

distribution across various score functions, we consider R replicates of

$$(\hat{\omega}_r/c_H - \omega_0, \hat{\alpha}_r/c_H - \alpha_0, \hat{\beta}_r - \beta_0)'$$

and use the following vectors to estimate the *standardized bias* and the *standardized MSE*:

$$(R^{-1}\sum_{r=1}^{R}\{\hat{\omega}_r/c_H - \omega_0\}, R^{-1}\sum_{r=1}^{R}\{\hat{\alpha}_r/c_H - \alpha_0\}, R^{-1}\sum_{r=1}^{R}\{\hat{\beta}_r - \beta_0\})', \tag{3.5.1}$$

$$(R^{-1}\sum_{r=1}^{R}\{\hat{\omega}_r/c_H - \omega_0\}^2, R^{-1}\sum_{r=1}^{R}\{\hat{\alpha}_r/c_H - \alpha_0\}^2, R^{-1}\sum_{r=1}^{R}\{\hat{\beta}_r - \beta_0\}^2)'.$$

In Tables 3.2 and 3.3, we report the standardized bias and MSE of Huber's and μ-estimator to guide our choice of the corresponding tuning parameters k and μ. The underlying data generating process (DGP) is the GARCH(1, 1) model with $\boldsymbol{\theta}_0 = (1.65 \times 10^{-5}, 0.0701, 0.901)'$ under three types of innovation distributions: the normal, DE and logistic distribution. The above value of the true parameter is motivated from the estimated parameter of the GARCH(1, 1) model for the Shanghai Stock Exchange (SSE) Index data which will be analyzed later in this chapter. We use (3.5.1) for the computation with sample size $n = 1000$ and $R = 150$ replications.

The simulation results in Table 3.2 and Table 3.3 show that the bias and MSE of Huber's k-estimator and μ-estimator do not vary widely for various values of k and μ. Therefore $k = 1.5$ and $\mu = 3$ are chosen for subsequent computations. Notice also that the minimum bias and MSE correspond to $\mu = 3$ in a number of cases.

M-estimators corresponding to different score functions for the GARCH(1, 1) models have been compared under various error distributions via simulation study in Iqbal and Mukherjee (2010). However, as mentioned by Francq and Zakoïan (2009), higher order GARCH models are also common for some real financial time series but the computation and simulation results for such models are not available

Table 3.2: The standardized bias and MSE of the Huber's estimator (with different k values being used) under various error distributions (sample size $n = 1000$; $R = 150$ replications).

	Standardized bias			**Standardized MSE**		
	ω	α	β	ω	α	β
Normal						
k=1	1.03×10^{-5}	-2.44×10^{-3}	-1.96×10^{-2}	2.62×10^{-10}	4.20×10^{-4}	1.54×10^{-3}
k=1.5	1.22×10^{-5}	2.47×10^{-3}	-1.98×10^{-2}	3.33×10^{-10}	4.55×10^{-4}	1.58×10^{-3}
k=2.5	1.14×10^{-5}	-4.33×10^{-4}	-2.02×10^{-2}	3.10×10^{-10}	3.71×10^{-4}	1.58×10^{-3}
DE						
k=1	7.24×10^{-6}	1.29×10^{-3}	-1.57×10^{-2}	1.87×10^{-10}	4.65×10^{-4}	1.58×10^{-3}
k=1.5	7.32×10^{-6}	1.67×10^{-3}	-1.63×10^{-2}	2.00×10^{-10}	4.82×10^{-4}	1.68×10^{-3}
k=2.5	8.27×10^{-6}	2.94×10^{-3}	-1.92×10^{-2}	2.79×10^{-10}	5.60×10^{-4}	2.22×10^{-3}
Logistic						
k=1	9.87×10^{-6}	2.15×10^{-3}	-2.03×10^{-2}	3.18×10^{-10}	5.25×10^{-4}	2.28×10^{-3}
k=1.5	1.00×10^{-5}	2.04×10^{-3}	-2.04×10^{-2}	3.11×10^{-10}	4.89×10^{-4}	2.22×10^{-3}
k=2.5	1.06×10^{-5}	2.18×10^{-3}	-2.16×10^{-2}	3.18×10^{-10}	4.84×10^{-4}	2.17×10^{-3}

Table 3.3: The standardized bias and MSE of μ-estimator (with different μ values being used) under various error distributions (sample size $n = 1000$; $R = 150$ replications).

	Standardized bias			**Standardized MSE**		
	ω	α	β	ω	α	β
Normal						
μ=2	1.17×10^{-5}	2.97×10^{-3}	-2.13×10^{-2}	4.05×10^{-10}	6.73×10^{-4}	2.16×10^{-3}
μ=2.5	1.14×10^{-5}	1.80×10^{-3}	-2.12×10^{-2}	3.77×10^{-10}	5.71×10^{-4}	2.04×10^{-3}
μ=3	1.14×10^{-5}	1.36×10^{-3}	-2.11×10^{-2}	3.68×10^{-10}	5.21×10^{-4}	1.97×10^{-3}
DE						
μ=2	7.39×10^{-6}	2.23×10^{-3}	-1.49×10^{-2}	2.74×10^{-10}	7.20×10^{-4}	2.21×10^{-3}
μ=2.5	7.36×10^{-6}	1.50×10^{-3}	-1.52×10^{-2}	2.68×10^{-10}	6.56×10^{-4}	2.16×10^{-3}
μ=3	7.40×10^{-6}	1.25×10^{-3}	-1.53×10^{-2}	2.62×10^{-10}	6.17×10^{-4}	2.09×10^{-3}
Logistic						
μ=2	7.73×10^{-6}	2.22×10^{-3}	-1.37×10^{-2}	2.45×10^{-10}	6.79×10^{-4}	1.99×10^{-3}
μ=2.5	7.66×10^{-6}	9.77×10^{-4}	-1.41×10^{-2}	2.48×10^{-10}	5.88×10^{-4}	1.97×10^{-3}
μ=3	7.72×10^{-6}	5.99×10^{-4}	-1.42×10^{-2}	2.54×10^{-10}	5.44×10^{-4}	1.94×10^{-3}

widely in the literature. Therefore, below we focus on comparing M-estimators with the underlying DGP being GARCH(2, 1) and GARCH(1, 2) models. We also evaluate the performance of M-estimators when the underlying DGP is the GARCH(1, 1) model but it is misspecified as the GARCH(2, 1) model. This is essentially the case where the parameter is at the boundary.

3.5.1 Simulation for GARCH(2, 1) models

We consider four types of innovation distributions: the normal, DE, logistic, $t(3)$ and $t(2.2)$. There are $R = 1000$ replications being generated with the sample size $n = 1000$ and

$$\boldsymbol{\theta}_0 = (4.46 \times 10^{-6}, 0.0525, 0.108, 0.832)',$$

a choice motivated by the QMLE computed using the R package `fGarch` for the FTSE 100 data which will be analyzed later.

The standardized bias and MSE of the various M-estimators are reported in Table 3.4. It is worth noting that under the normal distribution, the bias and MSE of other M-estimators are generally close to those of the QMLE. However, for more heavy-tailed distributions, the QMLE produces larger bias and MSE compared with other M-estimators. Under the $t(3)$ and $t(2.2)$ distributions, which do not admit finite fourth moment, the advantage of the M-estimators over the QMLE becomes more prominent. For instance, for the estimation of α_1, the MSE ratio of the QMLE with respect to the μ-estimator is 6.0 under the $t(3)$ distribution, and it increases significantly to 23.3 under the $t(2.2)$ distribution. Also, under the $t(2.2)$ distribution, the LAD and Huber's estimators perform poorly compared with the μ- and Cauchy-estimators since the former two yield significantly larger MSE than the latter two. Consequently, these provide evidence for (i) the robustness of the M-estimators for heavy-tailed distributions is not at the cost of losing much efficiency under the normal distribution and (ii) the μ- and Cauchy-estimators are relatively less sensitive to heavy-tails among these M-estimators.

Table 3.4: The standardized bias and MSE of the M-estimators for GARCH(2, 1) models under various error distributions (sample size $n = 1000$; $R = 1000$ replications).

	Standardized bias				**Standardized MSE**			
	ω	α_1	α_2	β	ω	α_1	α_2	β
Normal								
QMLE	3.55×10^{-6}	1.88×10^{-3}	3.05×10^{-3}	-2.02×10^{-2}	2.18×10^{-11}	1.53×10^{-3}	2.08×10^{-3}	1.36×10^{-3}
LAD	3.35×10^{-6}	3.55×10^{-3}	1.80×10^{-4}	-1.76×10^{-2}	2.08×10^{-11}	1.74×10^{-3}	2.36×10^{-3}	1.32×10^{-3}
Huber	3.53×10^{-6}	5.54×10^{-3}	4.37×10^{-3}	-1.71×10^{-2}	2.16×10^{-11}	1.84×10^{-3}	2.53×10^{-3}	1.27×10^{-3}
μ-estimator	2.84×10^{-6}	2.48×10^{-3}	1.16×10^{-3}	-1.60×10^{-2}	1.91×10^{-11}	2.18×10^{-3}	3.06×10^{-3}	1.65×10^{-3}
Cauchy	2.66×10^{-6}	1.60×10^{-3}	1.57×10^{-3}	-1.55×10^{-2}	2.03×10^{-11}	2.51×10^{-3}	3.58×10^{-3}	1.94×10^{-3}
DE								
QMLE	2.51×10^{-6}	1.42×10^{-2}	-1.23×10^{-2}	-1.77×10^{-2}	1.49×10^{-11}	2.59×10^{-3}	2.59×10^{-3}	1.35×10^{-3}
LAD	1.74×10^{-6}	1.14×10^{-2}	-1.09×10^{-2}	-1.31×10^{-2}	6.60×10^{-12}	1.45×10^{-3}	1.84×10^{-3}	8.53×10^{-4}
Huber's	1.73×10^{-6}	1.21×10^{-2}	-1.21×10^{-2}	-1.28×10^{-2}	6.73×10^{-12}	1.49×10^{-3}	1.92×10^{-3}	8.93×10^{-4}
μ-estimator	1.44×10^{-6}	1.25×10^{-2}	-7.18×10^{-3}	-1.12×10^{-2}	5.64×10^{-12}	1.80×10^{-3}	2.46×10^{-3}	8.97×10^{-4}
Cauchy	1.37×10^{-6}	1.36×10^{-2}	-5.67×10^{-3}	-1.12×10^{-2}	6.61×10^{-12}	2.43×10^{-3}	3.28×10^{-3}	1.03×10^{-3}
Logistic								
QMLE	3.83×10^{-6}	1.38×10^{-2}	-1.73×10^{-2}	-1.75×10^{-2}	2.64×10^{-11}	3.78×10^{-3}	3.01×10^{-3}	1.57×10^{-3}
LAD	2.97×10^{-6}	8.27×10^{-3}	-1.43×10^{-2}	-1.20×10^{-2}	1.55×10^{-11}	2.01×10^{-3}	2.16×10^{-3}	1.11×10^{-3}
Huber's	3.03×10^{-6}	8.42×10^{-3}	-1.23×10^{-2}	-1.25×10^{-2}	1.64×10^{-11}	2.01×10^{-3}	2.03×10^{-3}	1.12×10^{-3}
μ-estimator	2.50×10^{-6}	6.28×10^{-3}	-1.25×10^{-2}	-8.64×10^{-3}	1.33×10^{-11}	2.19×10^{-3}	2.98×10^{-3}	1.23×10^{-3}
Cauchy	2.41×10^{-6}	6.46×10^{-3}	-1.10×10^{-2}	-8.62×10^{-3}	1.42×10^{-11}	2.50×10^{-3}	3.49×10^{-3}	1.46×10^{-3}
$\boldsymbol{t(3)}$								
QMLE	1.67×10^{-6}	2.89×10^{-2}	-2.20×10^{-2}	-3.48×10^{-2}	2.74×10^{-11}	1.37×10^{-2}	1.56×10^{-2}	8.02×10^{-3}
LAD	1.00×10^{-6}	7.28×10^{-3}	-6.13×10^{-3}	-1.04×10^{-2}	5.62×10^{-12}	3.01×10^{-3}	4.58×10^{-3}	2.02×10^{-3}
Huber's	9.74×10^{-7}	8.20×10^{-3}	-8.00×10^{-3}	-1.05×10^{-2}	5.50×10^{-12}	2.99×10^{-3}	4.53×10^{-3}	2.01×10^{-3}
μ-estimator	6.62×10^{-7}	8.42×10^{-3}	-8.91×10^{-3}	-5.33×10^{-3}	3.93×10^{-12}	2.30×10^{-3}	3.59×10^{-3}	1.63×10^{-3}
Cauchy	5.89×10^{-7}	9.44×10^{-3}	-9.33×10^{-3}	-5.20×10^{-3}	4.33×10^{-12}	2.51×10^{-3}	3.91×10^{-3}	1.85×10^{-3}
$\boldsymbol{t(2.2)}$								
QMLE	-4.35×10^{-7}	9.90×10^{-2}	-4.39×10^{-2}	-1.54×10^{-1}	1.90×10^{-11}	1.34×10^{-1}	1.48×10^{-1}	8.10×10^{-2}
LAD	1.13×10^{-6}	3.16×10^{-2}	-8.87×10^{-5}	-3.48×10^{-2}	1.35×10^{-11}	3.30×10^{-2}	4.54×10^{-2}	1.38×10^{-2}
Huber	1.38×10^{-6}	5.30×10^{-2}	-1.08×10^{-2}	-4.40×10^{-2}	1.53×10^{-11}	4.43×10^{-2}	5.52×10^{-2}	1.58×10^{-2}
μ-estimator	4.55×10^{-7}	1.60×10^{-2}	-4.41×10^{-3}	-1.30×10^{-2}	5.51×10^{-12}	5.75×10^{-3}	9.33×10^{-3}	5.38×10^{-3}
Cauchy	4.69×10^{-7}	2.04×10^{-2}	-5.37×10^{-3}	-1.47×10^{-2}	6.74×10^{-12}	6.13×10^{-3}	1.06×10^{-2}	6.52×10^{-3}

3.5.2 Simulation under a misspecified GARCH model

It is important to check whether the M-estimators are consistent when a GARCH model is misspecified with a higher order as over-fitting can occur in practice. In this case, we are essentially fitting a GARCH model with some component(s) of the parameter at the boundary equal to zero. We simulate below data from the GARCH(1, 1) model under various error distributions; however, the data are fitted by the GARCH(2, 1) model. In simulation, we use $R = 1000$, $n = 1000$ and $\boldsymbol{\theta}_0 = (1.65 \times 10^{-5}, 0.0701, 0.901)'$, which is motivated by the QMLE obtained by using the `fGarch` package for the SSE data analyzed later in Section 3.7.

The standardized bias and MSE of the M-estimators are shown in Table 3.5. For all distributions considered, the bias are close to zero and the MSE are small indicating good performance of the M-estimators under this type of mis-specification. Similar to the results in Table 3.4, the QMLE is sensitive to the heavy-tailed distributions while other M-estimators are more robust as they yield smaller MSE values under the $t(3)$ distribution. For example, for the estimation of α_2, the MSE ratio of the QMLE with respect to the μ-estimator is 9.4 under the $t(3)$ distribution.

3.5.3 Simulation for the GARCH(1, 2) models

Since we did not come across a real data that can be fitted by the GARCH(1, 2) model, we resort to simulation results to study the performance of M-estimators for such models. We choose $\boldsymbol{\theta}_0 = (0.1, 0.1, 0.2, 0.6)'$, $R = 1000$ replications and $n = 1000$. The standardized bias and MSE of the M-estimators under various error distributions are reported in Table 3.6. We do not report results for the QMLE when data are generated under the $t(3)$ and $t(2.2)$ error distributions since the algorithm for computing the QMLE did not converge for most replications. Under the normal error distribution, the LAD and Huber's estimators produce MSE that is close to the QMLE while the μ- and Cauchy-estimators yield larger MSE corresponding for estimating ω and α. For the DE and logistic distributions, there is no significant difference between these estimators. Their difference becomes

Table 3.5: The standardized bias and MSE of the M-estimators under the misspecified model (sample size $n = 1000$; $R = 1000$ replications); the underlying DGP is the GARCH(1, 1) model whereas the model is misspecified as a GARCH(2, 1).

	Standardized bias				**Standardized MSE**			
	ω	α_1	α_2	β	ω	α_1	α_2	β
Normal								
QMLE	1.11×10^{-5}	-2.00×10^{-3}	5.97×10^{-3}	-2.38×10^{-2}	3.94×10^{-10}	1.55×10^{-3}	1.87×10^{-3}	2.64×10^{-3}
LAD	1.09×10^{-5}	-1.73×10^{-3}	5.65×10^{-3}	-2.43×10^{-2}	4.53×10^{-10}	1.73×10^{-3}	2.12×10^{-3}	3.09×10^{-3}
Huber's	1.22×10^{-5}	1.25×10^{-3}	6.08×10^{-3}	-2.43×10^{-2}	5.18×10^{-10}	1.82×10^{-3}	2.28×10^{-3}	3.13×10^{-3}
μ-estimator	1.11×10^{-5}	-5.36×10^{-4}	5.75×10^{-3}	-2.49×10^{-2}	5.27×10^{-10}	2.42×10^{-3}	2.99×10^{-3}	3.67×10^{-3}
Cauchy	1.13×10^{-5}	-5.85×10^{-4}	6.31×10^{-3}	-2.61×10^{-2}	6.26×10^{-10}	2.83×10^{-3}	3.57×10^{-3}	4.41×10^{-3}
DE								
QMLE	9.70×10^{-6}	-1.07×10^{-3}	7.12×10^{-3}	-2.45×10^{-2}	4.19×10^{-10}	2.82×10^{-3}	3.33×10^{-3}	3.78×10^{-3}
LAD	8.11×10^{-6}	6.07×10^{-4}	4.72×10^{-3}	-1.89×10^{-2}	2.91×10^{-10}	2.24×10^{-3}	2.60×10^{-3}	2.51×10^{-3}
Huber's	7.84×10^{-6}	-7.00×10^{-4}	4.79×10^{-3}	-1.94×10^{-2}	2.92×10^{-10}	2.20×10^{-3}	2.54×10^{-3}	2.58×10^{-3}
μ-estimator	7.21×10^{-6}	2.45×10^{-3}	3.15×10^{-3}	-1.69×10^{-2}	2.85×10^{-10}	2.59×10^{-3}	3.02×10^{-3}	2.59×10^{-3}
Cauchy	7.49×10^{-6}	3.86×10^{-3}	3.29×10^{-3}	-1.79×10^{-2}	3.48×10^{-10}	3.10×10^{-3}	3.65×10^{-3}	3.20×10^{-3}
Logistic								
QMLE	1.24×10^{-5}	-1.95×10^{-3}	9.70×10^{-3}	-2.68×10^{-2}	5.24×10^{-10}	2.14×10^{-3}	2.61×10^{-3}	3.28×10^{-3}
LAD	1.03×10^{-5}	-2.81×10^{-3}	8.40×10^{-3}	-2.30×10^{-2}	3.88×10^{-10}	1.82×10^{-3}	2.23×10^{-3}	2.63×10^{-3}
Huber's	1.00×10^{-5}	-3.27×10^{-3}	8.11×10^{-3}	-2.28×10^{-2}	3.83×10^{-10}	1.78×10^{-3}	2.14×10^{-3}	2.62×10^{-3}
μ-estimator	9.47×10^{-6}	-2.29×10^{-3}	8.31×10^{-3}	-2.21×10^{-2}	3.88×10^{-10}	2.15×10^{-3}	2.69×10^{-3}	2.86×10^{-3}
Cauchy	9.74×10^{-6}	-8.90×10^{-4}	8.56×10^{-3}	-2.26×10^{-2}	4.34×10^{-10}	2.53×10^{-3}	3.21×10^{-3}	3.23×10^{-3}
***t*(3)**								
QMLE	1.08×10^{-5}	1.64×10^{-2}	1.14×10^{-2}	-5.47×10^{-2}	1.15×10^{-9}	1.93×10^{-2}	2.67×10^{-2}	1.97×10^{-2}
LAD	4.50×10^{-6}	1.05×10^{-3}	2.96×10^{-3}	-2.08×10^{-2}	1.85×10^{-10}	3.01×10^{-3}	3.41×10^{-3}	3.39×10^{-3}
Huber's	5.46×10^{-6}	4.83×10^{-3}	2.64×10^{-3}	-2.03×10^{-2}	2.19×10^{-10}	3.33×10^{-3}	3.80×10^{-3}	3.50×10^{-3}
μ-estimator	4.47×10^{-6}	5.91×10^{-3}	4.41×10^{-4}	-1.51×10^{-2}	1.45×10^{-10}	2.55×10^{-3}	2.84×10^{-3}	2.25×10^{-3}
Cauchy	3.65×10^{-6}	3.85×10^{-3}	4.77×10^{-5}	-1.54×10^{-2}	1.45×10^{-10}	2.51×10^{-3}	2.86×10^{-3}	2.56×10^{-3}

clearer under heavy-tailed distributions: the μ- and Cauchy-estimators produce smaller MSE of ω under the $t(3)$ distribution and smaller MSE of α under the $t(2.2)$ distribution than the LAD and Huber's estimators.

Table 3.6: The standardized bias and MSE of the M-estimators for GARCH(1, 2) models under various error distributions (sample size $n = 1000$; $R = 1000$ replications).

	Standardized bias				**Standardized MSE**			
	ω	α	β_1	β_2	ω	α	β_1	β_2
Normal								
QMLE	5.53×10^{-2}	1.10×10^{-3}	9.65×10^{-2}	-1.52×10^{-1}	2.66×10^{-2}	1.17×10^{-3}	1.45×10^{-1}	1.38×10^{-1}
LAD	5.93×10^{-2}	7.15×10^{-4}	9.01×10^{-2}	-1.50×10^{-1}	3.21×10^{-2}	1.31×10^{-3}	1.55×10^{-1}	1.45×10^{-1}
Huber	6.49×10^{-2}	4.64×10^{-3}	9.77×10^{-2}	-1.57×10^{-1}	3.72×10^{-2}	1.37×10^{-3}	1.56×10^{-1}	1.47×10^{-1}
μ-estimator	7.45×10^{-2}	8.93×10^{-4}	1.11×10^{-1}	-1.86×10^{-1}	7.41×10^{-2}	1.84×10^{-3}	2.16×10^{-1}	2.01×10^{-1}
Cauchy	7.51×10^{-2}	1.25×10^{-3}	1.29×10^{-1}	-2.06×10^{-1}	6.30×10^{-2}	2.17×10^{-3}	2.43×10^{-1}	2.31×10^{-1}
DE								
QMLE	5.48×10^{-2}	2.93×10^{-3}	1.01×10^{-1}	-1.63×10^{-1}	3.15×10^{-2}	1.79×10^{-3}	1.62×10^{-1}	1.57×10^{-1}
LAD	3.73×10^{-2}	-1.93×10^{-3}	8.76×10^{-2}	-1.27×10^{-1}	1.20×10^{-2}	1.61×10^{-3}	1.46×10^{-1}	1.35×10^{-1}
Huber	3.83×10^{-2}	-1.22×10^{-3}	9.51×10^{-2}	-1.36×10^{-1}	1.21×10^{-2}	1.65×10^{-3}	1.53×10^{-1}	1.44×10^{-1}
μ-estimator	4.05×10^{-2}	1.15×10^{-3}	1.13×10^{-1}	-1.52×10^{-1}	1.72×10^{-2}	2.05×10^{-3}	1.73×10^{-1}	1.60×10^{-1}
Cauchy	4.74×10^{-2}	3.26×10^{-3}	1.18×10^{-1}	-1.66×10^{-1}	2.55×10^{-2}	2.48×10^{-3}	1.85×10^{-1}	1.72×10^{-1}
Logistic								
QMLE	5.77×10^{-2}	2.76×10^{-3}	1.06×10^{-1}	-1.61×10^{-1}	3.02×10^{-2}	1.49×10^{-3}	1.67×10^{-1}	1.59×10^{-1}
LAD	4.50×10^{-2}	-5.78×10^{-5}	7.27×10^{-2}	-1.18×10^{-1}	1.58×10^{-2}	1.37×10^{-3}	1.30×10^{-1}	1.18×10^{-1}
Huber	4.50×10^{-2}	-2.33×10^{-4}	8.85×10^{-2}	-1.34×10^{-1}	1.58×10^{-2}	1.36×10^{-3}	1.53×10^{-1}	1.39×10^{-1}
μ-estimator	4.52×10^{-2}	1.32×10^{-3}	9.39×10^{-2}	-1.40×10^{-1}	1.80×10^{-2}	1.72×10^{-3}	1.58×10^{-1}	1.44×10^{-1}
Cauchy	5.15×10^{-2}	2.91×10^{-3}	1.05×10^{-1}	-1.57×10^{-1}	2.98×10^{-2}	2.08×10^{-3}	1.85×10^{-1}	1.70×10^{-1}
***t*(3)**								
QMLE	-	-	-	-	-	-	-	-
LAD	2.93×10^{-2}	2.43×10^{-3}	1.08×10^{-1}	-1.40×10^{-1}	1.13×10^{-2}	2.49×10^{-3}	1.82×10^{-1}	1.59×10^{-1}
Huber	2.87×10^{-2}	1.50×10^{-3}	9.13×10^{-2}	-1.26×10^{-1}	1.18×10^{-2}	2.30×10^{-3}	1.60×10^{-1}	1.40×10^{-1}
μ-estimator	1.57×10^{-2}	8.75×10^{-5}	1.21×10^{-1}	-1.37×10^{-1}	5.59×10^{-3}	1.88×10^{-3}	1.63×10^{-1}	1.42×10^{-1}
Cauchy	1.50×10^{-2}	6.44×10^{-4}	1.38×10^{-1}	-1.54×10^{-1}	6.50×10^{-3}	2.15×10^{-3}	1.90×10^{-1}	1.65×10^{-1}
***t*(2.2)**								
QMLE	-	-	-	-	-	-	-	-
LAD	3.53×10^{-2}	2.57×10^{-2}	1.24×10^{-1}	-1.85×10^{-1}	1.30×10^{-2}	1.41×10^{-2}	2.41×10^{-1}	2.21×10^{-1}
Huber	4.86×10^{-2}	3.99×10^{-2}	7.81×10^{-2}	-1.66×10^{-1}	1.44×10^{-2}	1.63×10^{-2}	1.81×10^{-1}	1.79×10^{-1}
μ-estimator	1.72×10^{-2}	5.18×10^{-3}	1.51×10^{-1}	-1.78×10^{-1}	1.73×10^{-2}	4.27×10^{-3}	2.42×10^{-1}	2.12×10^{-1}
Cauchy	2.15×10^{-2}	9.68×10^{-3}	1.50×10^{-1}	-1.85×10^{-1}	2.05×10^{-2}	4.90×10^{-3}	2.34×10^{-1}	2.14×10^{-1}

3.6 Simulating the bootstrap distributions

To evaluate the finite sample performance of the bootstrap approximation, here we compare the bootstrap coverage rates with the nominal levels. In particular, we generate $R = 500$ data of sample size $n = 1000$ from the GARCH$(1, 1)$ model with parameter $\boldsymbol{\theta}_0 = (0.1, 0.1, 0.8)'$ under both the normal and $t(3)$ error distributions. For each data, we compute $B = 2000$ bootstrap estimates using the bootstrap

schemes M, E and U introduced in Section 3.3 and construct the bootstrap and asymptotic confidence intervals (CI) using (3.3.4) and (3.3.5), respectively. The coverage rates are computed as the proportions of the CI's that cover the true parameter. We report the coverage rates (in percentage) for the 90% and 95% nominal levels in Table 3.7.

Under the normal distribution, the coverage rates of the bootstrap approximation are generally close to the nominal levels. Also, the bootstrap approximation works better for the QMLE, LAD and Huber's estimators than the μ-estimator and Cauchy-estimator. However, under the $t(3)$ distribution, the bootstrap approximation works poorly for the QMLE while the coverage rates are reasonably well for other M-estimators. For both distributions, scheme U outperforms schemes M and E. In terms of the asymptotic approximation, it works well only for few cases and is outperformed by the bootstrap coverage rates for most cases and this indicates the usefulness of the bootstrap approximation.

3.7 Real data analysis

In this section, we analyse daily log-returns of three financial time series: (i) the Shanghai Stock Exchange (SSE) Index from January 2007 to December 2009 with $n = 752$; (ii) the Electric Fuel Corporation (EFCX) data from January 2000 to December 2001 with $n = 498$; (iii) the FTSE 100 Index data from January 2007 to December 2009 with $n = 783$. Based on exploratory data analysis, GARCH(1, 1) model fits well to the SSE and EFCX data. However, we fitted GARCH(2, 1) model to the FTSE 100 data for two reasons. First, when fitted by the GARCH(2, 1) model with `fGarch` package in R, α_2 is significant with p-value equal to 0.019; second, the Akaike information criterion (AIC) of the GARCH(2, 1) model is smaller than that of the GARCH(1, 1) model.

Table 3.7: The coverage rates (in percentage) of the bootstrap schemes M, E and U and asymptotic normal approximations for the M-estimators QMLE, LAD, Huber's, μ- and Cauchy-; the error distributions are normal and $t(3)$.

			90% nominal level			95% nominal level		
			ω	α	β	ω	α	β
Normal	QMLE	Scheme M	89.0	86.2	88.2	91.0	92.2	91.4
		Scheme E	87.2	83.8	86.8	90.2	88.4	91.2
		Scheme U	90.2	87.4	87.2	94.4	92.6	93.2
		Asymptotic	82.6	91.0	85.8	87.0	95.2	89.0
Normal	LAD	Scheme M	86.0	83.4	84.2	88.2	87.2	88.4
		Scheme E	88.0	87.2	87.2	91.0	91.2	90.2
		Scheme U	88.6	88.4	88.0	93.2	91.8	91.8
		Asymptotic	94.0	98.8	87.0	96.4	99.4	90.4
Normal	Huber's	Scheme M	88.8	85.4	86.6	91.2	89.8	91.2
		Scheme E	88.2	89.0	88.0	91.4	92.4	90.0
		Scheme U	89.6	90.4	88.4	93.6	93.6	91.8
		Asymptotic	87.6	95.4	86.2	90.6	96.6	90.4
Normal	μ-estimator	Scheme M	88.0	84.6	86.8	89.6	87.8	88.6
		Scheme E	87.4	84.8	86.6	89.4	88.4	88.4
		Scheme U	88.6	88.4	87.6	91.8	91.8	90.6
		Asymptotic	71.4	69.6	86.8	77.4	78.2	90.8
Normal	Cauchy	Scheme M	85.6	84.0	84.4	87.8	85.8	87.6
		Scheme E	81.4	82.2	80.2	82.8	86.2	84.2
		Scheme U	88.4	88.2	87.0	90.4	91.4	89.4
		Asymptotic	97.8	99.8	85.0	98.2	100.0	89.6
$t(3)$	QMLE	Scheme M	71.0	75.4	74.8	75.0	79.0	78.0
		Scheme E	67.6	72.4	66.8	73.4	76.2	72.4
		Scheme U	75.6	84.6	75.0	81.6	87.2	80.0
		Asymptotic	-	-	-	-	-	-
$t(3)$	LAD	Scheme M	84.4	80.6	83.0	85.4	83.8	87.8
		Scheme E	84.6	85.0	81.4	87.6	87.0	86.6
		Scheme U	81.6	86.2	79.2	87.4	89.2	84.8
		Asymptotic	98.0	99.8	88.8	99.6	100.0	91.2
$t(3)$	Huber's	Scheme M	83.0	80.6	81.8	85.6	83.2	86.6
		Scheme E	81.8	79.2	80.8	85.8	81.6	85.8
		Scheme U	86.2	88.0	86.0	90.2	91.4	90.2
		Asymptotic	96.8	99.0	88.4	97.8	99.6	92.8
$t(3)$	μ-estimator	Scheme M	82.4	84.8	83.8	86.2	88.4	88.2
		Scheme E	84.6	84.0	84.6	87.4	88.0	88.8
		Scheme U	82.6	83.6	80.4	88.8	88.2	86.4
		Asymptotic	86.6	91.8	80.8	90.6	95.6	86.4
$t(3)$	Cauchy	Scheme M	78.2	83.4	78.4	81.8	86.2	82.0
		Scheme E	83.4	85.6	82.6	85.4	89.0	87.2
		Scheme U	85.0	85.0	84.8	90.0	88.6	89.2
		Asymptotic	100.0	100.0	85.6	100.0	100.0	90.8

3.7.1 The SSE data and bootstrap estimates of the bias and MSE

Table 3.8 displays the QMLE computed using the R package `fGarch` and the QMLE and LAD estimates computed using the algorithm (3.4.1). The QMLEs given by `fGarch` and (3.4.1) are close. Also, the QMLE and LAD estimates of β are close.

Table 3.8: The M-estimates (QMLE and LAD) of the GARCH(1, 1) model for the SSE data; The QMLEs are obtained by using `fGarch` and (3.4.1).

	ω	α	β
`fGarch`	1.65×10^{-5}	7.01×10^{-2}	0.90
QMLE	2.88×10^{-5}	7.97×10^{-2}	0.87
LAD	1.62×10^{-5}	4.99×10^{-2}	0.86

To estimate the bias and MSE of the M-estimators of the GARCH(1, 1) parameters of the underlying DGP of the SSE data, notice that we know neither the underlying true parameters nor the error distribution. Moreover, a M-estimator based on H is consistent for the true parameter if and only if $c_H = 1$. This holds, in particular, when the QMLE is used if the underlying error distribution has unit variance. Hence to estimate population bias and MSE using simulation, we use population parameter as the one estimated from the SSE data using the QMLE computed from `fGarch`. Then we consider the DGP from GARCH$(1, 1)$ models with four possible error distributions, namely, the normal, DE, logistic and $t(3)$ distributions and for each scenario generate R replications of n observations. We estimate the *normalized bias* and *normalized MSE* of $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0H})$ by

$$(R^{-1}\sum_{r=1}^{R} n^{1/2}\{\hat{\omega}_r - c_H\omega_0\}, R^{-1}\sum_{r=1}^{R} n^{1/2}\{\hat{\alpha}_r - c_H\alpha_0\}, R^{-1}\sum_{r=1}^{R} n^{1/2}\{\hat{\beta}_r - \beta_0\})' \quad (3.7.1)$$

$$(R^{-1}\sum_{r=1}^{R}\{n^{1/2}(\hat{\omega}_r - c_H\omega_0)\}^2, R^{-1}\sum_{r=1}^{R}\{n^{1/2}(\hat{\alpha}_r - c_H\alpha_0)\}^2, R^{-1}\sum_{r=1}^{R}\{n^{1/2}(\hat{\beta}_r - \beta_0)\}^2)'. \quad (3.7.2)$$

Notice that the normalized bias and MSE are different from the standardized bias and MSE by some simple multiplicative factors involving c_H. The normalized bias and MSE are used to evaluate performance of the bootstrap approximation, while the standardized bias and MSE are used to compare performance of different M-estimators.

To obtain the bootstrap estimates of the normalized bias and MSE, we use three different bootstrap schemes to generate weights $\{w_{nt}; 1 \leq t \leq n\}$ B number of times. We compute the bootstrap estimates $\{\hat{\boldsymbol{\theta}}_{*nb}, 1 \leq b \leq B\}$ using (3.4.2) and consequently B bootstrap replicates (realizations)

$$\{\sigma_n^{-1} n^{1/2} (\hat{\boldsymbol{\theta}}_{*nb} - \hat{\boldsymbol{\theta}}_n); 1 \leq b \leq B\}$$

of the bootstrap distribution where $\hat{\boldsymbol{\theta}}_n$ is the M-estimate of the dataset computed using (3.4.1) based on the score function under consideration. The effect of the bootstrap scheme is reflected in the standardization through σ_n. The bootstrap estimates of the normalized bias and MSE are computed by

$$\text{Bias} = (1/B) \sum_{b=1}^{B} \{\sigma_n^{-1} n^{1/2} (\hat{\boldsymbol{\theta}}_{*nb} - \hat{\boldsymbol{\theta}}_n)\} \quad \text{and} \quad \text{MSE} = (1/B) \sum_{b=1}^{B} \{\sigma_n^{-1} n^{1/2} (\hat{\boldsymbol{\theta}}_{*nb} - \hat{\boldsymbol{\theta}}_n)\}^2. \tag{3.7.3}$$

Here the squares of vectors in the MSE above should be interpreted as entry-wise square.

Using (3.7.1) and (3.7.2) with $n = 752$ and $R = 500$, estimates of the normalized bias and MSE for the QMLE and LAD under various error distributions are shown in Table 3.9 and Table 3.10 respectively. Also, to evaluate the bootstrap approximation, we include the bootstrap estimates of the normalized bias and MSE in these tables. Note that for the LAD, all these bootstrap schemes have good approximation to the bias and MSE as they are generally of the same magnitude regardless of the underlying error distribution. For the QMLE under the DE, logistic and normal distributions, except for the bias of ω, scheme M provides good approximation while schemes E and U tend to underestimate the bias.

Table 3.9: The normalized bias and MSE of the QMLE and their bootstrap estimates for the SSE data.

	Normalized bias			**Normalized MSE**		
	ω	α	β	ω	α	β
Error Dist.						
DE	3.80×10^{-4}	0.15	-0.86	5.02×10^{-7}	0.70	3.51
Logisitic	4.39×10^{-4}	0.22	-0.90	5.90×10^{-7}	0.65	3.22
Normal	3.75×10^{-4}	0.12	-0.78	4.14×10^{-7}	0.48	2.59
$t(3)$	2.58×10^{-4}	0.49	-1.17	6.56×10^{-7}	6.60	11.70
Bootstrap						
Scheme M	4.68×10^{-5}	0.10	-0.18	7.58×10^{-7}	0.83	3.99
Scheme E	8.39×10^{-6}	5.47×10^{-2}	-5.96×10^{-2}	3.06×10^{-7}	0.68	2.35
Scheme U	4.90×10^{-6}	1.58×10^{-2}	-3.32×10^{-2}	1.23×10^{-7}	0.72	1.41

Table 3.10: The normalized bias and MSE of the LAD and their bootstrap estimates for the SSE data.

	Normalized bias			**Normalized MSE**		
	ω	α	β	ω	α	β
Error Dist.						
DE	1.45×10^{-4}	4.54×10^{-2}	-0.65	7.01×10^{-8}	0.12	2.24
Logisitic	2.33×10^{-4}	9.88×10^{-2}	-0.85	1.82×10^{-7}	0.20	3.05
Normal	2.28×10^{-4}	6.75×10^{-2}	-0.75	1.69×10^{-7}	0.21	2.63
$t(3)$	6.06×10^{-5}	6.12×10^{-2}	-0.48	2.82×10^{-8}	0.12	2.25
Bootstrap						
Scheme M	2.44×10^{-5}	6.12×10^{-2}	-0.17	7.79×10^{-8}	0.28	2.19
Scheme E	6.77×10^{-5}	0.11	-0.39	6.66×10^{-8}	0.25	1.92
Scheme U	1.29×10^{-5}	1.88×10^{-2}	-0.11	3.30×10^{-8}	0.22	1.14

Table 3.11: The M-estimates (QMLE, LAD, Huber's, μ- and Cauchy-) of the GARCH(2, 1) model using the FTSE 100 data; the QMLEs are obtained by using `fGarch` and (3.4.1).

	`fGarch`	QMLE	LAD	Huber's	μ-estimator	Cauchy
ω	4.46×10^{-6}	4.65×10^{-6}	3.13×10^{-6}	3.55×10^{-6}	1.02×10^{-5}	2.51×10^{-6}
α_1	5.25×10^{-2}	4.51×10^{-2}	2.46×10^{-2}	3.45×10^{-2}	4.95×10^{-2}	6.83×10^{-3}
α_2	0.11	9.00×10^{-2}	5.57×10^{-2}	6.42×10^{-2}	0.17	4.18×10^{-2}
β	0.83	0.85	0.84	0.86	0.81	0.80

3.7.2 The FTSE 100 data and the GARCH(2, 1) model

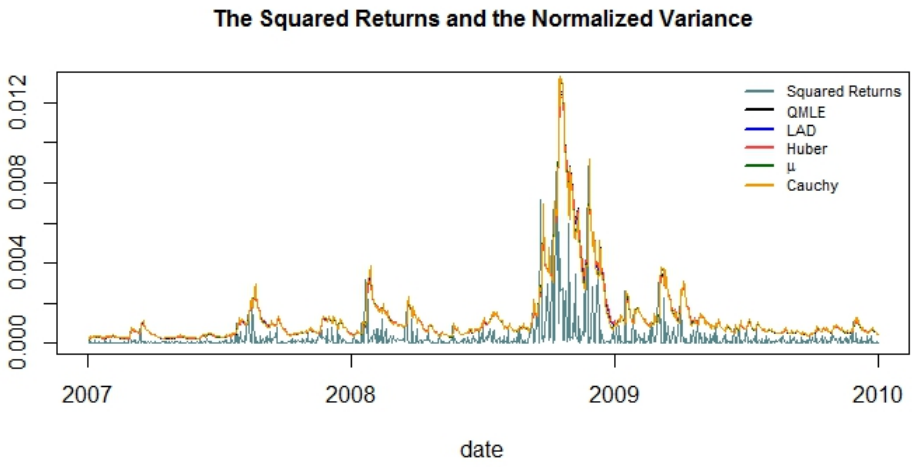
Here we fit the GARCH(2, 1) model with the FTSE 100 data. The estimates given by `fGarch` and by our M-estimators are shown in Table 3.11. The QMLE (based on algorithm (3.4.1)) and `fGarch` provide similar estimates for all components of the parameter. Also, the M-estimates of β do not vary much. For ω, α_1 and α_2, the M-estimates are quite different since c_H in (3.2.10) depends on the score function H used for the estimation.

For a GARCH(p, q) model, using (3.2.6) and the formulas for $\{c_j(\boldsymbol{\theta}); j \geq 0\}$ in Berkes et al. (2003) (Section 3), we have $\hat{v}_t(\boldsymbol{\theta}_{0H}) = c_H \hat{v}_t(\boldsymbol{\theta}_0)$. Since a M-estimator $\hat{\boldsymbol{\theta}}_n$ estimates $\boldsymbol{\theta}_{0H}$, $\hat{v}_t(\hat{\boldsymbol{\theta}}_n)$ estimates $c_H \hat{v}_t(\boldsymbol{\theta}_0)$ which is a scale-transformed estimate of the conditional variance. To examine the behavior of the market volatility after eliminating the effect of any particular M-estimator used, we define the following normalized volatility by

$$\hat{u}_t(\hat{\boldsymbol{\theta}}_n) := \hat{v}_t(\hat{\boldsymbol{\theta}}_n) / \sum_{i=1}^{n} \hat{v}_i(\hat{\boldsymbol{\theta}}_n);\ 1 \leq t \leq n. \tag{3.7.4}$$

Figure 3.1 shows the plot of $\{\hat{u}_t(\hat{\boldsymbol{\theta}}_n); 1 \leq t \leq n\}$ based on various M-estimators against the squared returns. Notice that although the M-estimates in Table 3.11 are different, the plot of the normalized volatilities almost overlap each other based on all M-estimators. Also, large values of the normalized volatilities and large squared returns occur at the same time. In this sense, the volatilities are well-modelled by using these M-estimators.

Figure 3.1: The plot of the squared returns and the estimated normalized conditional variances using various M-estimators for the FTSE 100 data



3.7.3 The EFCX data

Muler and Yohai (2008) fitted the GARCH(1, 1) model to the EFCX data and noted that the QMLE and LAD estimates of the parameter β are significantly different. Here in Table 3.12, we report estimates given by the `fGarch` and M-estimators. Note that in our previous analysis of the SSE and FTSE 100 data, `fGarch` estimates and our QMLE are quite close while their difference is much more significant for this data. It is also worth noting that while the LAD, Huber's, $\mu-$ and Cauchy-estimates of β are close to each other, they are all quite different from the corresponding estimate 0.84 of the QMLE when viewed as a M-estimate. We explain below that such interesting behavior might be related to the infinite fourth moment of the underlying innovation distribution.

Table 3.12: The M-estimates (QMLE, LAD, Huber's, μ- and Cauchy-) of the GARCH(1, 1) model for the EFCX data; the QMLEs are obtained by using `fGarch` and (3.4.1).

	`fGarch`	QMLE	LAD	Huber's	μ-estimator	Cauchy
ω	1.89×10^{-4}	6.28×10^{-4}	6.43×10^{-4}	8.37×10^{-4}	1.42×10^{-3}	2.97×10^{-4}
α	4.54×10^{-2}	7.20×10^{-2}	8.87×10^{-2}	0.10	0.27	6.35×10^{-2}
β	0.92	0.84	0.66	0.67	0.61	0.60

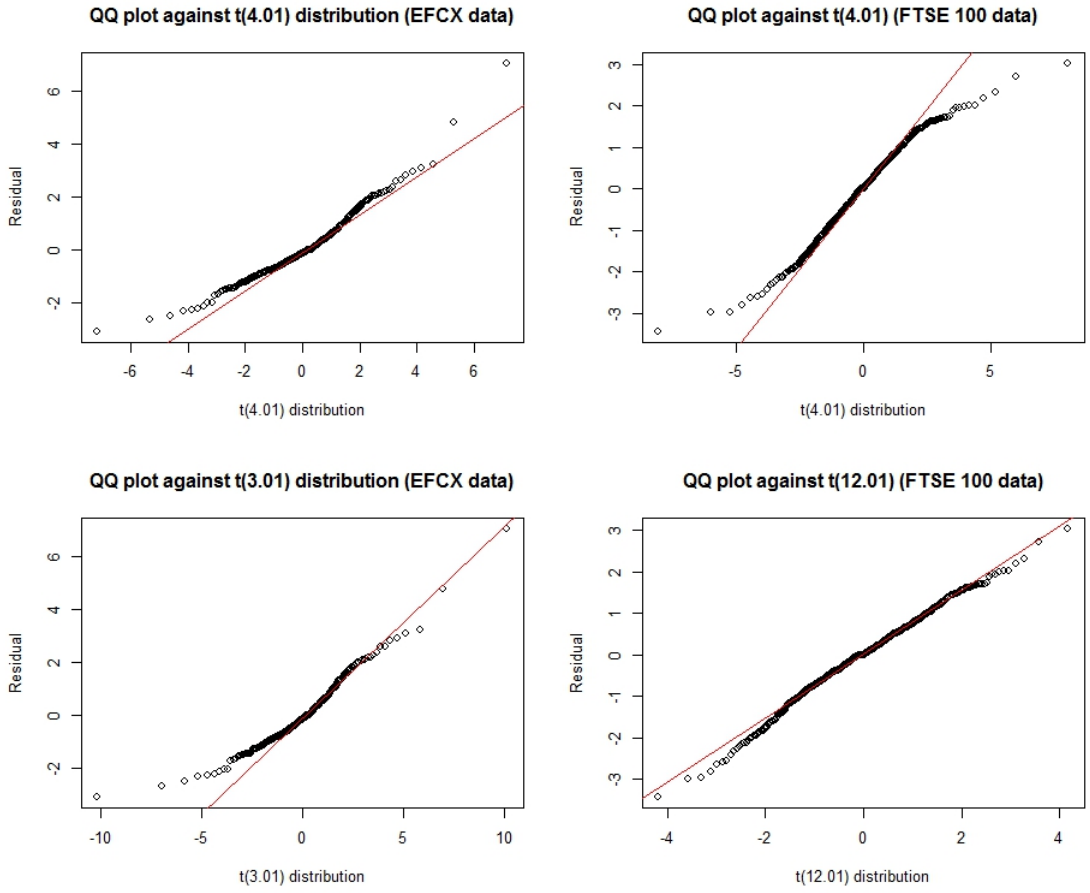
To examine whether the innovation distribution has finite fourth moment, we

use the QQ-plots of the residuals $\{X_t/\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_n); 1 \leq t \leq n\}$ based on the μ-estimator $\hat{\boldsymbol{\theta}}_n$ against the $t(d)$ distributions for various degrees of freedom d. We consider μ-estimator since it imposes mild moment assumption on the innovation distribution. The main idea behind the QQ-plots of the residuals against the $t(d)$ distribution is simple. Recall that if $\epsilon \sim t(d)$ distribution then $\mathrm{E}|\epsilon|^{\nu} < \infty$ if and only if $\nu < d$. Therefore, residuals with heavier tail than the $t(d)$ distribution correspond to the errors with the infinite d-th moment while those with lighter tail than the $t(d)$ distribution have the finite d-th error moment.

The top-left panel of Figure 3.2 shows the QQ-plot of the residuals against the $t(4.01)$ distribution for the EFCX data. The residuals have heavier right tail than the $t(4.01)$ distribution which implies that the fourth moment of the error term may not exist. On the other hand, the QQ-plot against the $t(3.01)$ distribution reveals lighter tail as shown at the bottom-left panel of Figure 3.2 and this implies that $\mathrm{E}|\epsilon|^3 < \infty$.

For the FTSE 100 data, the QQ-plot against the $t(4.01)$ distribution at the top-right panel of Figure 3.2 shows that the residuals have lighter tails than the $t(4.01)$ distribution. For the QQ-plot against the $t(12.01)$ distribution, as shown at the bottom-right panel of Figure 3.2, residuals fit the distribution better. Therefore, we may conclude that $\mathrm{E}|\epsilon|^4 < \infty$ holds for the FTSE 100 data and this explains why the other M-estimates of β in Table 3.11 are close.

Figure 3.2: The QQ-plot of the residuals against t distributions for the EFCX (left column) and FTSE 100 (right column) data.



3.8 Conclusion

We consider a class of M-estimators and the weighted bootstrap approximation of their distributions for the GARCH models. An iteratively re-weighted algorithm for computing the M-estimators and their bootstrap replicates are implemented. Both simulation and real data analysis demonstrate superior performance of the M-estimators for the GARCH(1, 1), GARCH(2, 1) and GARCH(1, 2) models. Under heavy-tailed error distributions, we show that the M-estimators are more robust than the routinely-applied QMLE. We also demonstrate through simulations that the M-estimators work well when the true GARCH(1, 1) model is misspecified as the GARCH(2, 1) model. Simulation results indicate that under the finite sample size, bootstrap approximation is better than the asymptotic normal approximation

of the M-estimators.

Chapter 4

R-estimation in GARCH models; asymptotics, applications and bootstrapping

4.1 Introduction

Estimation of parameters based on ranks of the residuals was discussed by Koul and Ossiander (1994) for the homoscedastic autoregressive model and Mukherjee (2007) for the heterscedastic models. For the GARCH model defined by (3.1.1) and (3.1.2), Andrews (2012) proposed a class of R-estimators using a log-transformation of the squared observations and then minimizing a rank-based residual dispersion function. However, our R-estimators for the GARCH model are defined through the one-step approach based on an asymptotic linearity result of a rank-based central sequence and uses data directly without requiring such transformation.

Similar to the linear regression and autoregressive models, the class of the R-estimators for the GARCH model are also asymptotically normal and highly efficient. However, unlike the commonly-used quasi-maximum likelihood estimator (QMLE) which is asymptotically normal under the finite fourth moment assump-

tion of the error distribution, the R-estimators proposed in this chapter are asymptotically normal under the assumption of only a finite $2+\delta$-th moment for some $\delta > 0$. The efficient property of the R-estimators is further confirmed based on the simulated data from the GARCH$(1,1)$ model and the higher order GARCH$(2,1)$ model to fill some void in the literature since the computation and empirical analysis for the higher order GARCH models are not considered widely. Analysis of real data shows that the numerical values of R-estimates can be different from the QMLE and the subsequent analysis of the GARCH residuals lead to conclude that such difference may be due to the assumption of the infinite fourth moment of the innovation distribution which may not hold and consequently leads to the failure of the QMLE.

Since the proposed class of the R-estimators are shown to converge to normal distributions, of which the covariance matrices do not have explicit forms, we employ a bootstrap method to approximate the distributions of the R-estimators. Chatterjee and Bose (2005) proved the consistency of the weighted bootstrap method for an estimator defined by smooth estimating equation. We consider weighted bootstrap with residual ranks that are integer-valued and non-smooth function. Our extensive simulation study provides evidence that the weighted bootstrap has good coverage rates even under a heavy-tailed distribution and for a small simple size.

Finally, we use the R-estimators for estimating parameters of the GJR(p,q) model proposed by Glosten et al. (1993), which is used to estimate the asymmetry effect of financial time series. Simulation results demonstrate good performance of the R-estimators for the GJR model.

The main contributions of the chapter are threefold. First, a new class of robust and efficient estimators for the GARCH model parameters is proposed. Second, the asymptotic distributions of the proposed estimators are derived based on weaker assumption on the error moment. Third, weighted bootstrap approximations of the distribution of the R-estimators are investigated empirically through the analysis

of real data and simulations. In particular, we propose algorithms for computing the R-estimators and the bootstrap replicates, which are computational friendly and easy to implement.

4.1.1 A motivating example

To illustrate the advantages of the class of R-estimators over the commonly-used QMLE, we consider below some simulation results corresponding to the GARCH$(2,1)$ model with underlying standardized innovation density (i) the normal distribution and (ii) the Student's t-distribution with $\nu = 3$ degrees of freedom, denoted as $t(3)$. We generate $R = 1000$ samples of size $n = 1000$ with parameter values $(\omega, \alpha_1, \alpha_2, \beta)' = (0.1, 0.1, 0.1, 0.6)'$. Simulation results described below are similar to various other choices of the true parameters. Consider three types of R-estimators given in Section 4.2.3. The boxplots of the QMLE and the proposed R-estimates based on the normal and $t(3)$ distributions are displayed in Figure 4.1 and Figure 4.2, respectively. Here the R-estimators and QMLE are standardized by multiplying with constant matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ defined in Section 4.2.5, so that all resulting estimators can estimate the parameter $\boldsymbol{\theta}_0$; see Section 4.2 for details.

An inspection of these two plots reveals overwhelming superiority of the R-estimators over the QMLE. Under the normal error distribution, the distribution patterns of the R-estimates of each parameter are quite similar to the QMLE around the true parameter value. However, under the $t(3)$ errors, the QMLE is lot more likely to deviate from the true parameter than the three R-estimates. Therefore, unlike the QMLE, the good performance of the R-estimators under the normal distribution is not at the cost of poor performance under the $t(3)$ distribution.

4.1.2 Outline of the chapter

The rest of the chapter is organized as follows. Section 4.2 defines the R-estimators based on an asymptotic linearity result of a rank-based central sequence. The

Figure 4.1: Boxplots of the QMLE and R-estimators (signs, van der Waerden, Wilcoxon) for the GARCH$(2,1)$ model under the normal distribution (sample size $n = 1000$; $R = 1000$ replications). The horizontal red line represents the actual parameter value.

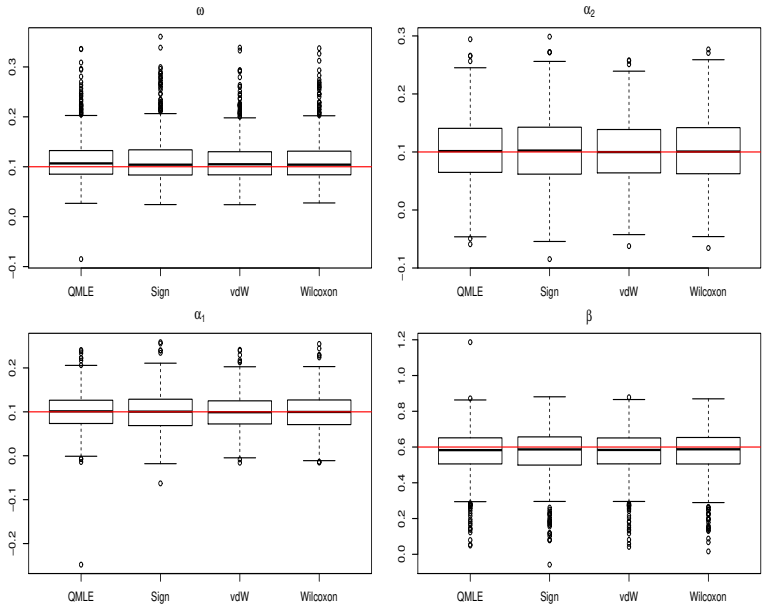
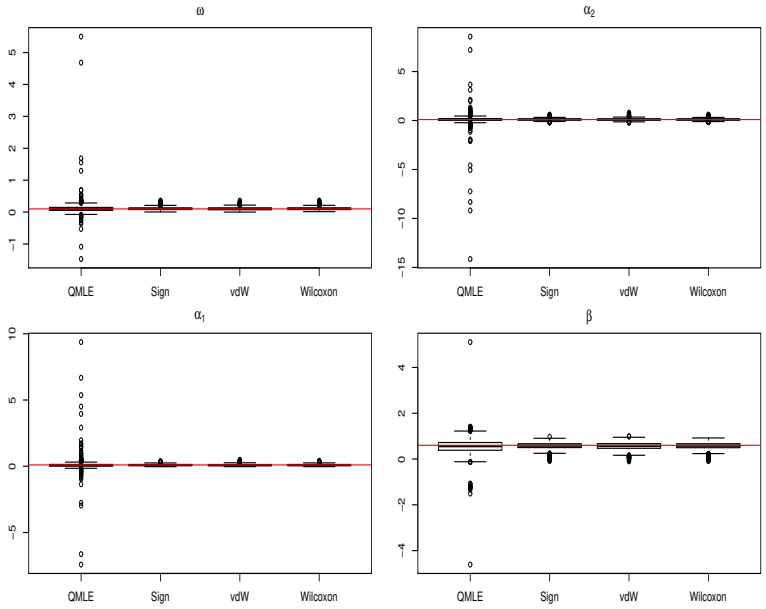


Figure 4.2: Boxplots of the QMLE and R-estimators (signs, van der Waerden, Wilcoxon) for the GARCH$(2,1)$ model under the $t(3)$ distribution (sample size $n = 1000$; $R = 1000$ replications). The horizontal red line represents the actual parameter value.



asymptotic distributions and efficiency of the R-estimators are discussed. Also, we give an algorithm for computing the R-estimators. Section 4.3 contains empirical and simulation results of the R-estimators. Section 4.4 describes the weighted

bootstrap for the R-estimators and includes extensive simulation results. Section 4.5 considers an application to the GJR model. Conclusion is given in Section 4.6. The technique used to establish the asymptotic distribution is included in Section 4.7.

4.2 The class of R-estimators for the GARCH model

In this section, we first define a central sequence of R-criteria $\{\hat{\boldsymbol{R}}_n(\boldsymbol{\theta})\}$ based on ranks of the residuals of the GARCH model. We prove the asymptotic uniform linear expansion (4.2.6) of this central sequence. Consequently, we define one-step R-estimator $\hat{\boldsymbol{\theta}}_n$ in (4.2.8). We propose a recursive algorithm for computation in Section 4.2.4.

Notations: Throughout the chapter, for a function g, we use $\dot{g}$ and $\ddot{g}$ to denote its first and second derivatives whenever they exist. We use c, b, c_1 to denote positive constants whose values can possibly change from line to line. Let ϵ be a generic random variable (r.v.) with the same distribution as $\{\epsilon_t\}$ and let F and f denote the cumulative distribution function (c.d.f.) and probability density function (p.d.f.) of ϵ, respectively. Let $\eta_t := \epsilon_t/\sqrt{c_\varphi}$, where $c_\varphi > 0$ satisfies (4.2.3), and η be a generic r.v. with the same distribution as $\{\eta_t\}$. Let G and g be the c.d.f. and p.d.f. of η, respectively. A sequence of stochastic process $\{Y_n(\cdot)\}$ is said to be $u_{\mathrm{P}}(1)$ (denoted by $Y_n = u_{\mathrm{P}}(1)$) if for every $c > 0$, $\sup\{|Y_n(\boldsymbol{b})|; ||\boldsymbol{b}|| \leq c\} = o_{\mathrm{P}}(1)$, where $||\cdot||$ stands for the Euclidean norm.

4.2.1 Rank-based central sequence

From Lemma 2.3 and Theorem 2.1 of Berkes et al. (2003), σ_t^2 of (3.1.2) has the unique almost sure representation $\sigma_t^2 = c_0 + \sum_{j=1}^{\infty} c_j X_{t-j}^2$, $t \in \mathbb{Z}$, where $\{c_j; j \geq 0\}$ are defined in (2.7)-(2.9) of Berkes et al. (2003).

Let $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})'$ denote the true parameter belonging

to a compact subset $\boldsymbol{\Theta}$ of $(0,\infty)^{1+p}\times(0,1)^q$. A typical element in $\boldsymbol{\Theta}$ is denoted by $\boldsymbol{\theta}=(\omega,\alpha_1,\ldots,\alpha_p,\beta_1,\ldots,\beta_q)'$. Define the variance function by

$$v_t(\boldsymbol{\theta})=c_0(\boldsymbol{\theta})+\sum_{j=1}^{\infty}c_j(\boldsymbol{\theta})X_{t-j}^2,\quad \boldsymbol{\theta}\in\boldsymbol{\Theta},t\in\mathbb{Z},$$

where the coefficients $\{c_j(\boldsymbol{\theta});j\geq 0\}$ are given in (3.1) of Berkes et al. (2003) with the property $c_j(\boldsymbol{\theta}_0)=c_j$, $j\geq 0$, so that the variance functions satisfy $v_t(\boldsymbol{\theta}_0)=\sigma_t^2$, $t\in\mathbb{Z}$ and

$$X_t=\{v_t(\boldsymbol{\theta}_0)\}^{1/2}\epsilon_t,\ \ 1\leq t\leq n.$$

Recall that $\{\hat{v}_t(\boldsymbol{\theta})\}$ is the observable approximation of $\{v_t(\boldsymbol{\theta})\}$, which is defined by

$$\hat{v}_t(\boldsymbol{\theta})=c_0(\boldsymbol{\theta})+I(2\leq t)\sum_{j=1}^{t-1}c_j(\boldsymbol{\theta})X_{t-j}^2,\ \boldsymbol{\theta}\in\boldsymbol{\Theta},\ 1\leq t\leq n.$$

Let $H^*(x)=x\{-\dot{f}(x)/f(x)\}$. The maximum likelihood estimator (MLE) is a solution of $\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta})=\mathbf{0}$, where

$$\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta}):=n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{v_t(\boldsymbol{\theta})}\left\{1-H^*\left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})}\right]\right\}.$$

However, f in H^* is usually unknown and we therefore consider an approximation to $\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta})$.

Let $\varphi:(0,1)\to\mathbb{R}$ be a score function satisfying some regularity conditions which will be discussed later. Let $R_{nt}(\boldsymbol{\theta})$ denote the rank of $X_t/v_t^{1/2}(\boldsymbol{\theta})$ among $\{X_j/v_j^{1/2}(\boldsymbol{\theta});1\leq j\leq n\}$. In linear regression models, the MLE has the same asymptotic efficiency as an R-estimator based on the score function $\varphi(u)=-\dot{f}(F^{-1}(u))/f(F^{-1}(u))$. For the estimation of the scale parameters, the MLE corresponds to the *central sequence*

$$\boldsymbol{R}_n(\boldsymbol{\theta}):=\boldsymbol{R}_{n,\varphi}(\boldsymbol{\theta})=n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{v_t(\boldsymbol{\theta})}\left\{1-\varphi\left[\frac{R_{nt}(\boldsymbol{\theta})}{n+1}\right]\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})}\right\}.\qquad(4.2.1)$$

However, since $v_t(\boldsymbol{\theta})$ is unobservable, we therefore replace it by $\hat{v}_t(\boldsymbol{\theta})$. Let $\hat{R}_{nt}(\boldsymbol{\theta})$

denote the rank of $X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})$ among $\{X_j/\hat{v}_j^{1/2}(\boldsymbol{\theta}); 1 \leq j \leq n\}$. We define *rank-based central sequence* as

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}) := \hat{\boldsymbol{R}}_{n,\varphi}(\boldsymbol{\theta}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})}{\hat{v}_t(\boldsymbol{\theta})} \left\{ 1 - \varphi \left[\frac{\hat{R}_{nt}(\boldsymbol{\theta})}{n+1} \right] \frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta})} \right\}. \tag{4.2.2}$$

4.2.2 One-step R-estimators and their asymptotic distributions

To define the R-estimator in terms of the classical Le Cam's one-step approach as in Hallin and La Vecchia (2017) and Hallin et al. (2020), we derive the asymptotic linearity of the rank-based central sequence under the following assumptions. Let $c_\varphi > 0$ be defined by

$$\sqrt{c_\varphi} = \mathrm{E}\left[\varphi\left(F\left(\epsilon_t\right)\right)\epsilon_t\right]$$

which satisfies

$$\mathrm{E}\left\{ \varphi\left[F\left(\epsilon_t\right)\right] \frac{\epsilon_t}{\sqrt{c_\varphi}} \right\} = 1. \tag{4.2.3}$$

Define $\mu(x) := \int_{-\infty}^{x} sg(s)ds$. Since $g(x) > 0$, $\mu(x)$ is strictly decreasing on $(-\infty, 0]$ with range $[\mu(0), 0]$ and strictly increasing on $[0, +\infty)$ with range $[\mu(0), 0]$. The functions $y \to \mu^{-1}(y)$ on $[\mu(0), 0]$ with ranges $(-\infty, 0]$ and $[0, +\infty)$ are well-defined when the ranges are considered separately.

The following conditions on the distribution of η_t are assumed for the proof of Theorem 4.7.1 on the approximation of a scale-perturbed weighted mixed-empirical process by its non-perturbed version.

Assumption (A1). (i). The p.d.f. $g(x)$, $xg(x)$ and $x^2g(x)$ are bounded on $x \in \mathbb{R}$; functions $y \to \mu^{-1}(y)g(\mu^{-1}(y))$ and $y \to (\mu^{-1}(y))^2 g(\mu^{-1}(y))$ are uniformly continuous on $[\mu(0), 0]$ when they are considered separately;

(ii).

$$\limsup_{\delta \to 0} \left\{ |x| \int_0^1 |xg(x) - (x + hx\delta)g(x + hx\delta)| dh; x \in \mathbb{R} \right\} = 0;$$

(iii). There is a $\delta > 0$ such that $\mathrm{E}|\eta_t|^{2+\delta} < \infty$.

We remark that Assumption (i) entails that $\mu(x)$ is uniformly Lipschitz continuous in scale in the sense that for some constant $0 < c < \infty$ and for every $s \in \mathbb{R}$, we have $\sup_{x\in\mathbb{R}} |\mu(x + xs) - \mu(x)| \leq c|s|$.

A more easily verifiable condition for Assumption (ii) can be obtained, for example, when g admits the derivative $\dot{g}$ which satisfies that for some $\delta > 0$,

$$\sup\{x^2 \sup |g(y) + y\dot{g}(y)|; x(1-\delta) < y < x(1+\delta)\} < \infty.$$

In particular, Assumptions (i), (ii) and (iii) in (A1) hold for a wide range of distributions, including normal, double-exponential, logistic and t-distributions with degrees of freedom more than 2.

We also need the following assumptions on the parameter space and the score function φ.

Assumption (A2). Let $\boldsymbol{\Theta}_0$ denote the set of interior points of $\boldsymbol{\Theta}$. We assume that $\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0\varphi} \in \boldsymbol{\Theta}_0$, where

$$\boldsymbol{\theta}_{0\varphi} = (c_\varphi \omega_0, c_\varphi \alpha_{01}, ..., c_\varphi \alpha_{0p}, \beta_{01}, ..., \beta_{0q})' \tag{4.2.4}$$

is a transformation of the true parameter $\boldsymbol{\theta}_0$.

Assumption (A3). The score function φ is non-decreasing, right-continuous with only a finite number of points of discontinuity and is bounded on $(0, 1)$.

To state the asymptotic linearity of $\hat{\boldsymbol{R}}_n(\boldsymbol{\theta})$, write v_t and $\dot{\boldsymbol{v}}_t$ for $v_t(\boldsymbol{\theta}_{0\varphi})$ and $\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})$, respectively. Let

$$\gamma(\varphi) := \int_0^1 \int_0^1 G^{-1}(u) G^{-1}(v) \left[\min\{u, v\} - uv\right] d\varphi(u) d\varphi(v), \ \boldsymbol{J} := \mathrm{E}(\dot{\boldsymbol{v}}_t \dot{\boldsymbol{v}}_t'/v_t^2)$$

$$\rho(\varphi) := \int_0^1 \{G^{-1}(u)\}^2 g\{G^{-1}(u)\} d\varphi(u), \ \sigma^2(\varphi) := \mathrm{E}\left\{\varphi\left[G(\eta_t)\right]\eta_t\right\}^2 - 1,$$

$$\lambda(\varphi) := \int_0^1 \int_0^1 G^{-1}(u) I(v \le u)(1 - G^{-1}(v)\varphi(v)) dv d\varphi(u). \tag{4.2.5}$$

Let Z be the r.v. $Z := \int_0^1 G^{-1}(u) B(u) d\varphi(u)$, where $B(.)$ is the standard Brownian bridge. Then Z has mean zero and variance $\gamma(\varphi)$; see the proof in Lemma 4.7.5 for details. Let $\tilde{G}_n(x)$, $x \in \mathbb{R}$ be the empirical distribution function of $\{\eta_t\}$ (which is unobservable),

$$\boldsymbol{Q}_n := \int_0^1 n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t}{v_t} \left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u)) \right] d\varphi(u),$$

$$\boldsymbol{N}_n := n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t}{v_t} \left\{ 1 - \eta_t \varphi\left[G(\eta_t) \right] \right\}.$$

The following proposition states the asymptotic uniform linearity of $\hat{\boldsymbol{R}}_n(\boldsymbol{\theta})$.

Proposition 4.2.1. *Let Assumptions (A1)-(A3) hold. Then for $\boldsymbol{b} \in \mathbb{R}^{1+p+q}$ with $||\boldsymbol{b}|| < c$,*

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi} + n^{-1/2}\boldsymbol{b}) - \hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) = (1/2 + \rho(\varphi)/2)\boldsymbol{J}\boldsymbol{b} + u_{\mathrm{P}}(1). \tag{4.2.6}$$

Moreover,

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{Q}_n + \boldsymbol{N}_n + u_{\mathrm{P}}(1), \tag{4.2.7}$$

where $\boldsymbol{Q}_n$ converges in distribution to $\mathrm{E}(\dot{\boldsymbol{v}}_1/v_1)Z$ with mean zero and covariance matrix $\mathrm{E}(\dot{\boldsymbol{v}}_1/v_1)\mathrm{E}(\dot{\boldsymbol{v}}_1'/v_1)\gamma(\varphi)$ and $\boldsymbol{N}_n \to \mathcal{N}(\boldsymbol{0}, \boldsymbol{J}\sigma^2(\varphi))$.

The above asymptotic linearity allows us to define a class of R-estimators through the one-step approach. Let $\{\hat{\boldsymbol{\Upsilon}}_n\}$ be a sequence of consistent estimator of $\boldsymbol{\Upsilon}_{\varphi,g}(\boldsymbol{\theta}_{0\varphi}) := (1/2+\rho(\varphi)/2)\boldsymbol{J}$; see Section 4.2.4 for a construction of $\hat{\boldsymbol{\Upsilon}}_n$. Let $\bar{\boldsymbol{\theta}}_n$ be a root-n consistent and asymptotically discrete estimator of $\boldsymbol{\theta}_{0\varphi}$. More precisely, a sequence $\{\bar{\boldsymbol{\theta}}_n\}$ is called discrete if there exists $K \in \mathbb{N}$ such that independent of

$n \in \mathbb{N}$, $\bar{\boldsymbol{\theta}}_n$ takes on at most K different values in

$$\mathcal{Q}_n := \{\boldsymbol{\theta} \in \mathbb{R}^{1+p+q} : n^{-1/2} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq c\}, \quad c > 0 \text{ fixed};$$

see Kresis (1987, Section 4) for details. We remark that here asymptotically discreteness is only of theoretical interest since in practice $\bar{\boldsymbol{\theta}}_n$ always has a bounded number of digits; see Le Cam and Yang (2000, Chapter 6) and van der Vaart (1998, Section 5.7) for more details. Then the one-step R-estimator is defined as

$$\hat{\boldsymbol{\theta}}_n := \bar{\boldsymbol{\theta}}_n - n^{-1/2} \left(\hat{\boldsymbol{\Upsilon}}_n\right)^{-1} \hat{\boldsymbol{R}}_n(\bar{\boldsymbol{\theta}}_n). \tag{4.2.8}$$

Note that strictly speaking, the R-estimators based on this definition are not functions of the ranks of the residuals only. However, we borrow the terminology from the regression and the homoscedastic-autoregression settings and still call them (generalized) R-estimators. When, for example, $\varphi(u) = u - 1/2$, $\hat{\boldsymbol{\theta}}_n$ is an analogue of the Wilcoxon type R-estimator.

The following theorem shows that the R-estimator defined in (4.2.2) is $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_{0\varphi}$.

Theorem 4.2.1. *Let Assumptions (A1)-(A3) hold. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}\right) = -(1/2 + \rho(\varphi)/2)^{-1} \boldsymbol{J}^{-1}(\boldsymbol{Q}_n + \boldsymbol{N}_n) + o_{\mathrm{P}}(1). \tag{4.2.9}$$

Hence as $n \to \infty$, $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}\right)$ is normal with mean $\mathbf{0}$ and covariance matrix

$$\boldsymbol{J}^{-1} \frac{[4\gamma(\varphi) + 8\lambda(\varphi)]\, \mathrm{E}(\dot{\boldsymbol{v}}_1/v_1)\mathrm{E}\,(\dot{\boldsymbol{v}}_1'/v_1) + 4\sigma^2(\varphi)\boldsymbol{J}}{(1+\rho(\varphi))^2} \boldsymbol{J}^{-1}.$$

We remark that according to Theorem 4.2.1, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$ has a complicated form. Hence we consider bootstrap methods in Section 4.4 to approximate the limit distribution of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}\right)$.

Notice that similar to the M-estimator for the GARCH model, $\hat{\boldsymbol{\theta}}_n$ turns out to

be an estimator of transformed parameter $\boldsymbol{\theta}_{0\varphi}$ defined in (4.2.4), where the factor c_φ is unknown since the distribution of ϵ_t is unknown. Although c_φ is unknown, one can still apply the R-estimator to get a consistent estimate of some important quantity in financial risk management such as the Value at Risk (VaR) when the returns are modeled as GARCH(p, q). This is because c_φ is cancelled in the estimation process. See, e.g., Iqbal and Mukherjee (2010) for how to estimate the VaR based on the GARCH model.

4.2.3 Examples of the score functions

Below we cite examples of three commonly-used R-scores; for similar examples of scores in other models and settings, see Section 2.3 in Chapter 2, Mukherjee (2007) and Hallin and La Vecchia (2017).

Example 1 (sign score). Let $\varphi(u) = \text{sign}(u - 1/2)$. Then for symmetric innovation distribution, $c_\varphi = (\text{E}|\epsilon|)^2$, which coincides with the scale factor of the LAD estimator in Mukherjee (2008). Therefore, the sign R-estimator is expected to be close to the LAD estimator. This is demonstrated later in the real data analysis.

Example 2 (Wilcoxon score). Let $\varphi(u) = u - 1/2$ so that the range of $\varphi(u)$ is symmetric.

Example 3 (van der Waerden (vdW) or normal score). One might also set $\varphi(u) = \Phi^{-1}(u)$, with $\Phi(\cdot)$ denoting the c.d.f. of the standard normal distribution. Notice that unlike the sign and Wilcoxon score, the vdW score is not bounded as $u \to 0$ and $u \to 1$. It thus does not satisfy Assumption (A3). However, an approximating sequence of bounded score functions of φ on $(0, 1)$ can be constructed as in Andrews (2012). It is demonstrated later using both real data analysis and extensive simulation that the vdW has superior performance compared with the QMLE.

We now provide heuristics for the definition of the R-estimator in (4.2.1). When the underlying error distribution is known, one can obtain efficient R-estimator by

choosing the score function as $\varphi(u) = -\dot{f}(F^{-1}(u))/f(F^{-1}(u))$. Since for large n, the empirical distribution function $R_{nt}(\boldsymbol{\theta}_{0\varphi})/(n+1)$ of $\{\epsilon_j; 1 \leq j \leq n\}$ evaluated at ϵ_t is close to $F(\epsilon_t)$, we have

$$\varphi\left[\frac{R_{nt}(\boldsymbol{\theta})}{n+1}\right]\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})} \approx H^*\left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})}\right].$$

Therefore, the criteria function of the R-estimator gets close to the MLE which is efficient. This leads to the choice of the vdW, sign and Wilcoxon under the normal, double exponential (DE) and logistic distributions, respectively. This is observed later in simulation study of the R-estimator.

4.2.4 Computational aspects

Here we discuss some key computational aspects and propose an algorithm to compute $\hat{\boldsymbol{\theta}}_n$.

First, since c_φ depends on the unknown density f, it is difficult to have a $\sqrt{n}$-consistent initial estimator $\bar{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_{0\varphi}$. However, due to finite sample size in practice, the one-step procedure is usually iterated a number of times, taking $\hat{\boldsymbol{\theta}}_n$ as the new initial estimate, until it stabilizes numerically. This iteration process would mitigate the impact of different initial estimates; see van der Vaart (1998, Section 5.7) and Hallin and La Vecchia (2017) for similar comments. In fact, we observed during our extensive simulation study that irrespective of the choice of the QMLE, LAD or $\boldsymbol{\theta}_0$ as initial estimates, only few iterations result in the same estimates.

Second, to compute $\hat{\boldsymbol{\theta}}_n$ of (4.2.8), we need $\hat{\boldsymbol{\Upsilon}}_n$ which is a consistent estimator of $(1/2 + \rho(\varphi)/2)\boldsymbol{J}$. The matrix $\boldsymbol{J}$ can be consistently estimated by

$$\hat{\boldsymbol{J}}_n(\bar{\boldsymbol{\theta}}_n) := n^{-1}\sum_{t=1}^{n}\{\dot{\hat{\boldsymbol{v}}}_t(\bar{\boldsymbol{\theta}}_n)\dot{\hat{\boldsymbol{v}}}_t'(\bar{\boldsymbol{\theta}}_n)/\hat{v}_t^2(\bar{\boldsymbol{\theta}}_n)\}.$$

For estimating $\rho(\varphi)$ which is a function of the density g, we can use the asymptotic linearity in (4.2.6). Here with an arbitrarily chosen $\boldsymbol{b}$, we can substitute $\bar{\boldsymbol{\theta}}_n$

for $\boldsymbol{\theta}_{0\varphi}$ and then solve the equation for $\rho(\varphi)$ based on (4.2.6). A more delicate approach for estimating $\rho(\varphi)$ can be found in Cassart et al. (2010) and Hallin and La Vecchia (2017, Appendix C). Based on our extensive simulation study and real data analysis, it appears that different values of $\rho(\varphi)$ would finally lead to same estimate after some iterations. Consequently, we set $\rho(\varphi) = 1$ during the computation which is the value corresponding to the vdW score under the normal distribution.

In summary, we propose the following iterative algorithm to compute the R-estimator.

$$\hat{\boldsymbol{\theta}}_{(r+1)} = \hat{\boldsymbol{\theta}}_{(r)} - \left[\sum_{t=1}^{n} \frac{\dot{\hat{\boldsymbol{v}}}_t(\hat{\boldsymbol{\theta}}_{(r)})\dot{\hat{\boldsymbol{v}}}_t'(\hat{\boldsymbol{\theta}}_{(r)})}{\hat{v}_t^2(\tilde{\boldsymbol{\theta}}_{(r)})}\right]^{-1} \times \left\{\sum_{t=1}^{n} \frac{\dot{\hat{\boldsymbol{v}}}_t(\hat{\boldsymbol{\theta}}_{(r)})}{\hat{v}_t(\tilde{\boldsymbol{\theta}}_{(r)})}\left[1 - \varphi\left(\frac{R_{nt}(\hat{\boldsymbol{\theta}}_{(r)})}{n+1}\right)\frac{X_t}{\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_{(r)})}\right]\right\}, \quad \text{for} \quad r = 0, 1, ..., \tag{4.2.10}$$

with $\hat{\boldsymbol{\theta}}_{(0)} = \bar{\boldsymbol{\theta}}_n$ being the initial estimator.

4.2.5 Asymptotic relative efficiency

In the linear regression and autoregressive models, the asymptotic relative efficiency (ARE) of the R-estimators is high with respect to (wrt) the least squares estimator for a wide array of error distributions. For the GARCH model, we compare the ARE of the R-estimator wrt the QMLE based on Theorem 4.2.1.

Define a diagonal matrix of order $(1 + p + q) \times (1 + p + q)$ by

$$\boldsymbol{A} := \text{diag}(\underbrace{c_\varphi^{-1}, ..., c_\varphi^{-1}}_{1+p}\underbrace{1, ..., 1}_{q}).$$

Then $\boldsymbol{A}\hat{\boldsymbol{\theta}}_n$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$ for all score functions φ. Using the forms of $\{c_j(\boldsymbol{\theta}); j \geq 0\}$ in (3.1) of Berkes et al. (2003),

$$v_t(\boldsymbol{A}^{-1}\boldsymbol{\theta}) = c_\varphi v_t(\boldsymbol{\theta}), \quad \dot{\boldsymbol{v}}_t(\boldsymbol{A}^{-1}\boldsymbol{\theta}) = c_\varphi \boldsymbol{A}\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}).$$

Since $\boldsymbol{\theta}_{0\varphi} = \boldsymbol{A}^{-1}\boldsymbol{\theta}_0$ with $\boldsymbol{J}_0 := \mathrm{E}[\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_0)\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_0)/v_t^2(\boldsymbol{\theta}_0)]$, we have

$$\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} = \boldsymbol{A}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_0)}{v_t(\boldsymbol{\theta}_0)}, \quad \boldsymbol{J} = \boldsymbol{A}\boldsymbol{J}_0\boldsymbol{A}.$$

Thus Theorem 4.2.1 implies that as $n \to \infty$, $n^{1/2}\left(\boldsymbol{A}\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ converges to the normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\boldsymbol{C}_1 := \boldsymbol{J}_0^{-1}\frac{[4\gamma(\varphi) + 8\lambda(\varphi)]\,\mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_0)/v_1(\boldsymbol{\theta}_0))\mathrm{E}\,(\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_0)/v_1(\boldsymbol{\theta}_0)) + 4\sigma^2(\varphi)\boldsymbol{J}_0}{(1+\rho(\varphi))^2}\boldsymbol{J}_0^{-1}.$$

For the QMLE $\hat{\boldsymbol{\theta}}_{\mathrm{QMLE}}$, we can derive a similar result as follows. Define a $(1+p+q)\times(1+p+q)$ diagonal matrix $\boldsymbol{B} = \mathrm{diag}(\underbrace{(\mathrm{E}\epsilon^2)^{-1}, .., (\mathrm{E}\epsilon^2)^{-1}}_{1+p}\underbrace{1, ..., 1}_{q})$. When $\mathrm{E}\epsilon^4 < \infty$,

$$\sqrt{n}\left(\boldsymbol{B}\hat{\boldsymbol{\theta}}_{\mathrm{QMLE}} - \boldsymbol{\theta}_0\right) \to N(\mathbf{0}, \boldsymbol{C}_2),$$

where $\boldsymbol{C}_2 = (\mathrm{E}\epsilon^4/(\mathrm{E}\epsilon^2)^2 - 1)\boldsymbol{J}_0^{-1}$. The ARE of the R-estimator wrt the QMLE is

$$\begin{aligned}\boldsymbol{C}_1^{-1}\boldsymbol{C}_2 = \boldsymbol{J}_0 &\left\{[4\gamma(\varphi) + 8\lambda(\varphi)]\,\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_0)}{v_1(\boldsymbol{\theta}_0)}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_0)}{v_1(\boldsymbol{\theta}_0)}\right) + 4\sigma^2(\varphi)\boldsymbol{J}_0\right\}^{-1} \\ &\times (1+\rho(\varphi))^2\left(\frac{E\epsilon^4}{(\mathrm{E}\epsilon^2)^2} - 1\right). \qquad (4.2.11)\end{aligned}$$

For the sign R-estimator, $\gamma(\varphi)$, $\lambda(\varphi)$ and $\rho(\varphi)$ are all zeros. Hence $\boldsymbol{C}_1^{-1}\boldsymbol{C}_2$ reduces to

$$(\mathrm{E}\epsilon^4/(\mathrm{E}\epsilon^2)^2 - 1)/(4\sigma^2(\varphi))\boldsymbol{I}_{1+p+q},$$

where $\boldsymbol{I}_{1+p+q}$ is the $(1+p+q)\times(1+p+q)$ identity matrix. Consequently, the ARE of the sign R-estimator wrt the QMLE equals

$$(\mathrm{E}\epsilon^4/(\mathrm{E}\epsilon^2)^2 - 1)/(4\sigma^2(\varphi)),$$

which is 0.876 under the normal distribution. This corresponds to the classical result of the ARE of the mean absolute deviation wrt the mean square deviation; see, e.g., Huber and Ronchetti (2009, Chapter 1).

For the vdW and Wilcoxon R-estimators, the AREs are more difficult to calculate since $\gamma(\varphi)$ and $\lambda(\varphi)$ are non-zero. However, in the following simulation study in Table 4.2, the estimated AREs reveal that the vdW R-estiamtor, compared with the QMLE, does not lose any efficiency even under the normal distribution.

Scale-transformation invariance of the ARE. The following proposition states that the ARE of the R-estimator wrt the QMLE enjoys the invariance property of the scale-transformation in terms of the innovation. Hence, with underlying distributions of the same type but with different variances, the AREs remain the same. In particular, let $e(\epsilon)$ denote the ARE when the innovation term is $\{\epsilon_t\}$. Then we have the following proposition.

Proposition 4.2.2. *Under Assumptions (A1)-(A3), $e(\tilde{\epsilon}) = e(\epsilon)$, where $\tilde{\epsilon} = c\epsilon$ for a constant c.*

Proof. Let $\tilde{c}_\varphi$, $\tilde{\eta}_t$, $\tilde{v}_t(\boldsymbol{\theta})$, $\tilde{\sigma}^2(\varphi)$, $\tilde{\rho}(\varphi)$, $\tilde{\gamma}(\varphi)$, $\tilde{\lambda}(\varphi)$ denote the counterparts of c_φ, η_t, $v_t(\boldsymbol{\theta})$, $\sigma^2(\varphi)$, $\rho(\varphi)$, $\gamma(\varphi)$, $\lambda(\varphi)$ when the innovation term is $\{\tilde{\epsilon}_t\}$. Since $\tilde{c}_\varphi = c^2 c_\varphi$, we have $\tilde{\eta}_t = \eta_t$ and this implies that

$$\tilde{\sigma}^2(\varphi) = \sigma^2(\varphi), \tilde{\rho}(\varphi) = \rho(\varphi), \tilde{\gamma}(\varphi) = \gamma(\varphi), \tilde{\lambda}(\varphi) = \lambda(\varphi), \mathrm{E}\tilde{\epsilon}^4/(\mathrm{E}\tilde{\epsilon}^2)^2 = \mathrm{E}\epsilon^4/(\mathrm{E}\epsilon^2)^2.$$

Thus, in view of (4.2.11), we obtain $e(\tilde{\epsilon}) = e(\epsilon)$. □

Using the similar method of proving Proposition 4.2.2, it is easy to show that the scale-transformation invariance of the ARE also holds in terms of the score function. Specifically, let $e(\varphi)$ denote the ARE when the score function is φ. Then $e(\tilde{\varphi}) = e(\varphi)$ for $\tilde{\varphi} = c\varphi$ for a constant $c > 0$.

4.3 Real data analysis and simulation results

This section examines the performance of the R-estimators and compare them with the QMLE by analysing three financial time series and by carrying out extensive Monte Carlo simulation.

4.3.1 Real data analysis

In this section we fit GARCH$(1,1)$ model to three financial time series and compare the proposed three R-estimators with the M-estimators QMLE and LAD discussed in Mukherjee (2008).

In an earlier work, Muler and Yohai (2008) fitted the the GARCH$(1,1)$ model to the Electric Fuel Corporation (EFCX) time series for the period of January 2000 to December 2001 with sample size $n = 498$. The parameters of the model are estimated by M-estimators based on various score functions. It turned out that the M-estimates of the parameter β differ widely depending on the score functions and so it is difficult to assess which estimate should be relied on in similar situations. Here we compare various M-estimates and R-estimates of the GARCH$(1,1)$ parameters for the EFCX series again shedding light on which could be some possible reasons for the difference in estimates and finally which estimation methods can be relied upon. We also compare M-estimates of the GARCH$(1,1)$ parameters when fitted to two other dataset, namely, the S&P 500 stock index from June 2013 to May 2017 with $n = 1005$ and the GBP/USD exchange rate from June 2013 to May 2017 with $n = 998$ to illustrate that the M- and R-estimates of β do not differ widely when the underlying theoretical assumptions hold in general.

In Table 4.1, we report the QMLE computed using the `fGarch` package in R program, the M-estimates QMLE and LAD and the R-estimates proposed in Examples 1-3 of Section 4.2.3. For the EFCX data, the R-estimates of β for all score functions are quite close to the LAD estimate, but they are very different than the QMLE. On the contrary, for the S&P 500 and GBP/USD data, all these estimates of β are close to each other. We can also find that the LAD estimates of all parameters are quite close to the sign estimates and this is consistent with the discussion in Example 1 of Section 4.2.3. Note also that for ω and α, the R-estimates are quite different since c_φ's of these scores have different values.

To investigate why the QMLE of β is different from the other R-estimates and LAD for the EFCX data, we check the assumption $\mathrm{E}\epsilon^4 < \infty$ for this data

Table 4.1: The QMLE, LAD and R-estimates (sign, Wilcoxon and vdW) of the GARCH(1, 1) parameter for the EFCX, S&P 500 and GBP/USD data.

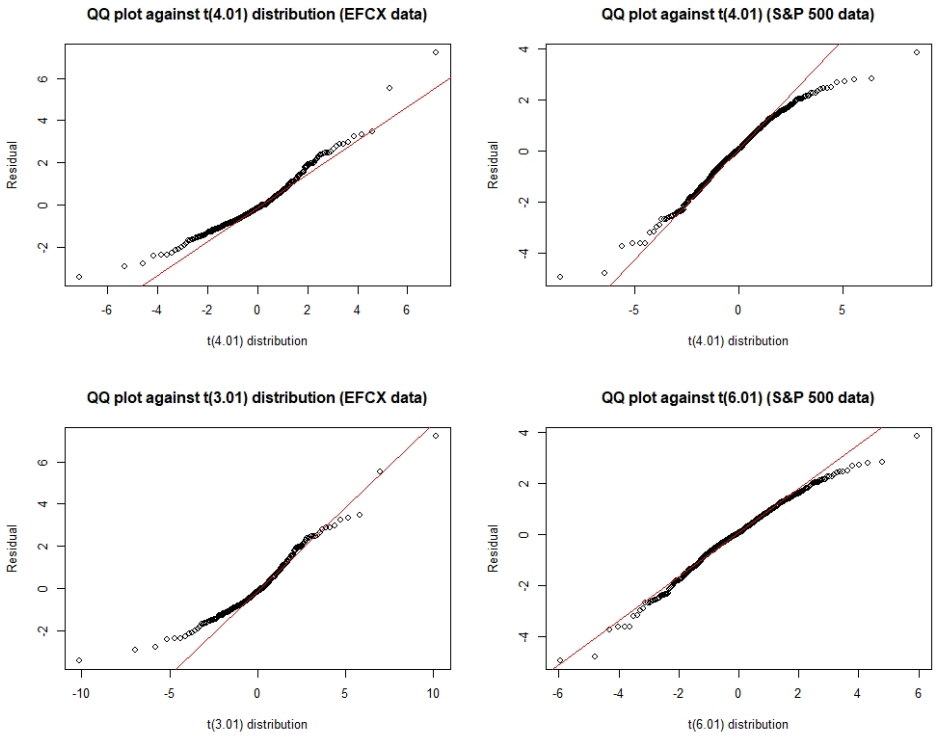
	`fGarch`	QMLE	LAD	sign	Wilcoxon	vdW
EFCX						
ω	1.89×10^{-4}	6.28×10^{-4}	6.43×10^{-4}	6.27×10^{-4}	8.44×10^{-5}	1.12×10^{-3}
α	4.54×10^{-2}	7.20×10^{-2}	8.87×10^{-2}	8.81×10^{-2}	1.09×10^{-2}	0.12
β	0.92	0.84	0.66	0.65	0.67	0.69
S&P 500						
ω	6.50×10^{-6}	7.02×10^{-6}	3.02×10^{-6}	3.02×10^{-6}	4.02×10^{-7}	5.97×10^{-6}
α	0.18	0.18	0.11	0.11	1.41×10^{-2}	0.18
β	0.72	0.70	0.73	0.73	0.73	0.72
GBP/USD						
ω	5.32×10^{-7}	1.02×10^{-6}	3.64×10^{-7}	3.64×10^{-7}	5.23×10^{-8}	8.73×10^{-7}
α	0.11	0.13	3.74×10^{-2}	3.74×10^{-2}	5.18×10^{-3}	8.53×10^{-2}
β	0.88	0.85	0.91	0.91	0.91	0.89

by using the QQ-plots of the residuals based on the QMLE and the R-estimates corresponding to the vdW score against t distributions. We consider the vdW score only since the R-estimates based on two other score functions and the LAD are close to the vdW estimates. For comparison, we have also provided QQ-plots for the S&P 500 data. The main idea behind the QQ-plots of the residuals against the $t(d)$ distribution is simple. Since if $\epsilon \sim t(d)$ distribution then $\mathrm{E}|\epsilon|^\nu < \infty$ if and only if $\nu < d$, residuals with heavier tail than the $t(d)$ distribution correspond to the errors with the infinite d-th moment while those with thinner tail than the $t(d)$ distribution have the finite d-th error moment.

The top-left panel of Figure 4.3 shows the QQ-plot of the residuals against the $t(4.01)$ distribution for the EFCX data. The residuals have heavier right tail than the $t(4.01)$ distribution which implies that the fourth moment of the error term may not exist. On the other hand, the QQ-plot against the $t(3.01)$ distribution reveals lighter tail as shown at the bottom-left panel of Figure 4.3 and this implies that $\mathrm{E}|\epsilon|^3 < \infty$.

For the S&P 500 data, the QQ-plot against $t(4.01)$ distribution at the top-right panel of Figure 4.3 shows that the residuals have lighter tails than $t(4.01)$

Figure 4.3: QQ-plots of the residuals against t-distributions for the EFCX (left column) and S&P 500 data (right column); the residuals are obtained by using the vdW R-estimator.



distribution. For the QQ-plot against $t(6.01)$ distribution, as shown at the bottom-right panel of Figure 4.3, the residuals fit the distribution better. Therefore, we may conclude that $\mathrm{E}|\epsilon|^4 < \infty$ holds for the S&P 500 data.

4.3.2 Simulation study of the R-estimators

We now evaluate the performance of the R-estimators based on simulated data from various error distributions. Apart from the GARCH$(1,1)$ model we consider the GARCH$(2,1)$ model also as the computation for higher order models are not considered frequently in the literature. Let R denote the number of replications and $\hat{\boldsymbol{\theta}}_{ni} = (\hat{\omega}_i, \hat{\alpha}_{i1}, ..., \hat{\alpha}_{ip}, \hat{\beta}_{i1}, ..., \hat{\beta}_{iq})'$ denote the R-estimator computed from the i-th data, $1 \leq i \leq n$. Note that $\hat{\boldsymbol{\theta}}_{ni}$ is an estimator of $\boldsymbol{\theta}_{0\varphi}$, which depends on the score function used in the estimation. To compare R-estimates based on different

score functions fairly, we consider the standardized bias defined by

$$\frac{1}{R}\sum_{i=1}^{R}\left(\frac{\hat{\omega}_i}{c_\varphi}-\omega_0,\frac{\hat{\alpha}_{i1}}{c_\varphi}-\alpha_{01},...,\frac{\hat{\alpha}_{ip}}{c_\varphi}-\alpha_{0p},\hat{\beta}_{i1}-\beta_{01},...,\hat{\beta}_{iq}-\beta_{0q}\right)',$$

and the standardized MSE defined by

$$\frac{1}{R}\sum_{i=1}^{R}\left(\left(\frac{\hat{\omega}_i}{c_\varphi}-\omega_0\right)^2,\left(\frac{\hat{\alpha}_{i1}}{c_\varphi}-\alpha_{01}\right)^2,...,\left(\frac{\hat{\alpha}_{ip}}{c_\varphi}-\alpha_{0p}\right)^2,\left(\hat{\beta}_{i1}-\beta_{01}\right)^2,...,\left(\hat{\beta}_{iq}-\beta_{0q}\right)^2\right)'.$$

We also compare the relative efficiency of the R-estimators wrt the QMLE under a finite sample size, as an estimate of the ARE, by using the formula

$$\widehat{\mathrm{ARE}}_{\mathrm{R/QMLE}}=\widehat{\mathrm{MSE}}_{\mathrm{QMLE}}/\widehat{\mathrm{MSE}}_{\mathrm{R}}.$$

Simulation for the GARCH$(1,1)$ model. Here we run simulation with $R=500, n=1000$ and $\boldsymbol{\theta}_0=(6.50\times10^{-6},0.177,0.716)'$, where our choice of $\boldsymbol{\theta}_0$ is motivated by the estimate given by the `fGarch` for the S&P 500 data in Table 4.1. The estimates of the standaradized bias and MSE of the R-estimators and QMLE under various error distributions are reported in Table 4.2. The estimates of the ARE are shown in the parentheses. Notice that under $t(3)$ distribution, the QMLE does not converge for many replications, while the R-estimators always converge. Therefore, the bias and MSE are obtained using the replications where the QMLE converges.

The bias of the M-estimators is close to zero under all distributions considered in Table 4.2, indicating consistency of the M-estimators. It is worth noting that the vdW achieves almost the same efficiency as the QMLE under the normal distribution, and the vdW is more efficient under heavier-tailed distributions. In general, the sign score is most efficient under the DE and $t(3)$ distributions, while the Wilcoxon score is optimal under the logistic distribution. Under the $t(3)$ error distribution with infinite fourth moment, the R-estimators outperform the QMLE in terms of both bias and MSE. For instance, for α, the bias ratio of the QMLE

Table 4.2: The estimates of the standardized bias, MSE and ARE of the R-estimators (sign, Wilcoxon and vdW) and the QMLE for the GARCH$(1,1)$ model under various error distributions with sample size $n = 1000$ based on $R = 500$ replications.

	Standardized bias			**Standardized MSE and ARE**		
	ω	α	β	ω	α	β
Normal						
QMLE	8.96×10^{-7}	-4.42×10^{-4}	-1.54×10^{-2}	6.45×10^{-12}	1.41×10^{-3}	4.14×10^{-3}
Sign	9.30×10^{-7}	1.74×10^{-3}	-1.54×10^{-2}	8.39×10^{-12} (0.77)	1.62×10^{-3} (0.87)	5.16×10^{-3} (0.80)
Wilcoxon	1.02×10^{-6}	3.09×10^{-3}	-1.61×10^{-2}	8.52×10^{-12} (0.76)	1.54×10^{-3} (0.91)	4.93×10^{-3} (0.84)
vdW	9.05×10^{-7}	4.55×10^{-4}	-1.55×10^{-2}	6.44×10^{-12} (1.00)	1.43×10^{-3} (0.98)	4.15×10^{-3} (1.00)
DE						
QMLE	1.02×10^{-6}	3.56×10^{-3}	-2.26×10^{-2}	8.60×10^{-12}	2.37×10^{-3}	6.29×10^{-3}
Sign	5.82×10^{-7}	-3.42×10^{-3}	-1.69×10^{-2}	6.22×10^{-12} (1.38)	1.74×10^{-3} (1.36)	5.15×10^{-3} (1.22)
Wilcoxon	6.24×10^{-7}	-2.93×10^{-3}	-1.74×10^{-2}	6.34×10^{-12} (1.36)	1.76×10^{-3} (1.35)	5.12×10^{-3} (1.23)
vdW	6.22×10^{-7}	-4.13×10^{-3}	-1.96×10^{-2}	6.51×10^{-12} (1.32)	1.88×10^{-3} (1.26)	5.45×10^{-3} (1.15)
Logistic						
QMLE	1.05×10^{-6}	2.51×10^{-3}	-1.51×10^{-2}	7.44×10^{-12}	1.63×10^{-3}	4.28×10^{-3}
Sign	6.85×10^{-7}	-2.65×10^{-3}	-1.17×10^{-2}	5.40×10^{-12} (1.38)	1.42×10^{-3} (1.15)	3.66×10^{-3} (1.17)
Wilcoxon	6.82×10^{-7}	-2.91×10^{-3}	-1.19×10^{-2}	5.24×10^{-12} (1.42)	1.38×10^{-3} (1.18)	3.56×10^{-3} (1.20)
vdW	7.06×10^{-7}	-3.80×10^{-3}	-1.34×10^{-2}	5.66×10^{-12} (1.31)	1.42×10^{-3} (1.14)	3.83×10^{-3} (1.12)
***t*(3)**						
QMLE	9.96×10^{-7}	2.99×10^{-2}	-5.46×10^{-2}	2.53×10^{-11}	2.74×10^{-2}	2.81×10^{-2}
Sign	4.33×10^{-7}	4.82×10^{-3}	-1.80×10^{-2}	6.78×10^{-12} (3.73)	3.72×10^{-3} (7.37)	7.73×10^{-3} (3.64)
Wilcoxon	4.15×10^{-7}	4.41×10^{-3}	-1.83×10^{-2}	7.10×10^{-12} (3.57)	3.86×10^{-3} (7.10)	8.18×10^{-3} (3.44)
vdW	3.92×10^{-7}	3.77×10^{-3}	-2.57×10^{-2}	9.38×10^{-12} (2.70)	5.33×10^{-3} (5.14)	1.14×10^{-2} (2.47)

with respect to the vdW is 7.9; the MSE ratio of the QMLE with respect to the sign is 7.37.

To strengthen the point that the R-estimators behave better than the QMLE under a heavy-tailed distribution, we have reported simulation results for larger sample sizes $n = 3000$ and $n = 5000$ under $t(3)$ distribution in Table 4.3. The QMLE failed to converge for large sample size; for example, with $n = 5000$ around 8% replications do not converge. From Table 4.3, when n increases, the performance of the R-estimators becomes even better in terms of both the bias and MSE.

Overall, the vdW dominates the QMLE and other R-estimators sacrifice small efficiency under the normal error distribution while they achieve much higher efficiency when tails become much heavier. This provides a strong support for using the R-estimators.

Simulation for the GARCH$(2,1)$ model. It was reported in Francq and

Table 4.3: The estimates of the standardized bias, MSE and ARE of the R-estimators (sign, Wilcoxon and vdW) and the QMLE for the GARCH$(1,1)$ model under the $t(3)$ error distribution with larger sample sizes $n = 3000, 5000$ based on $R = 500$ replications.

	Standardized bias			**Standardized MSE and ARE**		
	ω	α	β	ω	α	β
n = 3000						
QMLE	6.34×10^{-7}	1.80×10^{-2}	-3.48×10^{-2}	1.14×10^{-11}	1.61×10^{-2}	1.25×10^{-2}
Sign	1.52×10^{-7}	1.46×10^{-3}	-9.99×10^{-3}	1.65×10^{-12} (6.89)	1.29×10^{-3} (12.47)	2.10×10^{-3} (5.93)
Wilcoxon	1.61×10^{-7}	1.47×10^{-3}	-1.03×10^{-2}	1.76×10^{-12} (6.46)	1.35×10^{-3} (11.95)	2.22×10^{-3} (5.63)
vdW	1.58×10^{-7}	1.01×10^{-3}	-1.39×10^{-2}	2.46×10^{-12} (4.63)	1.89×10^{-3} (8.49)	3.15×10^{-3} (3.96)
n = 5000						
QMLE	3.66×10^{-7}	1.20×10^{-2}	-2.07×10^{-2}	8.21×10^{-12}	1.20×10^{-2}	8.22×10^{-3}
Sign	6.95×10^{-11}	-2.00×10^{-3}	-3.86×10^{-3}	1.01×10^{-12} (8.09)	7.21×10^{-4} (16.67)	1.16×10^{-3} (7.10)
Wilcoxon	-3.01×10^{-10}	-1.81×10^{-3}	-3.98×10^{-3}	1.06×10^{-12} (7.73)	7.56×10^{-4} (15.90)	1.20×10^{-3} (6.85)
vdW	-1.57×10^{-8}	-2.37×10^{-3}	-5.86×10^{-3}	1.54×10^{-12} (5.33)	1.13×10^{-3} (10.64)	1.77×10^{-3} (4.64)

Zakoïan (2009) that higher order GARCH models may fit some financial time series better than the GARCH$(1,1)$ model. Therefore, here we examine the performance of the R-estimators under the GARCH$(2,1)$ model by running simulations with $R = 500, n = 1000$. For choosing the true model parameters for the simulations, we fitted the FTSE 100 data from January 2007 to December 2009 to the by GARCH$(2,1)$ model using the `fGarch` package. It turned out that α_2 is significant with p-value $= 0.019$ and the Akaike information criterion (AIC) of the GARCH$(2,1)$ is smaller than that of the GARCH$(1,1)$. Since the `fGarch` estimate of the true parameter is $\boldsymbol{\theta}_0 = (4.46 \times 10^{-6}, 0.0525, 0.108, 0.832)'$, we choose this $\boldsymbol{\theta}_0$ to generate sample from the GARCH$(2,1)$ model with various error distributions. The R-estimators and QMLE are compared through the standardized bias and MSE and the corresponding estimates are reported in Table 4.4. Similar to the GARCH$(1,1)$ case, the advantage of the R-estimators over the QMLE becomes prominent under heavy-tailed distributions, especially under the $t(3)$ distribution, where the bias and MSE of the R-estimators have smaller order of magnitude than the those of the QMLE.

Table 4.4: The estimates of the standardized bias and MSE of the R-estimators (sign, Wilcoxon and vdW scores) and the QMLE for the GARCH$(2,1)$ model under various error distributions (sample size $n = 1000$; $R = 500$ replications).

	Standardized bias				**Standardized MSE**			
	ω	α_1	α_2	β	ω	α_1	α_2	β
Normal								
QMLE	3.80×10^{-6}	8.85×10^{-3}	-3.16×10^{-3}	-2.01×10^{-2}	2.50×10^{-11}	1.71×10^{-3}	1.93×10^{-3}	1.35×10^{-3}
Sign	3.79×10^{-6}	1.05×10^{-2}	-5.76×10^{-3}	-1.84×10^{-2}	2.65×10^{-11}	1.90×10^{-3}	2.19×10^{-3}	1.30×10^{-3}
Wilcoxon	3.68×10^{-6}	9.91×10^{-3}	-6.51×10^{-3}	-1.81×10^{-2}	2.42×10^{-11}	1.74×10^{-3}	2.01×10^{-3}	1.20×10^{-3}
vdW	3.95×10^{-6}	1.03×10^{-2}	-7.49×10^{-3}	-1.96×10^{-2}	2.67×10^{-11}	1.74×10^{-3}	1.94×10^{-3}	1.25×10^{-3}
DE								
QMLE	2.61×10^{-6}	4.43×10^{-3}	2.14×10^{-3}	-1.99×10^{-2}	3.11×10^{-11}	2.53×10^{-3}	4.01×10^{-3}	2.33×10^{-3}
Sign	1.96×10^{-6}	5.02×10^{-3}	2.57×10^{-4}	-1.61×10^{-2}	9.96×10^{-12}	1.85×10^{-3}	2.88×10^{-3}	1.66×10^{-3}
Wilcoxon	1.85×10^{-6}	3.03×10^{-3}	-2.03×10^{-3}	-1.65×10^{-2}	9.57×10^{-12}	1.79×10^{-3}	2.85×10^{-3}	1.73×10^{-3}
vdW	1.95×10^{-6}	1.80×10^{-3}	-1.96×10^{-3}	-1.81×10^{-2}	1.10×10^{-11}	1.92×10^{-3}	3.14×10^{-3}	1.97×10^{-3}
Logistic								
QMLE	4.72×10^{-6}	5.44×10^{-3}	8.41×10^{-4}	-1.98×10^{-2}	5.24×10^{-11}	3.75×10^{-3}	4.49×10^{-3}	2.06×10^{-3}
Sign	3.17×10^{-6}	3.23×10^{-3}	-2.32×10^{-3}	-1.49×10^{-2}	2.09×10^{-11}	1.75×10^{-3}	2.50×10^{-3}	1.39×10^{-3}
Wilcoxon	3.24×10^{-6}	2.93×10^{-3}	-1.97×10^{-3}	-1.51×10^{-2}	2.20×10^{-11}	1.73×10^{-3}	2.48×10^{-3}	1.42×10^{-3}
vdW	3.62×10^{-6}	2.49×10^{-3}	-1.97×10^{-3}	-1.72×10^{-2}	2.76×10^{-11}	1.91×10^{-3}	2.67×10^{-3}	1.72×10^{-3}
***t*(3)**								
QMLE	1.78×10^{-6}	3.06×10^{-2}	-2.07×10^{-2}	-3.12×10^{-2}	2.85×10^{-11}	7.88×10^{-2}	7.65×10^{-2}	1.08×10^{-2}
Sign	9.92×10^{-7}	3.18×10^{-3}	-3.92×10^{-3}	-1.29×10^{-2}	5.67×10^{-12}	3.25×10^{-3}	5.25×10^{-3}	2.42×10^{-3}
Wilcoxon	9.78×10^{-7}	3.69×10^{-3}	-4.87×10^{-3}	-1.28×10^{-2}	5.70×10^{-12}	3.51×10^{-3}	5.58×10^{-3}	2.50×10^{-3}
vdW	9.86×10^{-7}	5.10×10^{-3}	-9.49×10^{-3}	-1.56×10^{-2}	7.59×10^{-12}	5.66×10^{-3}	8.08×10^{-3}	3.57×10^{-3}

4.4 Bootstrapping the R-estimators

Since the asymptotic covariance matrix of the R-estimators are of complicated forms, in this section we employ the weighted bootstrap technique discussed by Chatterjee and Bose (2005) in the context of M-estimators to approximate the distributions of the R-estimators and we compute corresponding coverage probabilities to exhibit the effectiveness of such bootstrap approximations. The weighted bootstrap in this context is attractive for its computational simplicity since at each bootstrap replication, only the weights need to be generated instead of resampling the data components to compute the replicates of the bootstrapped R-estimate.

In this context, the weighted bootstrap version of the rank-based central sequence is

$$\hat{\boldsymbol{R}}^{*}_{n,\varphi}(\boldsymbol{\theta}) := \hat{\boldsymbol{R}}^{*}_{n}(\boldsymbol{\theta}) = n^{-1/2}\sum_{t=1}^{n} w_{nt}\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})}{\hat{v}_t(\boldsymbol{\theta})}\left\{1-\varphi\left[\frac{\hat{R}_{nt}(\boldsymbol{\theta})}{n+1}\right]\frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta})}\right\},$$

where $\{w_{nt}; 1 \leq t \leq n; n \geq 1\}$ is a triangular array of r.v.'s which satisfies the following conditions:

(i) The weights $\{w_{nt}; 1 \leq t \leq n\}$ are exchangeable and independent of the data $\{X_t; 1 \leq t \leq n\}$ and errors $\{\epsilon_t; 1 \leq t \leq n\}$;

(ii) For all $t \geq 1, w_{nt} \geq 0; \mathrm{E}(w_{nt}) = 1; \mathrm{Corr}(w_{n1}; w_{n2}) = O(1/n); \mathrm{Var}(w_{nt}) = \sigma_n^2$, where $0 < c_1 < \sigma_n^2 = o(n)$, with $c_1 > 0$ being a constant.

Among various schemes of the weights satisfying the above conditions, we compare the following three types of weights:

(i) Scheme M: $\{w_{n1}, \ldots, w_{nn}\}$ have a multinomial $(n, 1/n, \ldots, 1/n)$ distribution, which is essentially the classical paired bootstrap.

(ii) Scheme E: $w_{nt} = (nE_t)/\sum_{i=1}^n E_i$, where $\{E_t\}$ are i.i.d. exponential r.v.'s with mean 1.

(iii) Scheme U: $w_{nt} = (nU_t)/\sum_{i=1}^n U_i$, where $\{U_t\}$ are i.i.d. uniform r.v.'s on $(0.5, 1.5)$.

We use the weighted version of (4.2.10) to compute the bootstrap estimator $\hat{\boldsymbol{\theta}}_{*n}$:

$$\hat{\boldsymbol{\theta}}_{*(r+1)} = \hat{\boldsymbol{\theta}}_{*(r)} - \left[\sum_{t=1}^n w_{nt} \frac{\dot{\hat{\boldsymbol{v}}}_t(\hat{\boldsymbol{\theta}}_{*(r)})\dot{\hat{\boldsymbol{v}}}_t'(\hat{\boldsymbol{\theta}}_{*(r)})}{\hat{v}_t^2(\hat{\boldsymbol{\theta}}_{*(r)})}\right]^{-1} \times \left\{\sum_{t=1}^n w_{nt} \frac{\dot{\hat{\boldsymbol{v}}}_t(\hat{\boldsymbol{\theta}}_{*(r)})}{\hat{v}_t(\hat{\boldsymbol{\theta}}_{*(r)})}\left[1 - \varphi\left(\frac{\hat{R}_{nt}(\hat{\boldsymbol{\theta}}_{*(r)})}{n+1}\right)\frac{X_t}{\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_{*(r)})}\right]\right\}, \quad \text{for} \quad r = 0, 1, ..., \tag{4.4.1}$$

with $\hat{\boldsymbol{\theta}}_{*(0)} = \hat{\boldsymbol{\theta}}_n$ being the initial estimator.

4.4.1 Bootstrap coverage probabilities

Chatterjee and Bose (2005) proved the consistency of the bootstrap for an estimator defined by smooth estimating equation. Since ranks are integer-valued discontinuous functions, the proof of the asymptotic validity of the bootstrapped R-estimator is a mathematically challenging problem which is beyond the scope

of this thesis. Instead, we resort to simulations to evaluate the performance of the bootstrap approximation of the R-estimators by comparing the distribution of $\sigma_n^{-1}\sqrt{n}(\hat{\boldsymbol{\theta}}_{*n} - \hat{\boldsymbol{\theta}}_n)$ with that of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi})$ in terms of coverage rates.

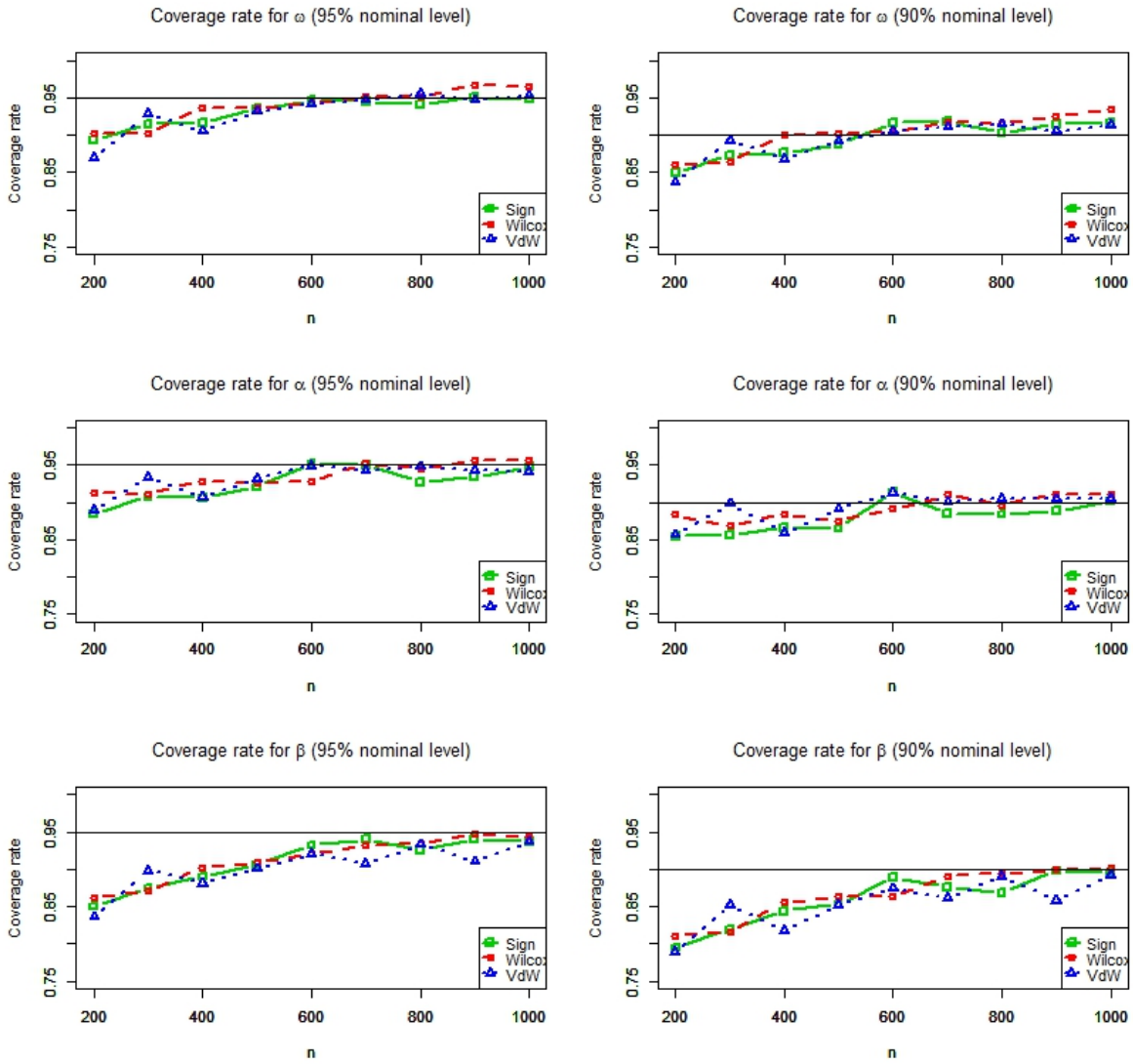
In particular, with the choice of the true parameter $\boldsymbol{\theta}_0 = (6.50\times10^{-6}, 0.177, 0.716)'$ as in the simulation study of the GARCH$(1,1)$ model of the previous section, we generate $R = 1000$ data each with sample size $n = 1000$ based on different error distributions. For each data, the exchangeable weights $\{w_{nt}; 1 \le t \le n\}$ are generated $B = 2000$ times. We consider cases where the error distributions are normal, DE, logistic and $t(3)$. The bootstrap weights are based on Schemes M, E and U. The bootstrap coverage rates (in percentage) for $95\%, 90\%$ nominal levels are reported in Table 4.5. Notice that all bootstrap schemes provide reasonable coverage rates under these error distributions. Scheme U is slightly better than the scheme M and E under the DE and $t(3)$ distributions.

To check the performance of the bootstrap under different sample sizes, we run simulation with $n = 200, 300, ..., 1000$ for the sign, Wilcoxon and vdW scores. There are $R = 1000$ replications being generated under the normal error distribution, and each replication is bootstrapped $B = 2000$ times with the scheme U. Figure 4.4 shows the bootstrap coverage rates for ω (first row), α (second row), β (third row) under 95% nominal level (left column) and 90% nominal level (right column). We notice that as the sample size increases, the coverage rates get close to the nominal levels for all parameters and all R-estimators, with only few exceptions. This tends to imply the consistency of the bootstrap approximation. With the sample size $n \ge 500$, the bootstrap coverage rates are generally close to the nominal levels.

Table 4.5: The bootstrap coverage rates (in percentage) for the R-estimators (sign, Wilcoxon and vdW) under various error distributions

			95% nominal level			90% nominal level		
			ω	α	β	ω	α	β
Normal	sign	Scheme M	94.1	93.6	92.8	90.8	88.5	89.2
		Scheme E	93.5	93.3	92.8	90.3	88.1	88.8
		Scheme U	94.8	94.7	93.7	91.7	90.2	89.6
Normal	Wilcoxon	Scheme M	96.4	96.3	93.7	93.8	90.3	89.7
		Scheme E	96.5	96.0	93.2	93.2	90.0	89.3
		Scheme U	96.5	95.6	94.3	93.5	91.1	90.1
Normal	vdW	Scheme M	94.3	92.3	93.6	91.3	89.1	89.0
		Scheme E	94.2	92.2	92.7	90.5	88.5	88.7
		Scheme U	95.3	94.1	93.7	91.4	90.5	89.2
DE	sign	Scheme M	90.8	90.5	91.6	87.8	86.0	86.8
		Scheme E	90.4	89.6	90.4	87.2	85.3	86.4
		Scheme U	91.9	92.7	92.7	88.8	88.9	87.8
DE	Wilcoxon	Scheme M	91.0	91.0	91.5	87.6	86.9	87.6
		Scheme E	90.7	90.2	90.4	87.2	86.2	86.6
		Scheme U	92.4	93.6	92.8	88.7	89.1	87.7
DE	vdW	Scheme M	90.9	87.6	89.7	87.5	83.9	85.3
		Scheme E	90.4	86.9	88.9	86.9	83.1	84.8
		Scheme U	92.4	90.2	91.0	89.7	85.6	86.0
Logistic	sign	Scheme M	93.0	91.1	92.1	89.0	87.6	88.6
		Scheme E	93.4	92.3	92.5	89.8	86.3	88.4
		Scheme U	93.0	92.3	91.9	88.7	87.5	87.1
Logistic	Wilcoxon	Scheme M	93.5	91.3	92.5	89.9	87.7	89.2
		Scheme E	93.7	89.4	91.7	90.0	85.8	87.1
		Scheme U	94.1	92.2	92.8	88.9	88.1	86.4
Logistic	vdW	Scheme M	93.1	91.2	92.3	89.3	88.0	87.0
		Scheme E	92.4	91.1	91.7	88.5	87.6	86.4
		Scheme U	94.4	93.6	92.2	90.4	90.8	86.8
$t(3)$	sign	Scheme M	88.3	85.3	88.3	86.0	82.6	83.5
		Scheme E	88.3	85.0	87.6	84.9	82.4	82.0
		Scheme U	91.8	89.0	90.6	87.5	85.6	86.4
$t(3)$	Wilcoxon	Scheme M	88.1	84.7	88.5	85.7	81.4	83.7
		Scheme E	88.0	84.5	87.7	85.0	80.7	82.8
		Scheme U	91.8	88.7	90.0	87.6	85.6	85.5
$t(3)$	vdW	Scheme M	85.6	82.3	86.1	81.9	79.7	81.0
		Scheme E	84.4	82.1	86.0	80.9	78.7	80.2
		Scheme U	90.3	85.4	88.9	86.6	81.0	83.3

Figure 4.4: Plot of the bootstrap coverage rates for the R-estimators (sign, Wilcoxon and vdW) at different sample sizes. The first, second and third rows are for ω, α and β respectively. The nominal levels are 95% (left column) and 90% (right column). Scheme U is employed and the errors have normal distribution.



4.5 Application of the R-estimator to the GJR model

Glosten et al. (1993) proposed the GJR(p, q) model for the asymmetric voatility observed in many financial dataset exhibit asymmetry property. The GJR(p, q) model is defined by

$$X_t = \sigma_t \epsilon_t$$

where

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \left[\alpha_{0i} + \gamma_{0i} I(X_{t-i} < 0)\right] X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2, \quad t \in \mathbb{Z},$$

with $\omega_0, \alpha_{0i}, \gamma_{0i}, \beta_{0j} > 0, \forall i, j$. Since σ_t^2 is linear in parameters, we define the R-estimators for the GJR model using the same rank-based central sequence as in (4.2.2). See also Iqbal and Mukherjee (2010) for the extension of M-estimators from the GARCH model to the GJR model. We do not prove any asymptotic theory for the R-estimators of the GJR model but present here empirical analysis using the same algorithm as in (4.2.10) to compute the R-estimators. The following extensive simulation study, similar to the GARCH case, demonstrates the superior performance of the R-estimators compared to the QMLE that is often used in this model. We also carry out simulation with increasing sample sizes to show the consistency of the R-estimators. Three types of R-estimators and the QMLE are compared below under various error distributions. We run simulations with the sample size $n = 1000$, number of replications $R = 500$ and true parameter

$$\boldsymbol{\theta}_0 = (3.45 \times 10^{-4}, 0.0658, 0.0843, 0.8182)',$$

which is motivated by the estimate in Tsay (2010) for the IBM stock monthly returns from 1926 to 2003. The estimates of the standardized bias and MSE of the QMLE and R-estimators and those of the ARE of the R-estimators wrt the

QMLE are reported in Table 4.6.

We remark that the results are consistent with those in Table 4.2: the vdW score still dominates the QMLE uniformly; the optimal scores under the DE and logistic distributions are also the sign and Wilcoxon, respectively. It is worth noting that under $t(3)$ distribution, the R-estimators are much more efficient than the QMLE for the parameter γ.

Table 4.6: The estimates of the standardized bias, MSE and ARE of the R-estimators (sign, Wilcoxon and vdW) and the QMLE for the GJR (1, 1) model under various error distributions (sample size $n = 1000$; $R = 500$ replications).

	Standardized bias				**Standardized MSE and ARE**			
	ω	α	γ	β	ω	α	γ	β
Normal								
QMLE	8.70×10^{-5}	-2.03×10^{-3}	7.47×10^{-3}	-2.03×10^{-2}	4.17×10^{-8}	7.63×10^{-4}	1.53×10^{-3}	3.80×10^{-3}
Sign	9.01×10^{-5}	-1.20×10^{-3}	8.43×10^{-3}	-2.00×10^{-2}	4.73×10^{-8}	9.40×10^{-4}	1.92×10^{-3}	4.30×10^{-3}
					(0.88)	(0.81)	(0.80)	(0.88)
Wilcoxon	9.35×10^{-5}	-8.15×10^{-4}	8.65×10^{-3}	-2.00×10^{-2}	4.96×10^{-8}	8.72×10^{-4}	1.76×10^{-3}	4.24×10^{-3}
					(0.84)	(0.87)	(0.87)	(0.90)
vdW	8.72×10^{-5}	-1.61×10^{-3}	7.59×10^{-3}	-2.01×10^{-2}	4.24×10^{-8}	7.72×10^{-4}	1.56×10^{-3}	3.82×10^{-3}
					(0.98)	(0.99)	(0.98)	(0.99)
DE								
QMLE	7.07×10^{-5}	-1.28×10^{-3}	1.13×10^{-2}	-1.87×10^{-2}	4.66×10^{-8}	1.23×10^{-3}	3.19×10^{-3}	4.72×10^{-3}
Sign	4.91×10^{-5}	-2.26×10^{-3}	6.99×10^{-3}	-1.62×10^{-2}	3.44×10^{-8}	9.58×10^{-4}	2.36×10^{-3}	4.10×10^{-3}
					(1.35)	(1.28)	(1.35)	(1.15)
Wilcoxon	5.06×10^{-5}	-2.23×10^{-3}	7.25×10^{-3}	-1.61×10^{-2}	3.45×10^{-8}	9.74×10^{-4}	2.41×10^{-3}	4.06×10^{-3}
					(1.35)	(1.26)	(1.32)	(1.16)
vdW	4.98×10^{-5}	-3.35×10^{-3}	6.73×10^{-3}	-1.71×10^{-2}	3.54×10^{-8}	1.02×10^{-3}	2.50×10^{-3}	4.23×10^{-3}
					(1.32)	(1.20)	(1.27)	(1.12)
Logistic								
QMLE	8.01×10^{-5}	7.88×10^{-4}	8.85×10^{-3}	-1.62×10^{-2}	3.86×10^{-8}	1.07×10^{-3}	2.40×10^{-3}	3.46×10^{-3}
Sign	6.06×10^{-5}	-2.59×10^{-4}	5.54×10^{-3}	-1.49×10^{-2}	3.26×10^{-8}	8.97×10^{-4}	1.96×10^{-3}	3.36×10^{-3}
					(1.18)	(1.19)	(1.23)	(1.03)
Wilcoxon	5.97×10^{-5}	-5.22×10^{-4}	5.38×10^{-3}	-1.44×10^{-2}	3.07×10^{-8}	8.76×10^{-4}	1.93×10^{-3}	3.18×10^{-3}
					(1.26)	(1.22)	(1.24)	(1.09)
vdW	6.27×10^{-5}	-1.18×10^{-3}	5.18×10^{-3}	-1.56×10^{-2}	3.25×10^{-8}	9.26×10^{-4}	2.04×10^{-3}	3.35×10^{-3}
					(1.19)	(1.15)	(1.17)	(1.03)
$\boldsymbol{t(3)}$								
QMLE	9.91×10^{-5}	-1.00×10^{-3}	9.21×10^{-2}	-6.45×10^{-2}	1.14×10^{-7}	4.38×10^{-3}	1.18×10^{-1}	2.31×10^{-2}
Sign	5.68×10^{-5}	4.23×10^{-4}	2.71×10^{-2}	-2.77×10^{-2}	3.78×10^{-8}	1.35×10^{-3}	4.69×10^{-3}	6.33×10^{-3}
					(3.02)	(3.24)	(25.19)	(3.65)
Wilcoxon	5.69×10^{-5}	3.94×10^{-5}	2.74×10^{-2}	-2.85×10^{-2}	3.93×10^{-8}	1.40×10^{-3}	4.80×10^{-3}	6.62×10^{-3}
					(2.91)	(3.12)	(24.57)	(3.49)
vdW	6.12×10^{-5}	-1.61×10^{-3}	3.43×10^{-2}	-3.71×10^{-2}	5.13×10^{-8}	1.93×10^{-3}	7.45×10^{-3}	9.58×10^{-3}
					(2.23)	(2.27)	(15.84)	(2.41)

Simulation under different sample size. We next investigate the behaviour of R-estimators by carrying out simulations with different sample sizes. The number of replications and true parameter are the same as those used for Table 4.6

Table 4.7: The standardized bias, MSE of the R-estimators (sign, Wilcoxon and vdW) for the GJR (1, 1) model under normal error distributions with different sample sizes ($R = 500$ replications).

	Standardized bias				**Standardized MSE**			
	ω	α	γ	β	ω	α	γ	β
Sign								
$n = 500$	1.78×10^{-4}	2.42×10^{-3}	1.88×10^{-2}	-4.84×10^{-2}	1.43×10^{-7}	1.64×10^{-3}	3.72×10^{-3}	1.33×10^{-2}
$n = 1000$	9.01×10^{-5}	-1.20×10^{-3}	8.43×10^{-3}	-2.00×10^{-2}	4.73×10^{-8}	9.40×10^{-4}	1.92×10^{-3}	4.30×10^{-3}
$n = 3000$	2.76×10^{-5}	-6.82×10^{-4}	1.97×10^{-3}	-4.47×10^{-3}	8.75×10^{-9}	3.05×10^{-4}	6.11×10^{-4}	8.73×10^{-4}
$n = 5000$	2.07×10^{-5}	-4.43×10^{-4}	2.05×10^{-3}	-3.43×10^{-3}	4.57×10^{-9}	1.70×10^{-4}	3.68×10^{-4}	4.62×10^{-4}
Wilcoxon								
$n = 500$	1.77×10^{-4}	2.52×10^{-3}	1.88×10^{-2}	-4.75×10^{-2}	1.37×10^{-7}	1.52×10^{-3}	3.45×10^{-3}	1.26×10^{-2}
$n = 1000$	9.35×10^{-5}	-8.15×10^{-4}	8.65×10^{-3}	-2.00×10^{-2}	4.96×10^{-8}	8.72×10^{-4}	1.76×10^{-3}	4.24×10^{-3}
$n = 3000$	3.01×10^{-5}	-2.94×10^{-5}	2.68×10^{-3}	-4.30×10^{-3}	8.18×10^{-9}	2.82×10^{-4}	5.60×10^{-4}	7.85×10^{-4}
$n = 5000$	2.45×10^{-5}	1.52×10^{-4}	2.82×10^{-3}	-3.54×10^{-3}	4.50×10^{-9}	1.63×10^{-4}	3.55×10^{-4}	4.29×10^{-4}
vdW								
$n = 500$	1.67×10^{-4}	1.42×10^{-3}	1.61×10^{-2}	-4.84×10^{-2}	1.31×10^{-7}	1.44×10^{-3}	3.03×10^{-3}	1.27×10^{-2}
$n = 1000$	8.72×10^{-5}	-1.61×10^{-3}	7.59×10^{-3}	-2.01×10^{-2}	4.24×10^{-8}	7.72×10^{-4}	1.56×10^{-3}	3.82×10^{-3}
$n = 3000$	2.88×10^{-5}	-1.72×10^{-4}	1.60×10^{-3}	-4.59×10^{-3}	7.42×10^{-9}	2.61×10^{-4}	4.90×10^{-4}	7.19×10^{-4}
$n = 5000$	2.42×10^{-5}	9.50×10^{-6}	2.42×10^{-3}	-4.02×10^{-3}	4.38×10^{-9}	1.49×10^{-4}	3.28×10^{-4}	4.26×10^{-4}

and the error distribution is normal. The estimates of the standardized bias and MSE of the R-estimators for the GJR$(1, 1)$ model are shown in Table 4.7. In general, for all R-estimators, both the bias and MSE decrease when the sample size increases from $n = 500$ to $n = 5000$. This tends to reflect that the R-estimators are consistent estimators of $\boldsymbol{\theta}_{0\varphi}$ for the GJR$(1, 1)$ model.

4.6 Conclusion

We propose a new class of R-estimators for the GARCH model and derive the asymptotic normality of these estimators under mild moment and smoothness conditions on the error distribution. We exhibit the robustness and efficiency of R-estimators with respect to the QMLE through simulation and real data analysis. We also consider a general type of weighted bootstrap for the R-estimators which is computational-friendly and easy-to-implement. The theoretical analysis such as the asymptotic validity of the weighted bootstrap is an interesting but challenging problem that can be explored in the future.

4.7 Proofs of Proposition 4.2.1 and Theorem 4.2.1

We will use the following facts from Berkes et al. (2003) for the proofs:

Fact 1. For any $\nu > 0$,

$$\mathrm{E}\left\{\sup\left[\left|\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta})}{v_1(\boldsymbol{\theta})}\right|^{\nu}; \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\right]\right\} < \infty. \tag{4.7.1}$$

and

$$\mathrm{E}\left\{\sup\left[\left|\frac{\ddot{\boldsymbol{v}}_1(\boldsymbol{\theta})}{v_1(\boldsymbol{\theta})}\right|^{\nu}; \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\right]\right\} < \infty.$$

Fact 2. There exist random variables Z_0, Z_1 and Z_2, all independent of $\{\epsilon_t; 1 \leq t \leq n\}$ and a number $0 < \rho < 1$, such that

$$0 < v_t(\boldsymbol{\theta}) - \hat{v}_t(\boldsymbol{\theta}) \leq \rho^t Z_0, \tag{4.7.2}$$

$$|\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}) - \dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})| \leq \rho^t Z_1, \tag{4.7.3}$$

$$|\ddot{\boldsymbol{v}}_t(\boldsymbol{\theta}) - \ddot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta})| \leq \rho^t Z_2.$$

Fact 3. Let $\{(A_t, B_t, C_t); t \geq 0\}$ be a sequence of identically distributed random variables. If $\mathrm{E}\log^+ A_0 + \mathrm{E}\log^+ B_0 + \mathrm{E}\log^+ C_0 < \infty$, then for any $|r| < 1$,

$$\sum_{t=0}^{\infty}(A_t + B_t C_t) r^t \quad \text{converges with probability 1.} \tag{4.7.4}$$

Idea of the proof of Theorem 4.2.1. We first derive the following Theorem 4.7.1, Corollary 4.7.1.1 and Theorem 4.7.2 on empirical processes where a scale-perturbed weighted mixed-empirical process is approximated by its non-perturbed version. With $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{0\varphi} + n^{-1/2}\boldsymbol{b}$, we derive asymptotic expansion of the difference between two quantities $\boldsymbol{T}_{1n}(\boldsymbol{\theta}_n)$ and $\boldsymbol{T}_{2n}(\boldsymbol{\theta}_n)$ which are defined later. We then show that $\boldsymbol{T}_{1n}(\boldsymbol{\theta}_n)$ can be approximated by a r.v., which is asymptotic normal, plus a term linear in $\boldsymbol{b}$. Also, we use $\boldsymbol{T}_{2n}(\boldsymbol{\theta}_n)$ to approximate $\boldsymbol{R}_n(\boldsymbol{\theta}_n)$ and show that asymptotically their difference is a r.v. with mean zero. Finally, we prove that the difference of $\boldsymbol{R}_n(\boldsymbol{\theta}_n)$ and $\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_n)$ converges in probability to zero. Using

these results, we are able to derive the asymptotic linearity of $\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_n)$ as shown in Proposition 4.2.1. Finally, using the definition of the one-step R-estimator in (5.4.8), we are able to derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$.

The following results state the uniform approximation of a scale-perturbed weighted mixed-empirical process by its non-perturbed version which is used for the derivation of the asymptotic distributions of the R-estimators.

Theorem 4.7.1, Corollary 4.7.1.1 and Theorem 4.7.2.

Let $\{(\eta_t,\, \gamma_{nt}, \delta_{nt}),\, 1 \le t \le n\}$ be an array of 3-tuple r.v.'s defined on a probability space such that $\{\eta_t,\, 1 \le t \le n\}$ are i.i.d. with c.d.f. G and η_t is independent of $(\gamma_{nt}, \delta_{nt})$ for each $1 \le t \le n$. Let $\{\mathcal{A}_{nt}; 1 \le t \le n\}$ be an array of increasing sub-σ-fields in both n and t so that $\mathcal{A}_{nt} \subset \mathcal{A}_{n(t+1)}$, $\mathcal{A}_{nt} \subset \mathcal{A}_{(n+1)t}$, $1 \le t \le n-1$, $n \ge 2$. Assume also that $(\gamma_{n1}, \delta_{n1})$ is $\mathcal{A}_{n1}$ measurable, and $\{\{(\gamma_{nt}, \delta_{nt}); 1 \le t \le j\}, \eta_1, \eta_2, \ldots, \eta_{j-1}\}$ are $\mathcal{A}_{nj}$ measurable, $2 \le j \le n$. For $x \in \mathbb{R}$, recall that $\mu(x) = \mathrm{E}[\eta I(\eta < x)] = \int_{-\infty}^{x} sg(s)ds$ and consider the following weighted mixed-empirical processes

$$\begin{aligned}
&\tilde{V}_n(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\eta_t I(\eta_t < x + x\delta_{nt}), \qquad (4.7.5)\\
&\tilde{J}_n(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\mu(x + x\delta_{nt}),\\
&V_n^*(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\eta_t I(\eta_t \le x), \qquad J_n^*(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\mu(x),\\
&\tilde{U}_n(x) := \tilde{V}_n(x) - \tilde{J}_n(x), \quad U_n^*(x) := V_n^*(x) - J_n^*(x).
\end{aligned}$$

Assume the following conditions on the weights $\{\gamma_{nt}\}$ and perturbations $\{\delta_{nt}\}$.

Let $C_n := \sum_{t=1}^{n} E|\gamma_{nt}|^q$ for some $q > 2$. Let a with $0 < a < q/2$ be such that

$$C_n/n^{q/2-a} = o(1). \tag{4.7.6}$$

$$\left(n^{-1}\sum_{t=1}^{n}\gamma_{nt}^2\right)^{1/2} = \gamma + o_{\mathrm{P}}(1) \text{ for a positive r.v.} \gamma. \tag{4.7.7}$$

$$\mathrm{E}\left(n^{-1}\sum_{t=1}^{n}\gamma_{nt}^2\right)^{q/2} = O(1). \tag{4.7.8}$$

$$\max_{1\le t\le n} n^{-1/2}|\gamma_{nt}| = o_{\mathrm{P}}(1). \tag{4.7.9}$$

$$\max_{1\le t\le n} |\delta_{nt}| = o_{\mathrm{P}}(1). \tag{4.7.10}$$

$$\frac{n^{q/2-\epsilon}}{C_n}\mathrm{E}\left[n^{-1}\sum_{t=1}^{n}\{\gamma_{nt}^2|\delta_{nt}|\}\right]^{q/2} = o(1). \tag{4.7.11}$$

$$n^{-1/2}\sum_{t=1}^{n}|\gamma_{nt}\delta_{nt}| = O_{\mathrm{p}}(1). \tag{4.7.12}$$

The following theorem shows that *uniformly* over the entire real line, the perturbed process $\tilde{U}_n$ can be approximated by U_n^*.

Theorem 4.7.1. *Under the above set-up and Assumptions (4.7.6)-(4.7.12) and (A1),*

$$\sup_{x\in\mathbb{R}}|\tilde{U}_n(x) - U_n^*(x)| \;=\; o_{\mathrm{P}}(1). \tag{4.7.13}$$

Proof. The proof is similar to the proof in Mukherjee (2007, Theorem 6.1). In particular, we show point-wise convergence for each x and then invoke the monotone structure of the mean processes to achieve the uniform convergence. For weighted empirical, the monotonically increasing mean process is given by the distribution function. Although μ in the present case is not a monotone function on $(-\infty, \infty)$, we use its monotone property separately on $(-\infty, 0]$ and $[0, \infty)$. □

We remark that this theorem is different from Koul and Ossiander (1994, Theorem 1.1) and Mukherjee (2007, Theorem 6.1) where weighted empirical processes were considered for the estimation of the mean parameters. For the estimation of the scale parameters, in this chapter we consider weighted mixed-empirical process

which is a weighted sum of the mixture of error and its indicator process.

The following corollary describes a Taylor-type expansion of the weighted sum of indicator functions $\tilde{V}_n(x)$.

Corollary 4.7.1.1. *Under the above setup and under the Assumptions (4.7.6)-(4.7.12) and (A1),*

$$\sup_{x\in\mathbb{R}} |\tilde{J}_n(x) - J_n^*(x) - x^2 g(x) n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\delta_{nt}| = o_{\mathrm{P}}(1). \tag{4.7.14}$$

Hence,

$$\sup_{x\in\mathbb{R}} |\tilde{V}_n(x) - V_n^*(x) - x^2 g(x) n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\delta_{nt}| = o_{\mathrm{P}}(1). \tag{4.7.15}$$

Proof. Here (4.7.15) follows from (4.7.14) and (4.7.13). Therefore, it remains to prove (4.7.14). Notice that the LHS of (4.7.14) equals

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \left| n^{-1/2} \sum_{t=1}^{n} \gamma_{nt} \left[x \int_x^{x+x\delta_{nt}} sg(s)ds - x^2 g(x)\delta_{nt} \right] \right| \\
=&\sup_{x\in\mathbb{R}} \left| n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\delta_{nt} \left[x \int_0^{1} (x + hx\delta_{nt}) g(x + hx\delta_{nt}) dh - x^2 g(x) \right] \right| \\
=&o_{\mathrm{P}}(1)
\end{aligned}$$

due to (4.7.12) and Assumption (A1). □

The next theorem provides an extended version of (4.7.13) when the weights are functions on appropriately scaled parameter space. We define the following processes of two arguments as follows.

Probabilistic framework: Let $\{\eta_t,\, 1 \le t \le n\}$ be i.i.d. with the c.d.f. G, $\{l_{nt}; 1 \le t \le n\}$ be an array of measurable functions from $\mathbb{R}^m$ to $\mathbb{R}$ such that for every $\boldsymbol{b} \in \mathbb{R}^m$ and $1 \le t \le n$, $(l_{nt}(\boldsymbol{b}),\, u_{nt}(\boldsymbol{b}))$ are independent of η_t. For $x \in \mathbb{R}$ and

$\boldsymbol{b} \in \mathbb{R}^m$, let

$$\tilde{\mathcal{V}}(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b}) \eta_t I\Big(\eta_t < x + x u_{nt}(\boldsymbol{b})\Big),$$
$$\tilde{\mathcal{J}}(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b})\, \mu\Big(x + x u_{nt}(\boldsymbol{b})\Big),$$
$$\tilde{\mathcal{U}}(x, \boldsymbol{b}) := \tilde{\mathcal{V}}(x, \boldsymbol{b}) - \tilde{\mathcal{J}}(x, \boldsymbol{b}),$$
$$\mathcal{V}^*(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b}) \eta_t I(\eta_t < x),\ \mathcal{J}^*(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b})\, \mu(x),$$
$$\mathcal{U}^*(x, \boldsymbol{b}) := \mathcal{V}^*(x, \boldsymbol{b}) - \mathcal{J}^*(x, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b}) \Big[\eta_t I(\eta_t < x) - \mu(x)\Big].$$

Here $\mathcal{U}^*(\cdot, \cdot)$ is a sequence of ordinary non-perturbed weighted mixed-empirical processes with weights $\{l_{nt}(\cdot)\}$ and $\tilde{\mathcal{U}}(\cdot, \cdot)$ is a sequence of perturbed weighted mixed-empirical processes with scale perturbations $\{u_{nt}(\cdot)\}$. In Theorem 4.7.2 below it is shown that $\tilde{\mathcal{U}}$ can be uniformly approximated by $\mathcal{U}^*$ under the following conditions (4.7.16)-(4.7.24) for $\{l_{nt}(\cdot)\}$ and $\{u_{nt}(\cdot)\}$. Note that the statements on assumptions and convergence hold point-wise for each fixed $\boldsymbol{b} \in \mathbb{R}^m$.

There exist numbers $q > 2$ and a (both free from $\boldsymbol{b}$) satisfying $0 < a < q/2$ such that with $C_n(\boldsymbol{b}) := \sum_{t=1}^{n} \mathrm{E}|l_{ni}(\boldsymbol{b})|^q$,

$$C_n(\boldsymbol{b})/n^{q/2-a} = o(1), \text{ for each } \boldsymbol{b} \in \mathbb{R}^m. \tag{4.7.16}$$

For some positive random process $\ell(\boldsymbol{b})$,

$$\left(n^{-1}\sum_{t=1}^{n} l_{nt}^2(\boldsymbol{b})\right)^{1/2} = \ell(\boldsymbol{b}) + o_{\mathrm{P}}(1), \qquad \boldsymbol{b} \in \mathbb{R}^m. \tag{4.7.17}$$

$$\mathrm{E}\left(n^{-1}\sum_{t=1}^{n} l_{ni}^2(\boldsymbol{b})\right)^{q/2} = O(1), \qquad \boldsymbol{b} \in \mathbb{R}^m. \tag{4.7.18}$$

$$\max_{1\le t\le n} n^{-1/2}|l_{nt}(\boldsymbol{b})| = o_{\mathrm{P}}(1), \qquad \boldsymbol{b} \in \mathbb{R}^m. \tag{4.7.19}$$

$$\max_{1\le t\le n}\{|u_{nt}(\boldsymbol{b})|\} = o_{\mathrm{P}}(1), \qquad \boldsymbol{b} \in \mathbb{R}^m. \tag{4.7.20}$$

$$\frac{n^{q/2-a}}{C_n(\boldsymbol{b})}\mathrm{E}\left[n^{-1}\sum_{t=1}^{n} l_{nt}^2(\boldsymbol{b})|u_{nt}(\boldsymbol{b})|\right]^{q/2} = o(1), \qquad \boldsymbol{b} \in \mathbb{R}^m. \tag{4.7.21}$$

$$n^{-1/2}\sum_{t=1}^{n} l_{nt}(\boldsymbol{b})u_{nt}(\boldsymbol{b}) = O_{\mathrm{p}}(1), \quad \boldsymbol{b} \in \mathbb{R}^m. \tag{4.7.22}$$

$$\forall\ b, \varepsilon > 0,\ \exists\, \delta > 0,\ \text{and}\ n_1 \in \mathbb{N}\,\text{whenever}\,\|\boldsymbol{s}\| \le b,\,\text{and}\,n > n_1, \tag{4.7.23}$$

$$P\left(n^{-1/2}\sum_{t=1}^{n} |l_{nt}(\boldsymbol{s})|\left\{\sup_{\|\boldsymbol{t}-\boldsymbol{s}\|<\delta}|u_{nt}(\boldsymbol{t}) - u_{nt}(\boldsymbol{s})|\right\} \le \varepsilon\right) > 1-\varepsilon.$$

$$\forall\ b\,\text{and}\,\varepsilon > 0,\ \exists\ \delta > 0,\ \text{and}\ n_2 \in \mathbb{N}\,\text{whenever}\,\|\boldsymbol{s}\| \le b,\,\text{and}\,n > n_2, \tag{4.7.24}$$

$$P\left(\sup_{\|\boldsymbol{t}-\boldsymbol{s}\|\le\delta} n^{-1/2}\sum_{t=1}^{n} |l_{nt}(\boldsymbol{t}) - l_{nt}(\boldsymbol{s})| \le \varepsilon\right) > 1-\varepsilon.$$

Conditions (4.7.16)-(4.7.24) are regularity conditions on the weights and perturbations of the two-parameters empirical processes. Conditions (4.7.23)-(4.7.24) are smoothness conditions on the weights and perturbations. Under stationarity and ergodicity, many of these conditions reduce to much simpler conditions based on existence of the moments.

The following theorem generalizes (4.7.13) when the weights are functions of $\boldsymbol{b}$.

Theorem 4.7.2. *Under the above framework, suppose that conditions (4.7.16)-(4.7.24) and Assumption (A1) hold. Then for every* $0 < b < \infty$,

$$\sup_{x\in\mathbb{R},\|\boldsymbol{b}\|\le b} |\tilde{\mathcal{U}}(x,\boldsymbol{b}) - \mathcal{U}^*(x,\boldsymbol{b})| \;=\; o_{\mathrm{P}}(1). \tag{4.7.25}$$

Proof. Clearly, under conditions (4.7.16)-(4.7.22), Theorem 4.7.1 entails that for

each fixed $\boldsymbol{b}$,

$$\sup_{x\in\mathbb{R}} |\tilde{\mathcal{U}}(x,\boldsymbol{b}) - \mathcal{U}^*(x,\boldsymbol{b})| = o_{\mathrm{P}}(1).$$

The uniform convergence with respect to $\boldsymbol{b}$ over compact sets can be proved as in Mukherjee (2007, Lemma 3.2) using conditions (4.7.23) and (4.7.24). □

The following facts are useful in the proofs of various results of this chapter. Let $m = 1 + p + q$ be the total number of parameters and fix $\boldsymbol{b} \in \mathbb{R}^m$. Let $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{0\varphi} + n^{-1/2}\boldsymbol{b}$,

$$u_{nt}(\boldsymbol{b}) = \frac{v_t^{1/2}(\boldsymbol{\theta}_n)}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})} - 1,\ v_{nt}(\boldsymbol{b}) = \frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_n)} - 1. \tag{4.7.26}$$

Then $\{u_{nt}(\boldsymbol{b})\}$ satisfies (4.7.20) since

$$u_{nt}(\boldsymbol{b}) = \frac{v_t(\boldsymbol{\theta}_n) - v_t(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\{v_t^{1/2}(\boldsymbol{\theta}_n) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}} = \frac{n^{-1/2}\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}^*)\boldsymbol{b}}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\{v_t^{1/2}(\boldsymbol{\theta}_n) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}}, \tag{4.7.27}$$

for some $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*(n,t,\boldsymbol{b})$ in the neighbourhood of $\boldsymbol{\theta}_{0\varphi}$ for large n. The $n^{-1/2}$-factor is used later for deriving convergence of some sequence of random vectors. Similarly, for some $\boldsymbol{\theta}^*$,

$$v_{nt}(\boldsymbol{b}) = \frac{v_t(\boldsymbol{\theta}_{0\varphi}) - v_t(\boldsymbol{\theta}_n)}{v_t^{1/2}(\boldsymbol{\theta}_n)\{v_t^{1/2}(\boldsymbol{\theta}_n) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}} = \frac{-n^{-1/2}\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}^*)\boldsymbol{b}}{v_t^{1/2}(\boldsymbol{\theta}_n)\{v_t^{1/2}(\boldsymbol{\theta}_n) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}} = n^{-1/2}\xi_{nt}, \tag{4.7.28}$$

say. Let $a_{nt}(\boldsymbol{b}) = v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})/v_t^{1/2}(\boldsymbol{\theta}_n) = 1 + v_{nt}(\boldsymbol{b}) = 1 + n^{-1/2}\xi_{nt}$. Then

$$\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} = a_{nt}(\boldsymbol{b})\eta_t = \eta_t + n^{-1/2}\eta_t\xi_{nt} = \eta_t + n^{-1/2}z_{nt},$$

where

$$z_{nt} = \eta_t\xi_{nt} = \eta_t \times \frac{-\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}^*)\boldsymbol{b}}{v_t^{1/2}(\boldsymbol{\theta}_n)\{v_t^{1/2}(\boldsymbol{\theta}_n) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}}.$$

For $\delta > 0$ in Assumption (A1) and any $c > 0$,

$$P\left[n^{-1/2}\max_{1\le t\le n}|z_{nt}| > c\right] \le \sum_{t=1}^{n} P\left[n^{-1/2}|z_{nt}| > c\right] \le n\frac{\mathrm{E}\left[n^{-1-\delta/2}|\eta_t|^{2+\delta}|\xi_{nt}|^{2+\delta}\right]}{c^{2+\delta}} = o(1)$$

since all moments of $\{|\xi_{nt}|\}$ are finite and η_t and ξ_{nt} are independent for all t. Therefore

$$\max_{1\le t\le n}\left|\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}-\eta_t\right|=o_{\mathrm{P}}(1). \tag{4.7.29}$$

If $\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)/v_t(\boldsymbol{\theta}_n)$ appears as the coefficients, we replace it by $\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi})$ and the difference is controlled as follows. Notice that all derivatives below exist with bounded moments and so

$$\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}-\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}=n^{-1/2}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}+n^{-1}\boldsymbol{A}^*_{tn}, \tag{4.7.30}$$

where $\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})=\ddot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi})-\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_{0\varphi})/\{v_t(\boldsymbol{\theta}_{0\varphi})\}^2$. Only the term $n^{-1/2}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}$ is of our interest since others are of higher order than n.

Take $l_{nt}(\boldsymbol{b})$ to be equal to the j-th coordinate ($1\le j\le m=1+p+q$) of

$$\boldsymbol{L}_{nt}(\boldsymbol{b})=\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\times\frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_n)} \tag{4.7.31}$$

and $u_{nt}(\boldsymbol{b})$ as in (4.7.26). We now show that (4.7.17)-(4.7.24) hold with such choice.

For each t with $1\le t\le n$, $\{\boldsymbol{L}_{nt}(\boldsymbol{b}),\, u_{nt}(\boldsymbol{b})\}$ are independent of η_t. Using a Taylor expansion of $l_{nt}(\boldsymbol{b})$ at $\boldsymbol{\theta}_{0\varphi}$ for each $1\le t\le n$ and noting the existence of all moments of $v_t(\boldsymbol{\theta}_{0\varphi})$ and its derivatives of all higher orders, (4.7.17) and (4.7.18) hold. Existence of all higher moments of $\{l_{nt}(\boldsymbol{b}),\, u_{nt}(\boldsymbol{b})\}$ ensure conditions (4.7.19)-(4.7.21).

To verify (4.7.22), we use (4.7.27) and that for each t, $v_t(\cdot)$ is a smooth function with derivatives of all order to conclude that

$$n^{-1/2}\sum_{t=1}^{n}\boldsymbol{L}_{nt}(\boldsymbol{b})u_{nt}(\boldsymbol{b})=\mathrm{E}[\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})/v_1^2(\boldsymbol{\theta}_{0\varphi})](\boldsymbol{b}/2)+o_{\mathrm{P}}(1)=\boldsymbol{J}\boldsymbol{b}/2+o_{\mathrm{P}}(1).$$

Conditions (4.7.23) and (4.7.24) can be verified using the mean value theorem.

The following lemmas and their proofs represent the intermediate steps in the proofs of Proposition 4.2.1 and Theorem 4.2.1.

Lemma 4.7.3, Lemma 4.7.4, Lemma 4.7.5 and Lemma 4.7.6.

Let

$$\boldsymbol{T}_{n1}(\boldsymbol{\theta}_n) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} \left\{ 1 - \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \varphi[G(\eta_t)] \right\},$$

$$\boldsymbol{T}_{n2}(\boldsymbol{\theta}_n) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} \left\{ 1 - \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] \right\}$$

and note that the difference in the definitions of these two quantities lies only in the argument of $\varphi(G(.))$. We show in Lemma 4.7.3 below that $\int_0^1 [\tilde{\mathcal{V}}(u, \boldsymbol{b}) - \mathcal{V}^*(u, \boldsymbol{b})] d\varphi(u) = \boldsymbol{T}_{n1}(\boldsymbol{\theta}_n) - \boldsymbol{T}_{n2}(\boldsymbol{\theta}_n)$. Using results on empirical processes in Theorem 4.7.2, $\int_0^1 [\tilde{\mathcal{V}}(u, \boldsymbol{b}) - \mathcal{V}^*(u, \boldsymbol{b})] d\varphi(u)$ is linear in $\boldsymbol{b}$. Consequently, we obtain the following uniform approximations of $\boldsymbol{T}_{n1}(\boldsymbol{\theta}_n) - \boldsymbol{T}_{n2}(\boldsymbol{\theta}_n)$ over $||\boldsymbol{b}|| \leq c$ where $c > 0$.

Lemma 4.7.3. *Let Assumptions (A1)-(A3) hold. Then, as $n \to \infty$,*

$$\boldsymbol{T}_{n2}(\boldsymbol{\theta}_n) - \boldsymbol{T}_{n1}(\boldsymbol{\theta}_n) = \boldsymbol{M}\boldsymbol{b} + u_{\mathrm{P}}(1), \tag{4.7.32}$$

where $\boldsymbol{M} = \boldsymbol{J}\rho(\varphi)/2$.

Proof. To use Theorem 4.7.2 in the proof, let $\boldsymbol{b} = n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{0\varphi})$ and $x = G^{-1}(u)$ for some $0 < u < 1$. For simplicity, we use the notation $\tilde{\mathcal{V}}(u, \boldsymbol{b})$ to denote $\tilde{\mathcal{V}}(G^{-1}(u), \boldsymbol{b})$ which is defined in the probabilistic framework above. Accordingly

$$\tilde{\mathcal{V}}(u, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} I \left[\eta_t < G^{-1}(u) \frac{v_t^{1/2}(\boldsymbol{\theta}_n)}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})} \right]$$

and

$$\mathcal{V}^*(u, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} I \left(\eta_t < G^{-1}(u) \right).$$

With the choice based on (4.7.31) and (4.7.26) and using

$$\frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_n)} \eta_t = \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)},$$

$$\begin{aligned}\tilde{\mathcal{V}}(u,\boldsymbol{b}) &= n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}I\left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}<G^{-1}(u)\right]\\ &= n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}I\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\right)<u\right].\end{aligned}$$

Similarly,

$$\mathcal{V}^*(u,\boldsymbol{b}) = n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}I\left(G(\eta_t)<u\right).$$

Since

$$\int_0^1 I\left\{G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\right)<u\right\}d\varphi(u) = \varphi(1)-\varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\right)\right],$$

we get

$$\int_0^1\tilde{\mathcal{V}}(u,\boldsymbol{b})d\varphi(u) = n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\left\{\varphi(1)-\varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\right)\right]\right\}$$

and

$$\int_0^1\mathcal{V}^*(u,\boldsymbol{b})d\varphi(u) = n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\Big(\varphi(1)-\varphi(G(\eta_t)\Big).$$

Cancelling $\varphi(1)$, $\int_0^1[\tilde{\mathcal{V}}(u,\boldsymbol{b})-\mathcal{V}^*(u,\boldsymbol{b})]d\varphi(u)$ equals

$$\begin{aligned}&n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\left\{-\varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\right)\right]+\varphi(G(\eta_t)\Big)\right\}\\ &=\boldsymbol{T}_{n2}(\boldsymbol{\theta}_n)-\boldsymbol{T}_{n1}(\boldsymbol{\theta}_n).\end{aligned}$$

Using (4.7.14) and (4.7.27) with

$$\tilde{\mathcal{J}}(u,\boldsymbol{b}) := n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_n)}\mu\left[G^{-1}(u)\frac{v_t^{1/2}(\boldsymbol{\theta}_n)}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}\right]$$

$$\mathcal{J}^*(u,\boldsymbol{b}) := n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_n)}\mu\left(G^{-1}(u)\right),$$

we have

$$\sup_{u\in(0,1)} |\tilde{\mathcal{J}}(u,\boldsymbol{b})-\mathcal{J}^*(u,\boldsymbol{b})-[G^{-1}(u)]^2 g(G^{-1}(u))n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_n)}u_{nt}(\boldsymbol{b})| = u_{\mathrm{P}}(1).$$

Also,

$$|n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_n)}u_{nt}(\boldsymbol{b}) - \boldsymbol{J}| = u_{\mathrm{P}}(1).$$

Hence,

$$\begin{aligned}\int_0^1 [\tilde{\mathcal{J}}(u,\boldsymbol{b}) - \mathcal{J}^*(u,\boldsymbol{b})]d\varphi(u) &= \int_0^1 [G^{-1}(u)]^2 g(G^{-1}(u))d\varphi(u)\boldsymbol{J}\boldsymbol{b}/2 + u_{\mathrm{P}}(1)\\ &= \boldsymbol{M}\boldsymbol{b} + u_{\mathrm{P}}(1) \qquad (4.7.33)\end{aligned}$$

by recalling that $\boldsymbol{M} = \boldsymbol{J}\rho(\varphi)/2$. Finally, (4.7.32) follows from Theorem 4.7.2. □

The following lemma states that the difference between $\boldsymbol{T}_{n1}(\boldsymbol{\theta}_n)$ and $\boldsymbol{N}_n$ is asymptotically linear in $\boldsymbol{b}$.

Lemma 4.7.4. *Let Assumptions (A1)-(A3) hold. Then, as $n \to \infty$,*

$$\boldsymbol{T}_{n1}(\boldsymbol{\theta}_n) - \boldsymbol{N}_n = \boldsymbol{J}\boldsymbol{b}/2 + u_{\mathrm{P}}(1), \qquad (4.7.34)$$

where

$$\boldsymbol{N}_n \to \mathcal{N}(\boldsymbol{0}, \boldsymbol{J}\sigma^2(\varphi)), \qquad (4.7.35)$$

with $\sigma^2(\varphi) = \mathrm{Var}\{\eta_1\varphi[G(\eta_1)]\}$.

Proof. The difference between $\boldsymbol{T}_{n1}$ and $\boldsymbol{N}_n$ lies in comparing $X_t/v_t^{1/2}(\boldsymbol{\theta}_n) = \eta_t + n^{-1/2}z_{nt}$ and η_t and involves smooth function of $\boldsymbol{b}$. So the proof follows easily with the details below. Notice that

$$\begin{aligned}\boldsymbol{T}_{n1}(\boldsymbol{\theta}_n) - \boldsymbol{N}_n &= n^{-1/2}\sum_{t=1}^{n}\left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\right]\{1 - a_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)]\}\\ &- n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}v_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)] = \boldsymbol{F}_{n1} - \boldsymbol{F}_{n2}.\end{aligned}$$

Using (4.7.30),

$$\begin{aligned}\boldsymbol{F}_{n1} &= n^{-1}\sum_{t=1}^{n}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}\left\{1-a_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)]\right\}+u_{\mathrm{P}}(1)\\ &= n^{-1}\sum_{t=1}^{n}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}\left\{1-\eta_t\varphi[G(\eta_t)]\right\}-n^{-1}\sum_{t=1}^{n}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}v_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)]+u_{\mathrm{P}}(1).\end{aligned}$$

Using the LLN, the first term in the above decomposition of $\boldsymbol{F}_{n1}$ is $u_{\mathrm{P}}(1)$ since

$$\mathrm{E}\left\{\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}\left\{1-\eta_t\varphi[G(\eta_t)]\right\}\right\}=\mathrm{E}[\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}]\,\mathrm{E}\left\{1-\eta_t\varphi[G(\eta_t)]\right\}=\boldsymbol{0}.$$

For the second term, using (4.7.28) we have $n^{-1/2}$ factor of $v_{nt}(\boldsymbol{b})$ and consequently it is $u_{\mathrm{P}}(1)$.

For $\boldsymbol{F}_{n2}$, we approximate $v_{nt}(\boldsymbol{b})$ by $-n^{-1/2}\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}/\{2v_t(\boldsymbol{\theta}_{0\varphi})\}$ and use $\mathrm{E}\{\eta_t\varphi[G(\eta_t)]\}=1$ to obtain $\boldsymbol{F}_{n2}=-\boldsymbol{J}\boldsymbol{b}/2+u_{\mathrm{P}}(1)$. Hence (4.7.34) is proved.

Using the independence of v_t and η_t for each t, $\boldsymbol{N}_n$ is a sum of the vectors of martingale differences and so (4.7.35) follows from the martingale CLT. □

Now consider the rank-based counterpart of $\boldsymbol{T}_{n2}(\boldsymbol{\theta}_n)$

$$\boldsymbol{R}_n(\boldsymbol{\theta}_n)=n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\left\{1-\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\varphi\left(\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1}\right)\right\}.$$

The following lemma provides the difference between $\boldsymbol{T}_{n2}(\boldsymbol{\theta}_n)$ and $\boldsymbol{R}_n(\boldsymbol{\theta}_n)$. It shows that the effect of replacing observations in $\boldsymbol{T}_{2n}(\boldsymbol{\theta}_n)$ by ranks is asymptotically a r.v. with mean zero.

Lemma 4.7.5. *Let Assumptions (A1)-(A3) hold. Then, as $n\to\infty$,*

$$\boldsymbol{R}_n(\boldsymbol{\theta}_n)-\boldsymbol{T}_{n2}(\boldsymbol{\theta}_n)=\boldsymbol{Q}_n+u_{\mathrm{P}}(1). \tag{4.7.36}$$

Also, $\boldsymbol{Q}_n$ converges in distribution to $\mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))Z$, where Z has mean zero and variance $\gamma(\varphi)$.

Proof. Consider the following decomposition

$$
\begin{aligned}
&\boldsymbol{R}_n(\boldsymbol{\theta}_n) - \boldsymbol{T}_{n2}(\boldsymbol{\theta}_n) \\
&= n^{-1/2}\sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \left\{ \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right] \right\} \\
&= n^{-1/2}\sum_{t=1}^{n} \left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \right] \eta_t \left\{ \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right] \right\} \\
&+ n^{-1/2}\sum_{t=1}^{n} \left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \right] v_{nt}(\boldsymbol{b})\eta_t \left\{ \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right] \right\} \\
&+ n^{-1/2}\sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right] \right\} \\
&+ n^{-1/2}\sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} v_{nt}(\boldsymbol{b})\eta_t \left\{ \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right] \right\} \\
&= \boldsymbol{D}_{n1} + \boldsymbol{D}_{n2} + \boldsymbol{D}_{n3} + \boldsymbol{D}_{n4}.
\end{aligned}
$$

Using the $n^{-1/2}$-factor in (4.7.30) and (4.7.28), $\boldsymbol{D}_{n1}$, $\boldsymbol{D}_{n2}$ and $\boldsymbol{D}_{n4}$ are $u_{\mathrm{P}}(1)$. We next prove that $\boldsymbol{D}_{n3} = \boldsymbol{Q}_n + u_{\mathrm{P}}(1)$ in detail. Recall that $\tilde{G}_n(x)$ is the empirical distribution function of $\{\eta_t\}$. Let $G_n(x), x \in \mathbb{R}$ be the empirical distribution function of $\{X_t/v_t^{1/2}(\boldsymbol{\theta}_n)\}$. Then

$$
\begin{aligned}
\boldsymbol{D}_{n3} &= n^{-1/2}\sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[G_n\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] \right\} \\
&= n^{-1/2}\sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[G(\eta_t) \right] \right\} \\
&- n^{-1/2}\sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi\left[G_n\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right) \right] - \varphi\left[\tilde{G}_n(\eta_t) \right] \right\} \\
&+ n^{-1/2}\sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi\left[G(\eta_t) \right] - \varphi\left[\tilde{G}_n(\eta_t) \right] \right\} \\
&= \boldsymbol{D}^*_{n1} - \boldsymbol{D}^*_{n2} + \boldsymbol{D}^*_{n3}.
\end{aligned}
$$

Since $\boldsymbol{D}^*_{n1}$ is the weighted sum of the difference of a c.d.f. evaluated at two different r.v.'s and integrated wrt φ, using the same technique for proving (4.7.33), $\boldsymbol{D}^*_{n1} = \boldsymbol{M}\boldsymbol{b} + u_{\mathrm{P}}(1)$.

Write $\boldsymbol{w}_t = \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi})$. Since $\boldsymbol{D}^*_{n2}$ is the weighted sum of the difference

of two different c.d.f.'s evaluated at two different r.v.'s and integrated wrt φ,

$$\begin{aligned}
\boldsymbol{D}_{n2}^{*} &= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \eta_t I\left[G_n\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\right) < u\right] - I\left[\tilde{G}_n(\eta_t) < u\right] d\varphi(u) \\
&= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \eta_t I\left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} < G_n^{-1}(u)\right] - I\left[\eta_t < \tilde{G}_n^{-1}(u)\right] d\varphi(u) \\
&= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \eta_t I\left[\eta_t < G_n^{-1}(u)\frac{1}{1+v_{nt}(\boldsymbol{b})}\right] - I\left[\eta_t < \tilde{G}_n^{-1}(u)\right] d\varphi(u).
\end{aligned}$$

Using (4.7.29), $\sup\left\{\left|G_n^{-1}(u) - \tilde{G}_n^{-1}(u)\right|; u \in (0,1)\right\} = u_{\mathrm{P}}(1)$. Hence, by Theorem 4.7.2,

$$\begin{aligned}
\boldsymbol{D}_{n2}^{*} &= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \left[\mu\left(\tilde{G}_n^{-1}(u)\frac{1}{1+v_{nt}(\boldsymbol{b})}\right) - \mu\left(\tilde{G}_n^{-1}(u)\right)\right] d\varphi(u) + u_{\mathrm{P}}(1) \\
&= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \dot{\mu}\left(\tilde{G}_n^{-1}(u)\right)\frac{-v_{nt}(\boldsymbol{b})}{1+v_{nt}(\boldsymbol{b})}\tilde{G}_n^{-1}(u)\, d\varphi(u) + u_{\mathrm{P}}(1) \\
&= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \left(\tilde{G}_n^{-1}(u)\right)^2 g\left(\tilde{G}_n^{-1}(u)\right)\frac{-v_{nt}(\boldsymbol{b})}{1+v_{nt}(\boldsymbol{b})} d\varphi(u) + u_{\mathrm{P}}(1) \\
&= \int_0^1 n^{-1} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_{0\varphi})}{2v_t^2(\boldsymbol{\theta}_{0\varphi})}\left(G^{-1}(u)\right)^2 g\left(G^{-1}(u)\right) d\varphi(u)\boldsymbol{b} + u_{\mathrm{P}}(1) \\
&= \boldsymbol{M}\boldsymbol{b} + u_{\mathrm{P}}(1).
\end{aligned}$$

Finally consider $\boldsymbol{D}_{n3}^{*}$ written as

$$\begin{aligned}
\boldsymbol{D}_{n3}^{*} &= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \eta_t \left\{I\left[\eta_t \le G^{-1}(u)\right] - I\left[\eta_t \le \tilde{G}_n^{-1}(u)\right]\right\} d\varphi(u) \\
&= \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \eta_t \left\{I\left[\eta_t \le G^{-1}(u)\right] - I\left[\eta_t \le G^{-1}(G(\tilde{G}_n^{-1}(u)))\right]\right\} d\varphi(u) \\
&= \int_0^1 \left[\boldsymbol{M}_n(u) - \boldsymbol{M}_n(G(\tilde{G}_n^{-1}(u)))\right] d\varphi(u) \\
&+ \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u))\right] d\varphi(u),
\end{aligned}$$

where $\boldsymbol{M}_n(u) := n^{-1/2}\sum_{t=1}^{n} \boldsymbol{w}_t \{\eta_t I\left[\eta_t \le G^{-1}(u)\right] - \mu(G^{-1}(u))\}$. We show that

$$\left|\int_0^1 \left[\boldsymbol{M}_n(u) - \boldsymbol{M}_n(G(\tilde{G}_n^{-1}(u)))\right] d\varphi(u)\right| = o_{\mathrm{P}}(1), \tag{4.7.37}$$

$$\boldsymbol{Q}_n = \int_0^1 n^{-1/2} \sum_{t=1}^{n} \boldsymbol{w}_t \left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u))\right] d\varphi(u) \to \mathrm{E}(\dot{\boldsymbol{v}}_1/v_1)Z. \quad (4.7.38)$$

For (4.7.37), note that $\{\boldsymbol{M}_n(.)\}$ converges weakly to a Brownian Bridge on $(0,1)$ since for each fixed u, $\boldsymbol{M}_n(u)$ converges to a normal distribution using the martingale CLT and it is tight using the bound on the moment of the difference process in Billingsley (1968, Theorem 12.3).

Since $\sup\{|u - G(\tilde{G}_n^{-1}(u))|; u \in (0,1)\} = \sup\{|G(x) - \tilde{G}_n(x)|; x \in \mathbb{R}\} = o_{\mathrm{P}}(1)$, by the Arzela-Ascoli theorem,

$$\sup\left\{\left|\boldsymbol{M}_n(u) - \boldsymbol{M}_n(G(\tilde{G}_n^{-1}(u)))\right|; u \in (0,1)\right\} = o_{\mathrm{P}}(1),$$

and consequently, (4.7.37) is proved. For (4.7.38), we use the Bahadur representation; see Bahadur (1966) and Ghosh (1971) for details. Since $\dot{g}(x)$ is bounded and g is positive on $\mathbb{R}$,

$$n^{1/2}\left(G^{-1}(u)) - \tilde{G}_n^{-1}(u)\right) - n^{-1/2}\sum_{i=1}^{n} \frac{I\{\eta_i \le G^{-1}(u)\} - u}{g\left(G^{-1}(u)\right)} = o(1) \quad \text{a.s..}$$

Applying the mean value theorem,

$$n^{1/2}\left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u))\right] - \dot{\mu}(G^{-1}(u))n^{-1/2}\sum_{i=1}^{n} \frac{I\{\eta_i \le G^{-1}(u)\} - u}{g\left(G^{-1}(u)\right)} = o(1) \quad \text{a.s..} \quad (4.7.39)$$

Using $\dot{\mu}(x) = xg(x)$,

$$\begin{aligned}\boldsymbol{Q}_n &= n^{-1}\sum_{t=1}^{n} \boldsymbol{w}_t \int_0^1 \left[\dot{\mu}(G^{-1}(u))n^{-1/2}\sum_{i=1}^{n} \frac{I\{\eta_i \le G^{-1}(u)\} - u}{g\left(G^{-1}(u)\right)}\right] d\varphi(u) + o_{\mathrm{P}}(1) \\ &= n^{-1}\sum_{t=1}^{n} \boldsymbol{w}_t \int_0^1 \left[G^{-1}(u)n^{1/2}\left(\tilde{G}_n(G^{-1}(u)) - u\right)\right] d\varphi(u) + o_{\mathrm{P}}(1).\end{aligned}$$

Since $n^{-1}\sum_{t=1}^{n} \boldsymbol{w}_t \to \mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))$, and using van der Vaart (1998, Theo-

rem 19.3),

$$n^{1/2}\left(\tilde{G}_n(G^{-1}(u)) - u\right) \to B(u),$$

we obtain (4.7.38) with the r.v. Z having mean zero. The variance of Z is given by

$$\begin{aligned}
\mathrm{E}(Z^2) &= \mathrm{E}\left[\int_0^1\int_0^1 G^{-1}(u)G^{-1}(v)B(u)B(v)d\varphi(u)d\varphi(v)\right] \\
&= \int_0^1\int_0^1 G^{-1}(u)G^{-1}(v)\mathrm{E}[B(u)B(v)]d\varphi(u)d\varphi(v) \\
&= \int_0^1\int_0^1 G^{-1}(u)G^{-1}(v)\left[\min\{u,v\} - uv\right]d\varphi(u)d\varphi(v) \\
&= \gamma(\varphi).
\end{aligned}$$

□

Now recall the rank-based central sequence

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_n) = n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_n)}{\hat{v}_t(\boldsymbol{\theta}_n)}\left\{1 - \frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta}_n)}\varphi\left(\frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1}\right)\right\},$$

which is an approximation to $\boldsymbol{R}_n(\boldsymbol{\theta}_n)$. We have the following lemma dealing with the difference between $\boldsymbol{R}_n(\boldsymbol{\theta}_n)$ and $\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_n)$.

Lemma 4.7.6. *Let Assumptions (A1)-(A3) hold. Then, as $n \to \infty$,*

$$\boldsymbol{R}_n(\boldsymbol{\theta}_n) - \hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_n) = u_{\mathrm{P}}(1). \tag{4.7.40}$$

Proof. Note that $\boldsymbol{R}_n(\boldsymbol{\theta}_n) - \hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_n)$ equals

$$n^{-1/2}\sum_{t=1}^{n}\left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)} - \frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_n)}{\hat{v}_t(\boldsymbol{\theta}_n)}\right] \tag{4.7.41}$$

$$+ n^{-1/2}\sum_{t=1}^{n}\left[\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_n)}{\hat{v}_t(\boldsymbol{\theta}_n)}\frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta}_n)} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\right]\varphi\left(\frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1}\right) \tag{4.7.42}$$

$$- n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)}\left[\varphi\left(\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1}\right) - \varphi\left(\frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1}\right)\right]. \tag{4.7.43}$$

Due to (4.7.2), (4.7.3) and $\hat{v}_t(\boldsymbol{\theta}) \geq c_0(\boldsymbol{\theta}) > 0$, we have

$$\begin{aligned}\left|\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_n)}{\hat{v}_t(\boldsymbol{\theta}_n)} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\right| &= \left|\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_n) - \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{\hat{v}_t(\boldsymbol{\theta}_n)} + \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)\frac{v_t(\boldsymbol{\theta}_n) - \hat{v}_t(\boldsymbol{\theta}_n)}{\hat{v}_t(\boldsymbol{\theta}_n)v_t(\boldsymbol{\theta}_n)}\right| \\ &\leq \frac{|\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_n) - \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)|}{\hat{v}_t(\boldsymbol{\theta}_n)} + \frac{|v_t(\boldsymbol{\theta}_n) - \hat{v}_t(\boldsymbol{\theta}_n)|}{\hat{v}_t(\boldsymbol{\theta}_n)}\frac{|\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)|}{v_t(\boldsymbol{\theta}_n)} \\ &\leq C\rho^t\left[Z_1 + Z_0\frac{|\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)|}{v_t(\boldsymbol{\theta}_n)}\right]. \qquad (4.7.44)\end{aligned}$$

Hence, in view of (4.7.1) and (4.7.4), for every $0 < b < \infty$,

$$\sup_{||\boldsymbol{b}||<b}\sum_{t=1}^{n}\left|\frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_n)}{\hat{v}_t(\boldsymbol{\theta}_n)} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\right| = O_{\mathrm{p}}(1),$$

which implies that (4.7.41) is $u_{\mathrm{P}}(1)$. Since φ is bounded, (4.7.42) is $u_{\mathrm{P}}(1)$. For (4.7.43), since there is a $n^{-1/2}$ factor from

$$\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_n)}{v_t(\boldsymbol{\theta}_n)}\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\eta_t\,,$$

it suffices to prove that

$$\boldsymbol{K}_n := n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\eta_t\left[\varphi\left(\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1}\right) - \varphi\left(\frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1}\right)\right] = u_{\mathrm{P}}(1).$$

Let $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. We split the sum in $\boldsymbol{K}_n$ into two parts: in the first part, t runs till $\lfloor n^k \rfloor - 1$ where $0 < k < 1/2$. We show that this part is $u_{\mathrm{P}}(1)$ by noting that its expectation is of the form $n^k n^{-1/2} = o(1)$ multiplied by expectation of $\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi})\eta_t$ and a bounded quantity because φ is bounded. The number of summands in the second term is $n - n^k$ which is large but there we bound expectation of the sum of by a quantity of the form $n\rho^{\lfloor n^k \rfloor}$

with $0 < k < 1/2$ and $0 < \rho < 1$ and this is $o(1)$. Accordingly

$$\begin{aligned}\boldsymbol{K}_n = \quad & n^{-1/2} \sum_{t=1}^{\lfloor n^k \rfloor - 1} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left[\varphi\left(\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right) - \varphi\left(\frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1} \right) \right] \\ & + n^{-1/2} \sum_{t=\lfloor n^k \rfloor}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left[\varphi\left(\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right) - \varphi\left(\frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1} \right) \right]. \end{aligned} \tag{4.7.45}$$

To show (4.7.45) is $u_{\mathrm{P}}(1)$, we prove that for every $0 < b < \infty$,

$$\sup_{\substack{\lfloor n^k \rfloor \le t \le n \\ ||\boldsymbol{b}|| < b}} \left| \varphi\left(\frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} \right) - \varphi\left(\frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1} \right) \right| = O_{\mathrm{p}}(n^{k-1}). \tag{4.7.46}$$

Since sequences $\{R_{nt}(\boldsymbol{\theta}_n)\}$ and $\{\hat{R}_{nt}(\boldsymbol{\theta}_n)\}$ are permutations of $\{1, ..., n\}$, with the probability tending to one as $n \to \infty$, both $\{R_{nt}(\boldsymbol{\theta}_n)\}$ and $\{\hat{R}_{nt}(\boldsymbol{\theta}_n)\}$ are at points of continuity of φ that has a finite number of the points of discontinuity. Therefore, to prove (4.7.46), it suffices to prove

$$\sup_{\substack{\lfloor n^k \rfloor \le t \le n \\ ||\boldsymbol{b}|| < b}} \left| \frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} - \frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1} \right| = O_{\mathrm{p}}(n^{k-1}). \tag{4.7.47}$$

For $\lfloor n^k \rfloor \le t \le n$, we decompose ranks as

$$\begin{aligned} & \frac{R_{nt}(\boldsymbol{\theta}_n)}{n+1} - \frac{\hat{R}_{nt}(\boldsymbol{\theta}_n)}{n+1} \\ & = \frac{1}{n+1} \sum_{i=1}^{\lfloor n^k \rfloor - 1} \left\{ I\left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_n)} < \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right] - I\left[\frac{X_i}{\hat{v}_i^{1/2}(\boldsymbol{\theta}_n)} < \frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta}_n)} \right] \right\} \\ & \quad + \frac{1}{n+1} \sum_{i=\lfloor n^k \rfloor}^{n} \left\{ I\left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_n)} < \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right] - I\left[\frac{X_i}{\hat{v}_i^{1/2}(\boldsymbol{\theta}_n)} < \frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta}_n)} \right] \right\}, \end{aligned} \tag{4.7.48}$$

where the first sum is $O_{\mathrm{P}}(n^{k-1})$. For the second sum, writing

$$I\left[\frac{X_i}{\hat{v}_i^{1/2}(\boldsymbol{\theta}_n)} < \frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta}_n)} \right] = I\left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_n)} \frac{v_i^{1/2}(\boldsymbol{\theta}_n) \hat{v}_t^{1/2}(\boldsymbol{\theta}_n)}{\hat{v}_i^{1/2}(\boldsymbol{\theta}_n) v_t^{1/2}(\boldsymbol{\theta}_n)} < \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_n)} \right],$$

the modulus of (4.7.48) is bounded above by

$$\sup_{\substack{x\in\mathbb{R}\\ ||\boldsymbol{b}||<b}} \frac{1}{n+1}\sum_{i=\lfloor n^k\rfloor}^{n}\left|I\left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_n)}<x\right]-I\left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_n)}\frac{v_i^{1/2}(\boldsymbol{\theta}_n)\hat{v}_t^{1/2}(\boldsymbol{\theta}_n)}{\hat{v}_i^{1/2}(\boldsymbol{\theta}_n)v_t^{1/2}(\boldsymbol{\theta}_n)}<x\right]\right|.$$

Using $|I(A)-I(B)|\leq I(A\cap B^c)+I(A^c\cap B)$, this is bounded above by

$$\sup_{\substack{x\in\mathbb{R}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \frac{1}{n+1}\sum_{i=\lfloor n^k\rfloor}^{n} I(\mathcal{A}_{i,x,\boldsymbol{\theta}}),$$

where the set $\mathcal{A}_{i,x,\boldsymbol{\theta}}$ is defined as

$$\begin{aligned}\mathcal{A}_{i,x,\boldsymbol{\theta}} := &\left\{\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}<x, \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\frac{v_i^{1/2}(\boldsymbol{\theta})\hat{v}_t^{1/2}(\boldsymbol{\theta})}{\hat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})}\geq x\right\}\\ &\cup\left\{\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\geq x, \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\frac{v_i^{1/2}(\boldsymbol{\theta})\hat{v}_t^{1/2}(\boldsymbol{\theta})}{\hat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})}<x\right\}.\end{aligned}$$

Therefore, it suffices to prove that $\sum_{i=\lfloor n^k\rfloor}^{n} I(\mathcal{A}_{i,x,\boldsymbol{\theta}})=o_{\mathrm{P}}(1)$ uniformly with respect to both x and $\boldsymbol{\theta}$. We show this with sets containing $\mathcal{A}_{i,x,\boldsymbol{\theta}}$.

Recall that using (4.7.2), $\hat{v}_t(\boldsymbol{\theta})\geq c_0(\boldsymbol{\theta})>c$ for a positive constant c and so

$$0<v_t^{1/2}(\boldsymbol{\theta})-\hat{v}_t^{1/2}(\boldsymbol{\theta})\leq\frac{\rho^t Z_0}{v_t^{1/2}(\boldsymbol{\theta})+\hat{v}_t^{1/2}(\boldsymbol{\theta})}\leq\frac{\rho^t Z_0}{2c_0^{1/2}(\boldsymbol{\theta})}.$$

Now using the triangular inequality,

$$\begin{aligned}&\left|\frac{v_i^{1/2}(\boldsymbol{\theta})\hat{v}_t^{1/2}(\boldsymbol{\theta})}{\hat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})}-1\right|\\ \leq&\left|\frac{\hat{v}_t^{1/2}(\boldsymbol{\theta})\left(v_i^{1/2}(\boldsymbol{\theta})-\hat{v}_i^{1/2}(\boldsymbol{\theta})\right)}{\hat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})}\right|+\left|\frac{\hat{v}_i^{1/2}(\boldsymbol{\theta})\left(\hat{v}_t^{1/2}(\boldsymbol{\theta})-v_t^{1/2}(\boldsymbol{\theta})\right)}{\hat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})}\right|.\end{aligned}\tag{4.7.49}$$

Therefore (4.7.49) is bounded above by

$$\frac{\rho^i Z_0}{2c_0^{1/2}(\boldsymbol{\theta})}+\frac{\rho^t Z_0}{2c_0^{1/2}(\boldsymbol{\theta})}.$$

In view of (4.7.49), we get

$$\left|\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\frac{v_i^{1/2}(\boldsymbol{\theta})\hat{v}_t^{1/2}(\boldsymbol{\theta})}{\hat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})}-\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\right|\leq(\rho^i+\rho^t)Z_4\left|\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\right|,$$

where $Z_4=Z_0/(2C^{1/2})$.

Therefore, $\mathcal{A}_{i,x,\boldsymbol{\theta}}$ is a subset of

$$\begin{aligned}\mathcal{B}_{i,x,\boldsymbol{\theta}}:=&\left\{\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}<x,\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}+(\rho^i+\rho^t)Z_4\left|\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\right|\geq x\right\}\\&\cup\left\{\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\geq x,\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}-(\rho^i+\rho^t)Z_4\left|\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})}\right|<x\right\}\\=&\left\{\eta_i<x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})},\eta_i+(\rho^i+\rho^t)Z_4|\eta_i|\geq x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}\right\}\\&\cup\left\{\eta_i\geq x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})},\eta_i-(\rho^i+\rho^t)Z_4|\eta_i|<x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}\right\}\\=&\left\{x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}-(\rho^i+\rho^t)Z_4|\eta_i|\leq\eta_i<x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}\right\}\\&\cup\left\{x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}\leq\eta_i<x\frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}+(\rho^i+\rho^t)Z_4|\eta_i|\right\}.\end{aligned}$$

Consider r.v.s X and $L\geq 0$ with X independent of η. Then $\mathrm{P}[X<\eta<X+L]\leq\sup_{y\in\mathbb{R}}\{g(y)\}\mathrm{E}(L)$ where the p.d.f. g is bounded. Consequently,

$$\sup_{\substack{x\in\mathbb{R}\\\boldsymbol{\theta}\in\boldsymbol{\Theta}}}\mathrm{E}\left[\sum_{i=\lfloor n^k\rfloor}^{n}I(\mathcal{A}_{i,x,\boldsymbol{\theta}})\right]\leq\sup_{y\in\mathbb{R}}g(y)\sum_{i=\lfloor n^k\rfloor}^{n}\mathrm{E}\left\{(\rho^i+\rho^t)Z_4|\eta_i|\right\}.$$

Notice that since $\lfloor n^k\rfloor\leq t\leq n$,

$$\sum_{i=\lfloor n^k\rfloor}^{n}\mathrm{E}\left\{\rho^tZ_4|\eta_i|\right\}\leq n\rho^{\lfloor n^k\rfloor}\mathrm{E}(Z_4)\mathrm{E}|\eta|=o(1)$$

due to $0<\rho<1$ and $\mathrm{E}|\eta|<\infty$. Hence, $\sum_{i=\lfloor n^k\rfloor}^{n}I(\mathcal{A}_{i,x,\boldsymbol{\theta}})$ converges in mean to zero uniformly with respect to both x and $\boldsymbol{\theta}$, which entails $\sum_{i=\lfloor n^k\rfloor}^{n}I(\mathcal{A}_{i,x,\boldsymbol{\theta}})=o_{\mathrm{P}}(1)$ uniformly.

□

With all the results above, we can easily prove Proposition 4.2.1 and the asymptotic result for the R-estimator as follows.

Proof of Proposition 4.2.1.

Proof. Combining (4.7.34), (4.7.36), (4.7.32) and (4.7.40), we get

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_n) - \boldsymbol{M}\boldsymbol{b} - \boldsymbol{Q}_n - \boldsymbol{N}_n - \boldsymbol{J}\boldsymbol{b}/2 = u_{\mathrm{P}}(1), \tag{4.7.50}$$

which, by letting $\boldsymbol{b} = \boldsymbol{0}$, entails

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{Q}_n + \boldsymbol{N}_n + u_{\mathrm{P}}(1).$$

Hence, (4.2.6) follows by recalling that $\boldsymbol{M} = \boldsymbol{J}\rho(\varphi)/2$.

The proof of (4.2.7) follows directly from (4.7.35) and (4.7.38).

□

Proof of Theorem 4.2.1.

Proof. From the definition of $\hat{\boldsymbol{\theta}}_n$ in ((5.4.8), (4.2.6) and (4.2.7) in Proposition 4.2.1, consistency of $\hat{\boldsymbol{\Upsilon}}_n$ and the asymptotic discreteness of $\bar{\boldsymbol{\theta}}_n$ (which allows us to treat $n^{1/2}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi})$ as if it were a bounded constant: see Lemma 4.4 in Kreiss (1987)), we have

$$\begin{aligned}
&n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}) \\
=&n^{1/2}\left\{\bar{\boldsymbol{\theta}}_n - n^{-1/2}\left(\hat{\boldsymbol{\Upsilon}}_n\right)^{-1}\hat{\boldsymbol{R}}_n(\bar{\boldsymbol{\theta}}_n) - \boldsymbol{\theta}_{0\varphi}\right\} \\
=&n^{1/2}\left\{\bar{\boldsymbol{\theta}}_n - n^{-1/2}\left(\hat{\boldsymbol{\Upsilon}}_n\right)^{-1}\left[\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) + (1/2+\rho(\varphi)/2)\boldsymbol{J}n^{1/2}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi})\right] - \boldsymbol{\theta}_{0\varphi}\right\} + o_{\mathrm{P}}(1) \\
=&n^{1/2}\left\{\bar{\boldsymbol{\theta}}_n - n^{-1/2}(1/2+\rho(\varphi)/2)^{-1}\boldsymbol{J}^{-1}\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) - (\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}) - \boldsymbol{\theta}_{0\varphi}\right\} + o_{\mathrm{P}}(1) \\
=&-(1/2+\rho(\varphi)/2)^{-1}\boldsymbol{J}^{-1}\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) + o_{\mathrm{P}}(1).
\end{aligned}$$

In view of (4.2.7), we have

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}) = -(1/2 + \rho(\varphi)/2)^{-1}\boldsymbol{J}^{-1}(\boldsymbol{Q}_n + \boldsymbol{N}_n) + o_{\mathrm{P}}(1).$$

Now, it remains to obtain the asymptotic covariance matrix of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}\right)$. Recall (4.7.35) and (4.7.38). Since the asymptotic covariance matrices of $\boldsymbol{Q}_n$ and $\boldsymbol{N}_n$ have been derived, it remains to obtain the covariance matrix $\mathrm{Cov}(\boldsymbol{Q}_n, \boldsymbol{N}_n)$. Note that $\mathrm{E}[\boldsymbol{Q}_n\boldsymbol{N}_n']$ equals

$$\mathrm{E}\left\{\left[\int_0^1 n^{-1/2}\sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u))\right]d\varphi(u)\right]\right.$$
$$\left.\times\left[n^{-1/2}\sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\left[1 - \eta_t\varphi\left[G(\eta_t)\right]\right]\right]\right\}. \tag{4.7.51}$$

Using (4.7.39) and $n^{-1}\sum_{t=1}^n \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi}) \to \mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))$, as $n \to \infty$, (4.7.51) has the same limit as

$$\begin{aligned}
&\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\\
&\times\lim_{n\to\infty}\mathrm{E}\left\{\int_0^1 G^{-1}(u)\left\{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left[I\{\eta_i \le G^{-1}(u)\} - u\right]\frac{\dot{\boldsymbol{v}}_j'(\boldsymbol{\theta}_{0\varphi})}{v_j(\boldsymbol{\theta}_{0\varphi})}\left[1 - \eta_j\varphi\left[G(\eta_j)\right]\right]\right\}d\varphi(u)\right\}\\
&= \mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\\
&\quad\times\lim_{n\to\infty}\mathrm{E}\left\{\int_0^1 G^{-1}(u)\left\{\frac{1}{n}\sum_{i=1}^n\left[I\{\eta_i \le G^{-1}(u)\} - u\right]\frac{\dot{\boldsymbol{v}}_i'(\boldsymbol{\theta}_{0\varphi})}{v_i(\boldsymbol{\theta}_{0\varphi})}\left[1 - \eta_i\varphi\left[G(\eta_i)\right]\right]\right\}d\varphi(u)\right\}\\
&= \mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\int_0^1 G^{-1}(u)\mathrm{E}\left\{\left[I\{\eta_1 \le G^{-1}(u)\} - u\right]\left[1 - \eta_1\varphi\left[G(\eta_1)\right]\right]\right\}d\varphi(u)\\
&= \mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\int_0^1 G^{-1}(u)\mathrm{E}\left\{I\{\eta_1 \le G^{-1}(u)\}\left[1 - \eta_1\varphi\left[G(\eta_1)\right]\right]\right\}d\varphi(u),
\end{aligned}$$

where the first equality is due to independence of η_i and η_j for $i \neq j$, independence of v_j and η_j, and Assumption (A1). The second equality is due to independence of v_i and η_i. The last equality is due to Assumption (A1).

Recall the definition of $\lambda(\varphi)$ in (4.2.5), which can also be written as

$$\lambda(\varphi) = \int_0^1 G^{-1}(u)\mathrm{E}\left\{I\{\eta_1 \leq G^{-1}(u)\}\left[1 - \eta_1\varphi\left[G(\eta_1)\right]\right]\right\} d\varphi(u).$$

We then have

$$\lim_{n\to\infty} \mathrm{Cov}(\boldsymbol{Q}_n, \boldsymbol{N}_n) = \mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\lambda(\varphi).$$

Hence, by recalling (4.7.38) and in view of (5.4.2), the asymptotic covariance matrix of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}\right)$ is

$$\begin{aligned}
&\boldsymbol{J}^{-1}\frac{\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{Var}(Z) + \mathrm{Var}(\boldsymbol{N}) + 2\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\lambda(\varphi)}{(1/2 + \rho(\varphi)/2)^2}\boldsymbol{J}^{-1} \\
&= \boldsymbol{J}^{-1}\frac{[4\gamma(\varphi) + 8\lambda(\varphi)]\,\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right) + 4\sigma^2(\varphi)\boldsymbol{J}}{(1 + \rho(\varphi))^2}\boldsymbol{J}^{-1}.
\end{aligned}$$

□

Chapter 5

Center-outward R-estimation for semiparametric VARMA models

5.1 Introduction

5.1.1 Quasi-maximum likelihood and R-estimation

Gaussian *quasi*-likelihood methods are pervasive in several areas of statistics. Among them is time series analysis, univariate and multivariate, linear and nonlinear. In particular, quasi-maximum likelihood estimation (QMLE))[1] and correlogram-based testing are the daily practice golden standard for ARMA and VARMA models. They only require the specification of the first two conditional moments, which depend on an unknown Euclidean parameter, while a Gaussian (misspecified) innovation density is assumed. Their properties are generally considered as fully satisfactory: QMLEs, in particular, are root-n consistent, parametrically efficient under Gaussian innovations, and asymptotically normal under finite fourth-order moment assumptions.

Despite their popularity, QMLE methods are not without some undesirable consequences, though, which are often overlooked: *(i)* while achieving efficiency under Gaussian innovations, their asymptotic performance can be quite poor un-

[1] Unless otherwise stated, "QMLE" here throughout refers to "Gaussian QMLE."

der non-Gaussian ones; *(ii)* due to technical reasons (the *Fisher consistency* requirement), the choice of a quasi-likelihood is always the most pessimistic one: quasi-likelihoods automatically are based on the least favorable innovation density (here, a Gaussian one); *(iii)* root-n consistency is far from being uniform across innovation densities; *(iv)* actual fourth-order moments may be infinite.

In principle, the ultimate theoretical remedy to those problems is the semiparametric estimation method described in the monograph by Bickel et al. (1993), which yields uniformly, locally and asymptotically, semiparametrically efficient estimators. For VARMA models, the semiparametric approach does not specify the innovation density (an infinite-dimensional nuisance) and the estimators based on Bickel et al. (1993) methodology are uniformly, locally and asymptotically *parametrically* efficient (VARMA models are *adaptive*, thus semiparametric and parametric efficiency coincide). However, semiparametric estimation procedures are not easily implemented, since they rely on kernel-based estimation of the actual innovation density (hence the choice of a kernel, the selection of a bandwidth) and the use of sample splitting techniques. All these niceties require relatively large samples and are hard to put into practice even for univariate time series.

A more flexible and computationally less heavy alternative in the presence of unspecified noise or innovation densities is R-estimation, which reaches efficiency at some chosen reference density (not necessarily Gaussian or least favorable) or class of densities. R-estimation has been proposed first in the context of location (Hodges and Lehmann 1956) and regression models with independent observations (Jurečková 1971, Koul 1971, van Eeden 1972, Jaeckel 1972). Later on, it was extended to autoregressive time series (Koul and Saleh 1993, Koul and Ossiander 1994, Terpstra et al. 2001, Hettmansperger and McKean 2010, Mukherjee and Bai 2002, Andrews 2008, 2012) and non-linear time series (Mukherjee 2007, Andreou and Werker 2015, Hallin and La Vecchia 2017, 2019).

Multivariate extensions of these approaches, however, run into the major difficulty of defining an adequate concept of ranks in the multivariate context. This

is most regrettable, as the drawbacks of quasi-likelihood methods for observations in dimension $d = 1$ only get worse as the dimension d increases (see Section 5.1.2 for a numerical example in dimension $d = 2$) while the use of the semiparametric method of Bickel et al. becomes problematic: the higher the dimension, the more delicate multivariate kernel density estimation and the larger the required sample size. A natural question is thus: "Can R-estimation palliate the drawbacks of the QMLE and the Bickel et al. technique in dimension $d \geq 2$ the way it does in dimension $d = 1$?" This question immediately comes up against another one: "What are ranks and signs, hence, what is R-estimation, in dimension $d \geq 2$?" Indeed, starting with dimension two, the real space $\mathbb{R}^d$ is no longer canonically ordered.

The main contribution of this chapter is to provide an answer to these questions. To this end, we propose a multivariate version of R-estimation, establish its asymptotic properties (root-n consistency and asymptotic normality), and demonstrate its feasibility and excellent finite-sample performance in the context of semiparametric VARMA models. Our approach builds on Chernozhukov et al. (2017), Hallin (2017), and Hallin et al. (2020a), who introduce novel concepts of *center-outward ranks and signs* based on measure transportation ideas. These center-outward ranks and signs (see Section 5.3 for details) enjoy all the properties that make traditional univariate ranks a successful tool of inference. In particular, they are distribution-free (see Hallin et al. (2020a) for details), thus preserve the validity of rank-based procedures irrespective of the possible misspecification of the innovation density. Moreover, they are invariant with respect to shift and global scale factors and equivariant under orthogonal transformations; see Hallin et al. (2020b). Extensive numerical exercises reveal the finite-sample superiority of our R-estimators over the conventional QMLE in the presence of asymmetric innovation densities (skew-normal, skew-t, Gaussian mixtures) and in the presence of outliers. All these advantages, moreover, do not come at the cost of a loss of efficiency under symmetry.

Other notions of multivariate ranks and signs have been proposed in the statistical literature. Among them, the componentwise ranks (Puri and Sen 1971), the spatial ranks (Oja 2010), the depth-based ranks (Liu 1992; Liu and Singh 1993), and the Mahalanobis ranks and signs (Hallin and Paindaveine (2002a)). Those ranks and signs all have their own merits but also some drawbacks, which make them unsuitable for our needs (essentially, they are not distribution-free, or not maximally so); we refer to the introduction of Hallin et al. (2020a) for details. The Mahalanobis ranks and signs have been successfully considered for testing purposes in the time series context (Hallin and Paindaveine 2002b, 2004). However, no results on estimation are available, and their distribution-freeness property is limited to elliptical densities—a very strong symmetry assumption which we are dropping here.

Leaving aside Wasserstein-distance-based methods, our contribution constitutes the first inferential application of measure transportation ideas to semiparametric inference for multivariate time series. Measure transportation, which goes back to Gaspard Monge (1746-1818) and his 1781 *Mémoire sur la Thérorie des Déblais et des Remblais*, in the past few years has become one of the most active and fertile subjects in pure and applied contemporary mathematics. Despite some crucial forerunning contributions (Cuesta-Albertos and Matrán (1997); Rachev and Rüschendorf (1998)), statistics was somewhat slower to join. However, some recent papers on multiple-output quantile regression (Carlier et al. 2016), distribution-free tests of vector independence and multivariate goodness-of-fit (Boeckel et al. (2018); Deb and Sen (2019); Shi et al. (2019); Ghosal and Sen (2019)) demonstrate the growing interest of the statistical community in measure transportation results. We refer to Panaretos and Zemel (2019) for a review.

5.1.2 A motivating example

As a justification of the practical interest of our R-estimation method, let us consider the very simple but highly representative motivating example of a bivariate

VAR(1) model

$$(\boldsymbol{I}_d - \boldsymbol{A}L)\,\boldsymbol{X}_t = \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}, \tag{5.1.1}$$

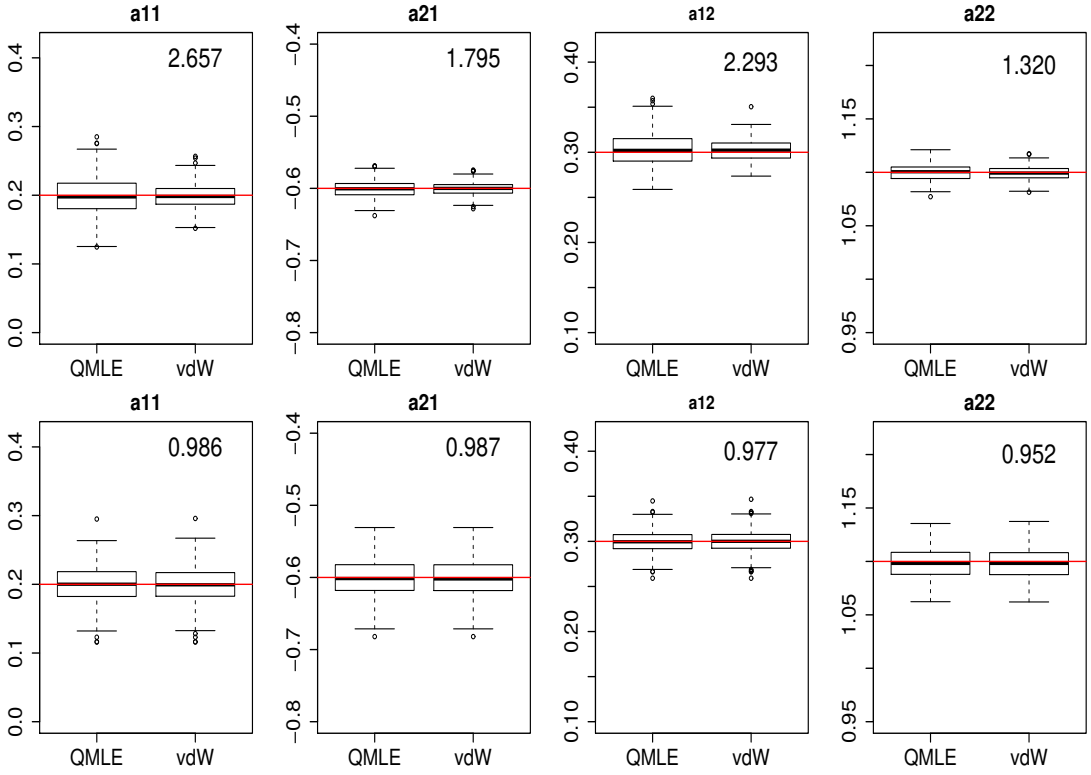
with parameter $\mathrm{vec}(\boldsymbol{A}) =: (a_{11}, a_{21}, a_{12}, a_{22})'$ taking the value $(0.2, -0.6, 0.3, 1.1)'$. We generated 300 replications of a realization of length $n = 1000$ of the stationary solution of (5.1.1) with two innovation densities—a spherical Gaussian one and a Gaussian mixture (see (5.5.2) for details)—which both satisfy the conditions for QMLE validity. The resulting boxplots of the QMLE and the Gaussian score (van der Waerden) R-estimator (see Section 5.4.2 for a definition) are shown in Figure 5.1, along with the mean squared error (MSE) ratios of the QMLE over the R-estimator.

Even a vary rapid inspection of the plots reveals that, under the mixture distribution, the R-estimator yields sizeably smaller MSE values than the QMLE. For instance, as far as the estimation of a_{11} is concerned, the MSE ratio is 2.657: the R-estimator is strikingly less dispersed than the QMLE. On the other hand, under Gaussian innovations (hence, with the QMLE coinciding with the MLE and achieving parametric efficiency), the QMLE and the R-estimator perform similarly, with MSE ratios extremely close to one for all the parameters. While our R-estimator quite significantly outperforms the QMLE under the mixture distribution, thus, this benefit comes at no cost under Gaussian innovations. Further numerical results are provided in Section 5.5 and Section 6.4; they all lead to the same conclusion.

5.1.3 Outline of the chapter

The rest of the chapter is organized as follows. Section 5.2 briefly recalls a local asymptotic normality result for the VARMA model with nonelliptical innovation density: an analytical form of the central sequence as a function of the residuals is provided, which indeed plays a key role in the construction of our estimators. In Section 5.3, we introduce the measure transportation-based notions of center-outward ranks and signs; for the sake of analogy, we also recall the definition

Figure 5.1: Boxplots of the QMLE and the R-estimator (van der Waerden) of the parameters a_{11}, a_{21}, a_{12}, and a_{22} of the bivariate VAR(1) (5.1.1) under the Gaussian mixture (5.5.2) (upper panel) and spherical Gaussian (lower panel) innovation densities, respectively (300 replications of length $n = 1000$). In each panel, the MSE ratio of the QMLE with respect to the R-estimator is reported. The horizontal line represents the actual parameter value.



of Mahalanobis ranks and signs, and shortly discuss their respective invariance properties. In Section 5.3.4, we explain the key idea of our construction of R-estimators, which consists in replacing the residuals appearing in central sequence with some adequate function of their center-outward ranks and signs, yielding a rank-based version of the latter: our R-estimators are obtained by incorporating that rank-based central sequence into a classical Le Cam one-step procedure. Root-n consistency and asymptotic normality are established in Proposition 5.4.2 under absolutely continuous innovation densities admitting finite second moment. Some standard score functions are discussed in Section 5.4.2. Section 5.5 presents simulation results under various densities of the various estimators; comparing their performance confirms the findings of the motivating example of Section 5.1.2. In Section 5.6, we show how our R-estimation method applies to a real dataset borrowed from econometrics, where a VARMA$(3, 1)$ model is identified. Finally, Section 5.7 concludes and provides some perspectives for future research.

To make this chapter more concise and readable, all proofs are postponed to Sections 6.1 and 6.2 of Chapter 6. Sections 5.2 and 5.3 are technical and can be skipped at first reading: the applied statistician can focus directly on the description of one-step R-estimation in (5.4.8) (implementation details are provided in Section 6.3 of Chapter 6) and the numerical results of Sections 5.5 and 5.6 (Section 6.4 of Chapter 6).

5.2 Local asymptotic normality

Local asymptotic normality (LAN) is an essential ingredient in the construction of our estimators and the derivation of their asymptotic properties. In this section, referring to results by Garel and Hallin (1995) and Hallin and Paindaveine (2004), we state, along with the required assumptions, the LAN property for stationary VARMA models, with an explicit expression for the central sequence to be used later on. The corresponding technical material is available in Sections 6.1 and 6.2 of Chapter 6.

5.2.1 Notation and assumptions

We throughout consider the d-dimensional VARMA(p, q) model

$$\Big(\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i L^i\Big)\boldsymbol{X}_t = \Big(\boldsymbol{I}_d + \sum_{j=1}^{q} \boldsymbol{B}_j L^j\Big)\boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}, \tag{5.2.1}$$

where $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_q$ are $d \times d$ matrices, L is the lag operator, and $\{\boldsymbol{\epsilon}_t; t \in \mathbb{Z}\}$ is an i.i.d. mean-zero innovation process with density f. The observed series is $\{\boldsymbol{X}_1^{(n)}, \ldots, \boldsymbol{X}_n^{(n)}\}$ (superscript $^{(n)}$ omitted whenever possible) and the $(p+q)d^2$-dimensional parameter of interest is

$$\boldsymbol{\theta} := \big((\mathrm{vec}\boldsymbol{A}_1)', \ldots, (\mathrm{vec}\boldsymbol{A}_p)', (\mathrm{vec}\boldsymbol{B}_1)', \ldots, (\mathrm{vec}\boldsymbol{B}_q)'\big)',$$

where $'$ indicates transposition. Letting $\boldsymbol{A}(L) := \boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i L^i$, and $\boldsymbol{B}(L) := \boldsymbol{I}_d + \sum_{j=1}^{q} \boldsymbol{B}_j L^j$, the following conditions are assumed to hold.

Assumption (A1). *(i)* All solutions of the determinantal equations

$$\det\left(\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i z^i\right) = 0 \quad \text{and} \quad \det\left(\boldsymbol{I}_d + \sum_{i=1}^{q} \boldsymbol{B}_i z^i\right) = 0, \quad z \in \mathbb{C}$$

lie outside the unit ball in $\mathbb{C}$; *(ii)* $|\boldsymbol{A}_p| \neq 0 \neq |\boldsymbol{B}_q|$; *(iii)* $\boldsymbol{I}_d$ is the greatest common left divisor of $\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i z^i$ and $\boldsymbol{I}_d + \sum_{i=1}^{q} \boldsymbol{B}_i z^i$.

Assumption (A1) is standard in the time series literature; the restrictions it imposes on the model parameter ensure the asymptotic stationarity of any solution to (5.2.1).

To proceed further, we assume that the innovation density f is non-vanishing over $\mathbb{R}^d$. More precisely we assume that, for all $c \in \mathbb{R}^+$, there exist constants $b_{c;f}$ and $a_{c;f}$ in $\mathbb{R}$ such that $0 < b_{c;f} \leq a_{c;f} < \infty$ and $b_{c;f} \leq f(\boldsymbol{x}) \leq a_{c;f}$ for $\|\boldsymbol{x}\| \leq c$: denote by $\mathcal{F}_d$ this family of densities.

Assumption (A2). The innovation density $f \in \mathcal{F}_d$ is such that *(i)* $\int \boldsymbol{x} f(\boldsymbol{x}) d\mu = \boldsymbol{0}$ and the covariance $\boldsymbol{\Xi} := \int \boldsymbol{x}\boldsymbol{x}' f(\boldsymbol{x}) d\mu$ is positive definite; *(ii)* there exists a square-integrable d-dimensional vector $\boldsymbol{D}f^{1/2}$ such that, for all sequence $\boldsymbol{h} \in \mathbb{R}^d$

such that $\mathbf{0} \neq \boldsymbol{h} \to \mathbf{0}$,

$$(\boldsymbol{h}'\boldsymbol{h})^{-1} \int [f^{1/2}(\boldsymbol{x}+\boldsymbol{h}) - f^{1/2}(\boldsymbol{x}) - \boldsymbol{h}'\boldsymbol{D}f^{1/2}(\boldsymbol{x})]^2 d\mu \to 0,$$

i.e., $f^{1/2}$ is mean-square differentiable, with mean square gradient $\boldsymbol{D}f^{1/2}$; *(iii)* letting

$$\boldsymbol{\varphi}_f(\boldsymbol{x}) := (\varphi_1(\boldsymbol{x}), \ldots, \varphi_d(\boldsymbol{x}))' := -2(\boldsymbol{D}f^{1/2})/f^{1/2}, \tag{5.2.2}$$

$\int \varphi_i^4(\boldsymbol{x}) f(\boldsymbol{x}) d\mu < \infty$, $i = 1, \ldots, d$; *(iv)* the score function $\boldsymbol{\varphi}_f$ is piecewise Lipschitz, i.e., there exists a finite measurable partition of $\mathbb{R}^d$ into J non-overlapping subsets I_j, $j = 1, \ldots, J$ and a constant $K < \infty$ such that $\|\boldsymbol{\varphi}_f(\boldsymbol{x}) - \boldsymbol{\varphi}_f(\boldsymbol{y})\| \leq K\|\boldsymbol{x} - \boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y}$ in I_j, $j = 1, \ldots, J$.

Assumption (A2)(i) requires the existence of the second moment of the innovations (a necessary condition for finite VARMA Fisher information). (A2)(ii) is a multivariate version of the classical one-dimensional quadratic mean differentiability assumption on $f^{1/2}$. Together, (A2)(i) and (A2)(iii) imply the existence and finiteness of the Fisher information matrix for location $\boldsymbol{\mathcal{I}}(f) = \int \boldsymbol{\varphi}_f(\boldsymbol{x})\boldsymbol{\varphi}_f'(\boldsymbol{x}) f(\boldsymbol{x}) d\mu$ appearing in Proposition 2.1 below. See Garel and Hallin (1995) for further discussion.

Let $\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta})$ denote the residuals computed from the initial values $\boldsymbol{\epsilon}_{-q+1}, \ldots, \boldsymbol{\epsilon}_0$ and $\boldsymbol{X}_{-p+1}, \ldots, \boldsymbol{X}_0$, the parameter value $\boldsymbol{\theta}$, and the observations $\boldsymbol{X}^{(n)} := \big(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\big)$; those residuals can be computed recursively, or from (6.1.1). Clearly, $\boldsymbol{X}^{(n)}$ is the finite realization of a solution of (5.2.1) with parameter value $\boldsymbol{\theta}$ iff $\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta})$ and $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ coincide. Denoting by $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ the distribution of $\boldsymbol{X}^{(n)}$ under parameter value $\boldsymbol{\theta}$ and innovation density f, the residuals $\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta})$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ are i.i.d. with density f.

The VARMA model (5.2.1) has no intercept—the observation yields no trend and is centered at the origin. Adding an intercept to the list of parameters would have no impact in the context of center-outward R-estimation since, as we shall see in Section 5.3.1, center-outward ranks and signs are shift-invariant. This is a ma-

jor advantage of center-outward R-estimators over their competitors (QMLEs and the elliptical R-estimators based on Mahalanobis ranks and signs of Section 5.3.2) which all require a preliminary centering of the residuals $\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})$. That centering is typically achieved by subtracting the residual mean $\overline{\boldsymbol{Z}}^{(n)}(\boldsymbol{\theta}) := n^{-1}\sum_{t=1}^n \boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})$ and has no asymptotic impact in view of the block-diagonal form of the information matrix (the location/trend and VARMA components of central sequences indeed are asymptotically mutually orthogonal). For the sake of notational simplicity, however, we are not formalizing that centering issue and avoid adding a d-dimensional intercept to the parameter $\boldsymbol{\theta}$.

5.2.2 LAN

Writing $L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta};f} := \log \mathrm{dP}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}^{(n)};f}/\mathrm{dP}^{(n)}_{\boldsymbol{\theta};f}$ for the log-likelihood ratio of $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}^{(n)};f}$ with respect to $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, where $\boldsymbol{\tau}^{(n)}$ is a bounded sequence of $\mathbb{R}^{(p+q)d^2}$, let

$$\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) := \boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}}\boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)\prime}\boldsymbol{\Gamma}_f^{(n)}(\boldsymbol{\theta}), \tag{5.2.3}$$

where $\boldsymbol{M}_{\boldsymbol{\theta}}$, $\boldsymbol{P}_{\boldsymbol{\theta}}$, and $\boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)}$ (see (6.1.2) and (6.1.3) in Section 6.1 for an explicit form) do not depend on f nor $\boldsymbol{\tau}^{(n)}$ and

$$\boldsymbol{\Gamma}_f^{(n)}(\boldsymbol{\theta}) := \big((n-1)^{1/2}(\mathrm{vec}\boldsymbol{\Gamma}_{1,f}^{(n)}(\boldsymbol{\theta}))', \ldots, (n-i)^{1/2}(\mathrm{vec}\boldsymbol{\Gamma}_{i,f}^{(n)}(\boldsymbol{\theta}))', \ldots, (\mathrm{vec}\boldsymbol{\Gamma}_{n-1,f}^{(n)}(\boldsymbol{\theta}))'\big)' \tag{5.2.4}$$

with the so-called *f-cross-covariance matrices*

$$\boldsymbol{\Gamma}_{i,f}^{(n)}(\boldsymbol{\theta}) := (n-i)^{-1}\sum_{t=i+1}^{n} \boldsymbol{\varphi}_f(\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}))\boldsymbol{Z}_{t-i}^{(n)\prime}(\boldsymbol{\theta}). \tag{5.2.5}$$

We then have the following LAN result (see Section 6.2 for a proof).

Proposition 5.2.1. *Let Assumptions (A1) and (A2) hold. Then, for any bounded sequence $\boldsymbol{\tau}^{(n)}$ in $\mathbb{R}^{(p+q)d^2}$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, as $n \to \infty$,*

$$L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta};f} = \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1) \tag{5.2.6}$$

with

$$\mathbf{\Lambda}_f(\boldsymbol{\theta}) := \boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}} \lim_{n\to\infty} \left\{ \boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)\prime} \left[\boldsymbol{I}_{n-1} \otimes \mathbf{\Xi} \otimes \boldsymbol{\mathcal{I}}(f) \right] \boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)} \right\} \boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}},$$

and $\mathbf{\Delta}_f^{(n)}(\boldsymbol{\theta})$ *is asymptotically normal, with mean* $\mathbf{0}$ *and variance* $\mathbf{\Lambda}_f(\boldsymbol{\theta})$*).*

The class $\mathcal{F}_d$ contains, among others, the elliptical densities. Recall that a d-dimensional random vector $\boldsymbol{Z}$ has centered elliptical distribution with *scatter matrix* $\mathbf{\Sigma}$ and *radial density* $\mathfrak{f}$ if its density has the form $f(\mathbf{z}) = \kappa_{d,\mathfrak{f}}^{-1}\big(\det\mathbf{\Sigma}\big)^{-1/2}\mathfrak{f}\big((\mathbf{z}'\mathbf{\Sigma}^{-1}\mathbf{z})^{1/2}\big)$ for some symmetric positive definite $\mathbf{\Sigma}$ and some function $\mathfrak{f}$: $\mathbb{R}^+ \to \mathbb{R}^+$ such that $\int_0^\infty r^{d-1}\mathfrak{f}(r)\,\mathrm{d}r < \infty$; $\kappa_{d,\mathfrak{f}} := \big(2\pi^{d/2}/\Gamma(d/2)\big)\int_0^\infty r^{d-1}\mathfrak{f}(r)\,\mathrm{d}r$ is a norming constant. When $\boldsymbol{Z}$ is elliptical with shape matrix $\mathbf{\Sigma}$ and radial density $\mathfrak{f}$, $\|\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}\|$ (where $\mathbf{\Sigma}^{1/2}$ stands for the symmetric root of $\mathbf{\Sigma}$) has density

$$f^\star_{d;\mathfrak{f}}(r) = (\mu_{d-1;\mathfrak{f}})^{-1} r^{d-1}\mathfrak{f}(r) I[r > 0]$$

where $\mu_{d-1;\mathfrak{f}} := \int_0^\infty r^{d-1}\mathfrak{f}(r)\mathrm{d}\,r$, and distribution function $F^\star_{d;\mathfrak{f}}$. Assumption (A2)(ii) on f then is equivalent to the mean square differentiability, with quadratic mean derivative $\mathrm{D}\mathfrak{f}^{1/2}$, of $r \mapsto \mathfrak{f}^{1/2}(r)$, $r \in \mathbb{R}_0^+$; letting $\varphi_{\mathfrak{f}} := -2\mathrm{D}\mathfrak{f}^{1/2}/\mathfrak{f}^{1/2}(r)$, we get $\mathcal{I}_{d;\mathfrak{f}} := \int_0^1 \Big(\varphi_{\mathfrak{f}} \circ \big(F^\star_{d;\mathfrak{f}}\big)^{-1}(u)\Big)^2 \mathrm{d}u < \infty$.

Elliptic random vectors admit the following representation in terms of spherical uniform variables. Denoting by $\mathbb{S}_d$ and $\mathcal{S}_{d-1}$ the open unit ball and the unit sphere in $\mathbb{R}^d$, respectively, define the spherical uniform distribution U_d over $\mathbb{S}_d$ as the product of the uniform measure over $\mathcal{S}_{d-1}$ with a uniform measure over the unit interval of distances to the origin. A d-dimensional random vector $\boldsymbol{Z}$ has centered elliptical distribution iff $F^\star_{d;\mathfrak{f}}(\|\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}\|)\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}/\|\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}\| \sim \mathrm{U}_d$. Putting $\boldsymbol{S}^{(n)}_{\mathbf{\Sigma},t} := \mathbf{\Sigma}^{-1/2}\boldsymbol{Z}_t^{(n)}/\|\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}_t^{(n)}\|$, where $\boldsymbol{Z}_t^{(n)} := \boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})$, it follows from Hallin and Paindaveine (2004) that the central sequence (5.2.3) for elliptical f considerably simplifies and takes the form (5.2.3) with

$$\mathbf{\Gamma}_{i,f}^{(n)}(\boldsymbol{\theta}) := (n-i)^{-1}\mathbf{\Sigma}^{-1/2}\sum_{t=i+1}^{n} \varphi_1(\|\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}_t^{(n)}\|)\varphi_2(\|\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}_{t-i}^{(n)}\|)\boldsymbol{S}^{(n)}_{\mathbf{\Sigma},t}\boldsymbol{S}^{(n)\prime}_{\mathbf{\Sigma},t-i}\mathbf{\Sigma}^{1/2} \tag{5.2.7}$$

where $\varphi_1(r) := \varphi_{\mathfrak{f}}(r)$ and $\varphi_2(r) := r$, $r \in \mathbb{R}^+$.

5.3 Center-outward ranks and signs

Parametrically optimal (in the Hájek-Le Cam asymptotic sense) rank-based inference procedures in LAN families is possible if the LAN central sequence can be expressed in terms of signs and ranks. In Section 5.3.4, we explain how to achieve this goal using the notions of multivariate ranks and signs proposed by Chernozhukov et al. (2017) (under the name of Monge-Kantorovich ranks and signs) and developed in Hallin (2017) and Hallin et al. (2020a) under the name of *center-outward ranks and signs*. This new concepts hinge on measure transportation theory; their empirical versions are based on an optimal coupling of the sample residuals $\mathbf{Z}_t^{(n)}$ with a regular grid over the unit ball.

5.3.1 Mapping the residuals to the unit ball

Let $\mathcal{P}_d$ denote the family of all distributions P with densities in $\mathcal{F}_d$—for this family the center-outward distribution functions defined below are continuous; see Hallin et al. (2020a). The *center-outward distribution function* $\boldsymbol{F}_{\pm}$ is defined as the a.e. unique gradient of convex function mapping $\mathbb{R}^d$ to $\mathbb{S}_d$ and pushing P forward to the spherical uniform distribution U_d. For $\mathrm{P} \in \mathcal{P}_d$, such mapping is a homeomorphism between $\mathbb{S}_d \setminus \{\mathbf{0}\}$ and $\mathbb{R}^d \setminus \boldsymbol{F}_{\pm}^{-1}(\{\mathbf{0}\})$ (Figalli 2018) and the corresponding *center-outward quantile function* is defined (letting, with a small abuse of notation, $\boldsymbol{Q}_{\pm}(\mathbf{0}) := \boldsymbol{F}_{\pm}^{-1}(\{\mathbf{0}\})$) as $\boldsymbol{Q}_{\pm} := \boldsymbol{F}_{\pm}^{-1}$. For any given distribution P, $\boldsymbol{Q}_{\pm}$ induces a collection of continuous, connected, and nested quantile contours and regions; the *center-outward median* $\boldsymbol{Q}_{\pm}(\mathbf{0})$ is a uniquely defined compact set of Lebesgue measure zero. We refer to Hallin et al. (2020a) for details.

Turning to the sample, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the residuals $\boldsymbol{Z}^{(n)}(\boldsymbol{\theta}) := (\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta}))$ under $\mathrm{P}_{\boldsymbol{\theta};f}^{(n)}$ are i.i.d. with density $f \in \mathcal{F}_d$ and center-outward distribution function $\boldsymbol{F}_{\pm}$. For the empirical counterpart $\boldsymbol{F}_{\pm}^{(n)}$ of $\boldsymbol{F}_{\pm}$, let n factorize into $n = n_R n_S + n_0$, for $n_R, n_S, n_0 \in \mathbb{N}$ and $0 \leq n_0 < \min\{n_R, n_S\}$, where $n_R \to \infty$ and $n_S \to \infty$ as $n \to \infty$, and consider a sequence of grids, where each grid consists of the intersection between an n_S-tuple $(\boldsymbol{u}_1, \ldots \boldsymbol{u}_{n_S})$ of unit vectors, and the n_R-hyperspheres

centered at the origin, with radii $1/(n_R+1), \ldots, n_R/(n_R+1)$, along with n_0 copies of the origin. The resulting grid is such that the discrete distribution with probability mass $1/n$ at each gridpoint and probability mass n_0/n at the origin converges weakly to the uniform U_d over the ball $\mathbb{S}_d$. Then, we define $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})$, for $t = 1, \ldots, n$ as the solution (optimal mapping) of a coupling problem between the residuals and the grid.

Specifically, the empirical center-outward distribution function is the (random) mapping

$$\boldsymbol{F}_{\pm}^{(n)} : \boldsymbol{Z}^{(n)} := (\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}) \mapsto (\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_1^{(n)}), \ldots, \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_n^{(n)}))$$

satisfying

$$\sum_{t=1}^{n} \|\boldsymbol{Z}_t^{(n)} - \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\|^2 = \min_{T \in \mathcal{T}} \sum_{t=1}^{n} \|\boldsymbol{Z}_t^{(n)} - T(\boldsymbol{Z}_t^{(n)})\|^2, \tag{5.3.1}$$

where $\boldsymbol{Z}_t^{(n)} = \boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})$, the set $\{\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) | t = 1, \ldots, n\}$ coincides with the n points of the grid, and $\mathcal{T}$ denotes the set of all possible bijective mappings between $\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$ and the n gridpoints. The sample counterpart of $\boldsymbol{Q}_{\pm}$ then is defined as $\boldsymbol{Q}_{\pm}^{(n)} := (\boldsymbol{F}_{\pm}^{(n)})^{-1}$ (again, with the small abuse of notation $\boldsymbol{Q}_{\pm}^{(n)}(\boldsymbol{0}) := (\boldsymbol{F}_{\pm}^{(n)})^{-1}(\{\boldsymbol{0}\})$). See Section 6.4 for a graphical illustration of these concepts.

Based on this empirical center-outward distribution function, the *center-outward ranks* and *signs* are

$$R_{\pm,t}^{(n)} := R_{\pm,t}^{(n)}(\boldsymbol{\theta}) := (n_R + 1)\|\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\|, \tag{5.3.2}$$

and (for $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) = \boldsymbol{0}$, let $\boldsymbol{S}_{\pm,t}^{(n)} := \boldsymbol{0}$)

$$\boldsymbol{S}_{\pm,t}^{(n)} := \boldsymbol{S}_{\pm,t}^{(n)}(\boldsymbol{\theta}) := \frac{\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})}{\|\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\|} I[\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) \neq \boldsymbol{0}], \tag{5.3.3}$$

respectively. It follows that $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})$ factorizes into

$$\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) = \frac{R_{\pm,t}^{(n)}}{n_R+1}\boldsymbol{S}_{\pm,t}^{(n)}, \quad \text{hence} \quad \boldsymbol{Z}_t^{(n)} = \boldsymbol{Q}_{\pm}^{(n)}\Big(\frac{R_{\pm,t}^{(n)}}{n_R+1}\boldsymbol{S}_{\pm,t}^{(n)}\Big). \tag{5.3.4}$$

Conditional on the grid (in case the latter is random), those ranks and signs are jointly distribution-free: more precisely, under $\mathrm{P}_{\boldsymbol{\theta};f}^{(n)}$, the n-tuple $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_1^{(n)}), \ldots, \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_n^{(n)})$ is uniformly distributed over the $n!$ permutations[2] of the n gridpoints, irrespective of $f \in \mathcal{F}_d$.

Empirical center-outward ranks and signs can be shown (Proposition 2.2 in Hallin et al. (2020b)) to enjoy the following invariance/equivariance properties. Denote by $\boldsymbol{F}_{\pm}^{\mathbf{Z}}$ the center-outward distribution function of $\boldsymbol{Z}$ and by $\boldsymbol{F}_{\pm}^{\mathbf{Z};(n)}$ the empirical distribution function, associated with some grid $\mathfrak{G}_n$, of the sample $\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$.

Proposition 5.3.1. *(Hallin et al. 2020b) Let $\boldsymbol{\mu} \in \mathbb{R}^d$ and denote by $\mathbf{O}$ a $d \times d$ orthogonal matrix. Then,*

(i) $\boldsymbol{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{OZ}}(\boldsymbol{\mu}+\mathbf{Oz}) = \mathbf{O}\boldsymbol{F}_{\pm}^{\mathbf{Z}}(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^d$;

(ii) denoting by $\boldsymbol{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{OZ};(n)}$ *the empirical distribution function of the sample*

$$\boldsymbol{\mu}+\mathbf{O}\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{\mu}+\mathbf{O}\boldsymbol{Z}_n^{(n)}$$

associated with the grid $\mathbf{O}\mathfrak{G}_n$ (hence, by $\boldsymbol{F}_{\pm}^{\mathbf{Z};(n)}$ the empirical distribution function of the sample $\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$ associated with the grid $\mathfrak{G}_n$),

$$\boldsymbol{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{OZ};(n)}(\boldsymbol{\mu}+\mathbf{O}\boldsymbol{Z}_i^{(n)}) = \mathbf{O}\boldsymbol{F}_{\pm}^{\mathbf{Z};(n)}(\boldsymbol{Z}_i^{(n)}), \quad i=1,\ldots,n; \tag{5.3.5}$$

(iii) the center-outward ranks $R_{i;\pm}^{(n)}$ and the cosines $\boldsymbol{S}_{i;\pm}^{(n)\prime}\boldsymbol{S}_{j;\pm}^{(n)}$ computed from $\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$ and the grid $\mathfrak{G}_n$ are the same as those computed from $\boldsymbol{\mu}+\mathbf{O}\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{\mu}+\mathbf{O}\boldsymbol{Z}_n^{(n)}$ and the grid $\mathbf{O}\mathfrak{G}_n$.

We refer to Sections 3.1 and 6 in Hallin et al. (2020a) for further details, comments on the main properties of $\boldsymbol{F}_{\pm}^{(n)}$ and $\boldsymbol{F}_{\pm}$, and for remarks on the minimal sufficiency and maximal ancillarity of the sub-σ-fields generated by the order

[2] Actually, for $n_0 > 1$, the $n!/n_0!$ permutations with repetitions.

statistic[3] and by the center-outward ranks and signs, in the fixed-$\boldsymbol{\theta}$ experiment $\{\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}|f \in \mathcal{F}_d\}$.

5.3.2 Mahalanobis ranks and signs

Definitions (5.3.2) and (5.3.3) call for a comparison with the earlier concepts of *elliptical* or *Mahalanobis ranks and signs* introduced in Hallin and Paindaveine (2002a and b, 2004), which we now describe. Associated with the centered elliptical distribution with scatter $\boldsymbol{\Sigma}$ and radial density $\mathfrak{f}$, consider the mapping

$$\boldsymbol{z} \mapsto \boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{z}) := F^{\star}_{d;\mathfrak{f}}(\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{z}\|)\boldsymbol{\Sigma}^{-1/2}\boldsymbol{z}/\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{z}\|$$

from $\mathbb{R}^d$ to $\mathbb{S}_d$. In measure transportation parlance, $\boldsymbol{F}_{\mathrm{ell}}$, just as $\boldsymbol{F}_{\pm}$, *pushes* the elliptical distribution of $\boldsymbol{Z}$ *forward to* the uniform U_d over the unit ball $\mathbb{S}_d$. This allows us to connect the Mahalanobis ranks and signs to the center-outward ones.

Denoting by $\widehat{\boldsymbol{\Sigma}}^{(n)}$ a consistent estimator of $\boldsymbol{\Sigma}$ measurable with respect to the order statistic[4] of the $\boldsymbol{Z}^{(n)}_t$'s and by $F^{\star(n)}$ the empirical distribution function of the moduli $\big\|\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}^{(n)}_t\big\|$, an empirical counterpart of $\boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t)$ is

$$\boldsymbol{F}^{(n)}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t) := F^{\star(n)}\big(\big\|\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}^{(n)}_t\big\|\big)\frac{\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}^{(n)}_t}{\big\|\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}^{(n)}_t\big\|} \tag{5.3.6}$$

with the *Mahalanobis ranks* (compare to (5.3.2))

$$R^{(n)}_{\mathrm{ell},t} := R^{(n)}_{\mathrm{ell},t}(\boldsymbol{\theta}) := (n+1)\|\boldsymbol{F}^{(n)}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t)\| = (n+1)F^{\star(n)}\big(\big\|\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}^{(n)}_t\big\|\big) \tag{5.3.7}$$

[3] An *order statistic* $\boldsymbol{Z}^{(n)}_{(\cdot)}$ of the un-ordered n-tuple $\boldsymbol{Z}^{(n)}$ is an arbitrarily ordered version of the same; see Section 6.4 and Hallin et al. (2020a). For instance, one may consider the ordering based on the first components.

[4] That is, a symmetric function of the $\boldsymbol{Z}_t$'s.

and *Mahalanobis signs* (compare to (5.3.3))

$$\begin{aligned}\boldsymbol{S}_{\text{ell},t}^{(n)} := \boldsymbol{S}_{\text{ell},t}^{(n)}(\boldsymbol{\theta}) &:= \frac{\boldsymbol{F}_{\text{ell}}^{(n)}(\boldsymbol{Z}_t^{(n)})}{\|\boldsymbol{F}_{\text{ell}}^{(n)}(\boldsymbol{Z}_t^{(n)})\|} I[\boldsymbol{F}_{\text{ell}}^{(n)}(\boldsymbol{Z}_t^{(n)}) \neq \mathbf{0}] \\ &= \frac{\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}_t^{(n)}}{\big\|\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}_t^{(n)}\big\|} I\big[\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}_t^{(n)} \neq \mathbf{0}\big] \end{aligned} \tag{5.3.8}$$

(for $\boldsymbol{F}_{\text{ell}}^{(n)}(\boldsymbol{Z}_t^{(n)}) = \mathbf{0} = \big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\boldsymbol{Z}_t^{(n)}$, let $\boldsymbol{S}_{\text{ell},t}^{(n)} := \mathbf{0}$). Similar to (5.3.4)), we have

$$\boldsymbol{F}_{\text{ell}}^{(n)}(\boldsymbol{Z}_t^{(n)}) = \frac{R_{\text{ell},t}^{(n)}}{n+1}\boldsymbol{S}_{\text{ell},t}^{(n)},$$

hence

$$\widehat{\boldsymbol{\Sigma}}^{(n)-1/2}\boldsymbol{Z}_t^{(n)} = (F^{\star(n)})^{-1}\left(\frac{R_{\text{ell},t}^{(n)}}{n+1}\right)\boldsymbol{S}_{\text{ell},t}^{(n)} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Z}_t^{(n)} + o_{\text{P}}(1).$$

5.3.3 Elliptical $\boldsymbol{F}_{\text{ell}}$, center-outward $\boldsymbol{F}_{\pm}$, and affine invariance

Both $\boldsymbol{F}_{\text{ell}}$ and $\boldsymbol{F}_{\pm}$ are *pushing* the elliptical distribution of $\boldsymbol{Z}$ *forward to* U_d. However, unless $\boldsymbol{\Sigma}$ is proportional to identity ($\boldsymbol{\Sigma} = c\boldsymbol{I}_d$ for some $c > 0$), $\boldsymbol{F}_{\text{ell}}$ and $\boldsymbol{F}_{\pm}$ are distinct, so that $\boldsymbol{F}_{\text{ell}}$ cannot be the gradient of a convex function. Moreover, both $\boldsymbol{F}_{\text{ell}}$ and $\boldsymbol{F}_{\pm}$ *sphericize* the distribution of $\boldsymbol{Z}$. Some key differences are worth to be mentioned, though.

First, while sphericization and probability integral transformation, in $\boldsymbol{F}_{\pm}$, are inseparably combined, $\boldsymbol{F}_{\text{ell}}$ proceeds in two separate steps: a *Mahalanobis sphericization* step (the parametric affine transformation $\boldsymbol{z} \mapsto \boldsymbol{z}_{\boldsymbol{\Sigma},\boldsymbol{\mu}} := \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{z} - \boldsymbol{\mu})$) first, followed by the spherical probability integral transformation

$$\boldsymbol{z}_{\boldsymbol{\Sigma},\boldsymbol{\mu}} \mapsto F_{d;\mathfrak{f}}^{\star}(\|\boldsymbol{z}_{\boldsymbol{\Sigma},\boldsymbol{\mu}}\|)\boldsymbol{z}_{\boldsymbol{\Sigma},\boldsymbol{\mu}}/\|\boldsymbol{z}_{\boldsymbol{\Sigma},\boldsymbol{\mu}}\|.$$

Second, Mahalanobis sphericization requires centering, hence the definition of a location parameter $\boldsymbol{\mu}$. Distinct choices of location (mean, spatial median, etc.) all yield the same result under ellipticity, but not under non-elliptical distributions. This is in sharp contrast with $\boldsymbol{F}_{\pm}$, which is location-invariant (see Hallin et

al. (2020b)). Similarly, all definitions and sensible estimators of the scatter yield the same results under elliptical symmetry but not under non-elliptical distributions.

Third, even under additional assumptions ensuring the identification of $\mathbf{\Sigma}$, the corresponding Mahalanobis sphericization, hence also $\boldsymbol{F}_{\text{ell}}$, only sphericizes the elliptical distributions, whilst $\boldsymbol{F}_{\pm}$ sphericizes them all.

Its preliminary Mahalanobis sphericization step, on the other hand, makes $\boldsymbol{F}_{\text{ell}}$ affine-invariant. Assuming that sensible choices of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are available, performing the same Mahalanobis transformation prior to determining $\boldsymbol{F}_{\pm}$ similarly would make center-outward distribution functions affine-invariant and the corresponding center-outward quantile functions affine-equivariant (in fact, for elliptical distributions, the resulting $\boldsymbol{F}_{\pm}$ then coincides with $\boldsymbol{F}_{\text{ell}}$). Whether this is desirable is a matter of choice. While affine-invariance, in view of the central role of the affine group in elliptical families, is quite natural under elliptical symmetry, its relevance is much less obvious away from ellipticity. A more detailed discussion of this fact, along with additional arguments related to the lack of affine invariance of non-elliptical local experiments, can be found in Hallin et al. (2020b).

Center-outward distribution functions, ranks, and signs, however, enjoy invariance/equivariance with respect to shift, global scale factors, and orthogonal transformations—see Proposition 5.3.1.

5.3.4 A center-outward sign- and rank-based central sequence

Efficient estimation in LAN experiments is based on central sequences and the so-called *Le Cam one-step method*. Our R-estimation is based on the same principles. Specifically, in the central sequence associated with some reference density f, we replace the residuals $\boldsymbol{Z}(\boldsymbol{\theta})$ with some adequate function of their ranks and their signs. Then, from the resulting rank-based statistic, we implement a suitable adaptation of the one-step method. If, under innovation density f, the substitu-

tion yields a genuinely rank-based, hence distribution-free, version of the central sequence, the resulting R-estimator achieves parametric efficiency under f while remaining valid under other innovation densities; see Hallin and Werker (2003) for a discussion.

In dimension $d = 1$, Allal et al. (2001), Hallin and La Vecchia (2017, 2019 and references therein) explain how to construct R-estimators for linear and nonlinear semiparametric time series models. In dimension $d > 1$, under elliptical innovations density, Hallin et al. (2006) exploit similar ideas for the estimation of shape matrices, based on the Mahalanobis ranks and signs. Hallin and Paindaveine (2004), in a hypothesis testing context, show that replacing $\boldsymbol{Z}_t^{(n)}$ in (5.2.7) with

$$\widehat{\boldsymbol{\Sigma}}^{(n)1/2} F_{d;\mathfrak{f}}^{\star -1}(R_{\text{ell},t}^{(n)}/(n+1))\boldsymbol{S}_{\text{ell},t}^{(n)} = \boldsymbol{F}_{\text{ell}}^{-1}((R_{\text{ell},t}^{(n)}/(n+1))\boldsymbol{S}_{\text{ell},t}^{(n)}) = \boldsymbol{F}_{\text{ell}}^{-1}(\boldsymbol{F}_{\text{ell}}^{(n)}(\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}))) \tag{5.3.9}$$

(where $R_{\text{ell},t}^{(n)} = R_{\text{ell},t}^{(n)}(\boldsymbol{\theta})$, $\boldsymbol{S}_{\text{ell},t}^{(n)} = \boldsymbol{S}_{\text{ell},t}^{(n)}(\boldsymbol{\theta})$, and $\widehat{\boldsymbol{\Sigma}}^{(n)}$ is a suitable estimator of the scatter matrix) yields a rank-based version of the central sequence associated with the elliptic density f—namely, a random vector $\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\boldsymbol{\theta})$ measurable with respect to the Mahalanobis ranks and signs (hence, distribution-free under ellipticity) such that, under f, $\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\boldsymbol{\theta}) - \boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) = o_{\text{P}}(1)$ as $n \to \infty$.

However, this construction is valid only for the family of elliptical innovation densities (in dimension one, the family of symmetric innovation densities), under which Mahalanobis ranks and signs are distribution-free. Elliptic symmetry is a severe limitation, which is unlikely to be satisfied in most applications. If the attractive properties of R-estimators in univariate semiparametric time series models are to be extended to dimension two and higher, center-outward ranks and signs, the distribution-freeness of which holds under any density $f \in \mathcal{F}_d$, are to be considered instead of the Mahalanobis ones.

Building on this remark, we propose to substitute $\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})$ in (5.2.5) with

$$\boldsymbol{F}_{\pm}^{-1}((R_{\pm,t}^{(n)}/(n_R+1))\boldsymbol{S}_{\pm,t}^{(n)}) = \boldsymbol{F}_{\pm}^{-1}(\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}))) = \boldsymbol{Q}_{\pm} \circ \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})), \tag{5.3.10}$$

where $R_{\pm,t}^{(n)} = R_{\pm,t}^{(n)}(\boldsymbol{\theta})$, $\boldsymbol{S}_{\pm,t}^{(n)} = \boldsymbol{S}_{\pm,t}^{(n)}(\boldsymbol{\theta})$, and $\boldsymbol{F}_{\pm}$ and $\boldsymbol{Q}_{\pm}$ are associated with some chosen reference innovation density $f \in \mathcal{F}_d$. This yields rank-based, hence distribution-free, f-cross-covariance matrices of the form $(i = 1, \ldots, n-1)$

$$\underset{\sim}{\boldsymbol{\Gamma}}_{i,f}^{(n)}(\boldsymbol{\theta}) := (n-i)^{-1} \sum_{t=i+1}^{n} \boldsymbol{\varphi}_f \left(\boldsymbol{F}_{\pm}^{-1} \left(\frac{R_{\pm,t}^{(n)}}{n_R+1} \boldsymbol{S}_{\pm,t}^{(n)} \right) \right) \boldsymbol{F}_{\pm}^{-1\prime} \left(\frac{R_{\pm,t-i}^{(n)}}{n_R+1} \boldsymbol{S}_{\pm,t-i}^{(n)} \right). \tag{5.3.11}$$

While this looks quite straightforward, practical implementation requires an analytical expression for $\boldsymbol{F}_{\pm}$, which typically is unavailable for general innovation densities. And, were such closed forms available, the problem of choosing an adequate multivariate reference density f remains.

Now, note that in the univariate case all standard reference densities are symmetric—think of Gaussian, logistic, double-exponential densities, leading to van der Waerden, Wilcoxon, or sign test scores. Therefore, in the sequel, we concentrate on rank-based cross-covariance matrices of the form $(i = 1, \ldots, n-1)$

$$\underset{\sim}{\boldsymbol{\Gamma}}_{i,J_1,J_2}^{(n)}(\boldsymbol{\theta}) := (n-i)^{-1} \sum_{t=i+1}^{n} J_1 \left(\frac{R_{\pm,t}^{(n)}}{n_R+1} \right) J_2 \left(\frac{R_{\pm,t-i}^{(n)}}{n_R+1} \right) \boldsymbol{S}_{\pm,t}^{(n)} \boldsymbol{S}_{\pm,t-i}^{(n)\prime} \tag{5.3.12}$$

to which $\underset{\sim}{\boldsymbol{\Gamma}}_{i,f}^{(n)}(\boldsymbol{\theta})$ in (5.3.11) reduces, with $J_1(u) = \varphi_{\mathfrak{f}}(F_{d;\mathfrak{f}}^{\star-1}(u))$ and $J_2(u) = F_{d;\mathfrak{f}}^{\star-1}(u)$, in the case of a spherical reference f with radial density $\mathfrak{f}$, yielding a rank-based version $\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}$ of the spherical central sequence $\boldsymbol{\Delta}_f^{(n)}$. More generally, we propose to use statistics of the form (5.3.12) with scores $J_1 : [0,1) \to \mathbb{R}$ and $J_2 : [0,1) \to \mathbb{R}$ which are not necessarily related to any spherical density. Then, the notation $\underset{\sim}{\boldsymbol{\Delta}}_{J_1,J_2}^{(n)}$ will be used in an obvious fashion, indicating that $\underset{\sim}{\boldsymbol{\Delta}}_{J_1,J_2}^{(n)}$ needs not be a central sequence.

The next section provides details on the choice of J_1 and J_2 and establishes the asymptotic properties (root-n consistency and asymptotic normality) of the related R-estimators.

5.4 R-estimation

5.4.1 One-step R-estimators: definition and asymptotics

We now proceed with a precise definition of our R-estimators and establish their asymptotic properties. Throughout, J_1 and J_2 are assumed to satisfy the following assumption.

Assumption (A3). The score functions J_1 and J_2 in (5.3.12) *(i)* are square-integrable, that is, $\sigma^2_{J_l} := \int_0^1 J_l^2(r)\mathrm{d}r < \infty$, $l = 1, 2$, and *(ii)* are continuous differences of two monotonic increasing functions.

Assumption (A3) is quite mild and it is satisfied, e.g., by all square-integrable functions with bounded variation. Define $\boldsymbol{J}_{J_2,f} := \int_{\mathbb{S}_d} J_2(\|\boldsymbol{u}\|)(\boldsymbol{u}/\|\boldsymbol{u}\|)\boldsymbol{F}_{\pm}^{-1\prime}(\boldsymbol{u})\mathrm{dU}_d(\boldsymbol{u})$, and

$$\boldsymbol{K}_{J_1,J_2,f} := \int_{\mathbb{S}_d} J_1(\|\boldsymbol{u}\|)\left[\boldsymbol{I}_d \otimes \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}\right]\boldsymbol{J}_{J_2,f}\left[\boldsymbol{I}_d \otimes \boldsymbol{\varphi}'_f\Big(\boldsymbol{F}_{\pm}^{-1}(\boldsymbol{u})\Big)\right]\mathrm{dU}_d(\boldsymbol{u}). \quad (5.4.1)$$

These two matrices under Assumptions (A2) and (A3) exist and are finite in view of the Cauchy–Schwarz inequality since $\boldsymbol{u}/\|\boldsymbol{u}\|$ is bounded.

R-estimation requires the asymptotic linearity of the rank-based objective function involved. Sufficient conditions for such linearity are available in the literature (see e.g. Jurečková (1971) and van Eeden (1972), Hallin and Puri (1994), Hallin and Paindaveine (2005) or Hallin et al. (2015)). In the same spirit, we introduce the following assumption on the rank-based statistics $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})$; the form of the linear term in the right-hand side of (5.4.2) follows from the form of the asymptotic shift in Lemma 6.2.3.

Decomposing the matrix $\boldsymbol{Q}^{(n)}_{\boldsymbol{\theta}}$ defined in (6.1.3) into $d^2 \times d^2(p+q)$ blocks (note that these blocks do not depend on n), write $\boldsymbol{Q}^{(n)}_{\boldsymbol{\theta}} = \big(\boldsymbol{Q}'_{1,\boldsymbol{\theta}} \cdots \boldsymbol{Q}'_{n-1,\boldsymbol{\theta}}\big)'$ and consider the following assumption.

Assumption (A4) For any positive integer i and $d^2(p+q)$-dimensional vector $\boldsymbol{\tau}$,

under actual density f, as $n \to \infty$

$$(n-i)^{1/2}\left[\mathrm{vec}(\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}))-\mathrm{vec}(\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))\right]=-\boldsymbol{K}_{J_1,J_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}+o_{\mathrm{P}}(1), \tag{5.4.2}$$

where $\mathbf{M}_{\boldsymbol{\theta}}$ and $\mathbf{P}_{\boldsymbol{\theta}}$, which do not depend on f, J_1 nor J_2, are given in (6.1.2) and (6.1.3) in Section 6.1.

Next, for $m \leq n-1$, consider

$$\underset{\sim}{\mathbf{\Gamma}}^{(m,n)}_{J_1,J_2}(\boldsymbol{\theta}):=((n-1)^{1/2}(\mathrm{vec}\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{1,J_1,J_2}(\boldsymbol{\theta}))',\ldots,(n-m)^{1/2}(\mathrm{vec}\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta}))')', \tag{5.4.3}$$

and the truncated version

$$\underset{\sim}{\mathbf{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta}):=\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}\underset{\sim}{\mathbf{\Gamma}}^{(m,n)}_{J_1,J_2}(\boldsymbol{\theta}) \quad \text{where} \quad \boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}:=\boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}}\boldsymbol{Q}^{(m+1)\prime}_{\boldsymbol{\theta}} \tag{5.4.4}$$

of $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$. This truncation is just a theoretical device required in the statement of asymptotic results and, as explained in Section 6.3.2, there is no need to implement it in practice. The asymptotic linearity (5.4.2) of $\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})$ entails, for $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$, the following result.

Proposition 5.4.1. *Let Assumptions (A1), (A2), (A3), and (A4) hold. Then, for any (m,n) such that $m \leq n-1$ and $m \to \infty$ (hence also $n \to \infty$),*

$$\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau})-\underset{\sim}{\mathbf{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta})=-\mathbf{\Upsilon}^{(m+1)}_{J_1,J_2,f}(\boldsymbol{\theta})\boldsymbol{\tau}+o_{\mathrm{P}}(1), \tag{5.4.5}$$

where $\mathbf{\Upsilon}^{(m+1)}_{J_1,J_2,f}(\boldsymbol{\theta}):=\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}(\boldsymbol{I}_m\otimes\boldsymbol{K}_{J_1,J_2,f})\boldsymbol{T}^{(m+1)\prime}_{\boldsymbol{\theta}}$.

With the above asymptotic linearity result, we are now ready to define our R-estimators. First, let us introduce some notations. Under Assumption (A1), let $\mathbf{\Upsilon}_{J_1,J_2,f}(\boldsymbol{\theta}):=\lim_{n\to\infty}\mathbf{\Upsilon}^{(n)}_{J_1,J_2,f}(\boldsymbol{\theta})$ and define the *cross-information matrix*

$$\mathbf{I}_{J_1,J_2,f}(\boldsymbol{\theta}):=\lim_{n\to\infty}\mathrm{E}_{\boldsymbol{\theta},f}\left[\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})\mathbf{\Delta}^{(n)}_{f}(\boldsymbol{\theta})'\right]. \tag{5.4.6}$$

Let

$$\bar{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}) := (n-i)^{-1} \sum_{t=i+1}^{n} J_1(\|\boldsymbol{F}_{\pm,t}\|) J_2(\|\boldsymbol{F}_{\pm,t-i}\|) \boldsymbol{S}_{\pm,t} \boldsymbol{S}'_{\pm,t-i} \tag{5.4.7}$$

with $\boldsymbol{S}_{\pm,t} := \boldsymbol{F}_{\pm,t}/\|\boldsymbol{F}_{\pm,t}\|$ representing the "sign" of $\boldsymbol{F}_{\pm,t} := \boldsymbol{F}_{\pm}(\boldsymbol{Z}^{(n)}_t(\boldsymbol{\theta}))$. Denote by $\bar{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$ the central sequence resulting from substituting $\bar{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})$ for $\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})$ in $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$. Following the proofs in Lemma 6.2.4 and Lemma 6.2.3, it is easy to see that the difference between $\bar{\mathbf{\Delta}}^{(n)}_{J_1,J_2}$ and $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}$ converges to zero in quadratic mean as $n \to \infty$. Therefore, $\mathbf{\Upsilon}_{J_1,J_2,f}(\boldsymbol{\theta})$ coincides with the cross-information matrix (5.4.6) when Assumptions (A1), (A2) and (A3) hold; see the proof of Lemmas 6.2.1 and 6.2.4 in Section 6.2.

Let $\hat{\mathbf{\Upsilon}}^{(n)}_{J_1,J_2}$ denote a consistent (under innovation density f) estimator of $\mathbf{\Upsilon}_{J_1,J_2,f}(\boldsymbol{\theta})$; such an estimator is provided in (5.4.5), see Section 6.3 for details. Also, denote by $\hat{\boldsymbol{\theta}}^{(n)}$ a preliminary root-n consistent and asymptotically discrete[5] estimator of $\boldsymbol{\theta}$. Our one-step R-estimator then is defined as

$$\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)} := \hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2} \left(\hat{\mathbf{\Upsilon}}^{(n)}_{J_1,J_2}\right)^{-1} \underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\hat{\boldsymbol{\theta}}^{(n)}). \tag{5.4.8}$$

The following proposition establishes its root-n consistency and asymptotic normality.

Proposition 5.4.2. *Let Assumptions (A1), (A2), (A3), and (A4) hold. Let*

$$\mathbf{\Omega}^{(n)} := d^{-2}\sigma^2_{J_1}\sigma^2_{J_2} \left(\mathbf{\Upsilon}^{(n)}_{J_1,J_2,f}(\boldsymbol{\theta})\right)^{-1} \boldsymbol{T}^{(n)}_{\boldsymbol{\theta}} \boldsymbol{T}^{(n)\prime}_{\boldsymbol{\theta}} \left(\mathbf{\Upsilon}^{(n)\prime}_{J_1,J_2,f}(\boldsymbol{\theta})\right)^{-1}.$$

Then, denoting by $\left(\mathbf{\Omega}^{(n)}\right)^{-1/2}$ *the symmetric square root of* $\mathbf{\Omega}^{(n)}$,

$$n^{1/2}\left(\mathbf{\Omega}^{(n)}\right)^{-1/2}(\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)} - \boldsymbol{\theta}) \to \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{d^2(p+q)}), \tag{5.4.9}$$

[5] Asymptotic discreteness is only a theoretical requirement since, in practice, $\hat{\boldsymbol{\theta}}^{(n)}$ anyway only has a bounded number of digits; see Le Cam and Yang (2000, Chapter 6) and van der Vaart (1998, Section 5.7) for details.

under innovation density f, as both n_R and n_S tend to infinity.

See Section 6.2 for the proof. Section 6.3 discusses the computational aspects of the procedure and describes the algorithm we are using. Codes are available from the authors' `GitHub` page `https://github.com/HangLiu10/RestVARMA`.

5.4.2 Some standard score functions

The rank-based cross-covariance matrices $\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{J_1,J_2}$, hence also the resulting R-estimator, depend on the choice of score functions J_1 and J_2. We provide three examples of sensible choices extending scores that are widely applied in the univariate (see e.g. Hallin and La Vecchia (2020)) and the elliptical multivariate setting (see Hallin and Pandaveine (2004)).

Example 1 (*Sign test* scores). Setting $J_1(u) = 1 = J_2(u)$ yields the center-outward sign-based cross-covariance matrices

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\text{sign}}(\boldsymbol{\theta}) = (n-i)^{-1} \sum_{t=i+1}^{n} \boldsymbol{S}^{(n)}_{\pm,t}(\boldsymbol{\theta}) \boldsymbol{S}^{(n)\prime}_{\pm,t-i}(\boldsymbol{\theta}), \quad i = 1, \ldots, n-1. \tag{5.4.10}$$

The resulting $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\text{sign}}(\boldsymbol{\theta})$ entirely relies on the center-outward signs $\boldsymbol{S}^{(n)}_{\pm,t}(\boldsymbol{\theta})$, which should make them particularly robust and explains the terminology *sign test* scores.

Example 2 (*Spearman* scores). Another simple choice is $J_1(u) = J_2(u) = u$. The corresponding rank-based cross-covariance matrices are

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\text{Sp}}(\boldsymbol{\theta}) = (n-i)^{-1} \sum_{t=i+1}^{n} \boldsymbol{F}^{(n)}_{\pm,t} \boldsymbol{F}^{(n)\prime}_{\pm,t-i}, \quad i = 1, \ldots, n-1, \tag{5.4.11}$$

with $\boldsymbol{F}^{(n)}_{\pm,t} := \boldsymbol{F}^{(n)}_{\pm}(\mathbf{Z}^{(n)}_t(\boldsymbol{\theta}))$, reducing, for $d = 1$, to Spearman autocorrelations, whence the terminology *Spearman* scores.

Example 3 (*van der Waerden* or *normal* scores). Finally, $J_1(u) = J_2(u) = \left((F^{\chi^2}_d)^{-1}(u)\right)^{1/2}$, where $F^{\chi^2}_d$ denotes the chi-square distribution function with d degrees of freedom, yields the *van der Waerden* (vdW) *rank scores*, with cross-

covariance matrices

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\mathrm{vdW}}(\boldsymbol{\theta}) = (n-i)^{-1}\sum_{t=i+1}^{n}\left[\left(F_d^{\chi^2}\right)^{-1}\left(\frac{R^{(n)}_{\pm,t}(\boldsymbol{\theta})}{n_R+1}\right)\right]^{1/2}\left[\left(F_d^{\chi^2}\right)^{-1}\left(\frac{R^{(n)}_{\pm,t-i}(\boldsymbol{\theta})}{n_R+1}\right)\right]^{1/2} \times \boldsymbol{S}^{(n)}_{\pm,t}(\boldsymbol{\theta})\boldsymbol{S}^{(n)\prime}_{\pm,t-i}(\boldsymbol{\theta}), \quad i=1,\ldots,n-1. \tag{5.4.12}$$

Adequate choices of J_1 and J_2, namely,

$$J_1 = \varphi_{\mathfrak{f}} \circ \left(F^{\star}_{d;\mathfrak{f}}\right)^{-1} \quad \text{and} \quad J_2 = \left(F^{\star}_{d;\mathfrak{f}}\right)^{-1}, \tag{5.4.13}$$

yield asymptotic efficiency of $\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)}$ under spherical distributions with radial density $\mathfrak{f}$. Indeed, it is shown in Chernozhukov et al. (2017) that, for spherical distributions, $\boldsymbol{F}_{\pm}$ actually coincides with $\boldsymbol{F}_{\mathrm{ell}}$. Hence, $\bar{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}$, under spherical density f, coincides with the central sequence $\boldsymbol{\Delta}^{(n)}_f$. Therefore, due to the convergence in quadratic mean of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}$ to $\bar{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}$, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}$ and $\boldsymbol{\Delta}^{(n)}_f$ are asymptotically equivalent and $\boldsymbol{\Upsilon}_{J_1,J_2,f}(\boldsymbol{\theta})$ coincides with the Fisher information matrix; $\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)}$ then achieves (parametric) asymptotic efficiency under innovation density f.

Condition (5.4.13) is satisfied by the van der Waerden scores for Gaussian $\mathfrak{f}$: the corresponding R-estimator, thus, is parametrically efficient under spherical Gaussian innovations. If the residuals are sphericized prior to the computation of center-outward ranks and signs, then parametric efficiency is reached under any Gaussian innovation density; we have explained in Section 5.3.3 why this may be desirable or not. Neither the Spearman nor the sign test scores satisfy (5.4.13) for any $\mathfrak{f}$. Efficiency, however, is just one possible criterion for the selection of J_1 and J_2 and many alternative options are available, based on ease-of-implementation (as in Examples 1 and 2) or robustness (as in Example 1).

5.5 Numerical illustration

A numerical study of the performance of our R-estimators was conducted in dimensions $d=2$ (Sections 5.5.1, 5.5.2, and 5.5.3) and $d=3$ (Section 5.5.4). Further

results are available in Section 6.4 of Chapter 6.

In dimension $d = 2$, we considered the bivariate VAR(1) model

$$(\boldsymbol{I}_d - \boldsymbol{A}L)\,\boldsymbol{X}_t = \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z} \tag{5.5.1}$$

with the same parameter of interest $\boldsymbol{\theta} := \mathrm{vec}\boldsymbol{A} = (a_{11}, a_{21}, a_{12}, a_{22})' = (0.2, -0.6, 0.3, 1.1)'$ as in the motivating example of Section 5.1.2 and spherical Gaussian, spherical t_3, skew-normal, skew-t_3, Gaussian mixture, and non-spherical Gaussian innovations, respectively. The skew-normal and skew-t_3 distributions are described in Section 6.4.2 of Chapter 6; the Gaussian mixture is of the form

$$\frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + \frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) + \frac{1}{4}\mathcal{N}(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3), \tag{5.5.2}$$

with $\boldsymbol{\mu}_1 = (-5, 0)'$, $\boldsymbol{\mu}_2 = (5, 0)'$, $\boldsymbol{\mu}_3 = (0, 0)'$, $\boldsymbol{\Sigma}_1 = \begin{pmatrix} 7 & 5 \\ 5 & 5 \end{pmatrix}$, $\boldsymbol{\Sigma}_2 = \begin{pmatrix} 7 & -6 \\ -6 & 6 \end{pmatrix}$, and $\boldsymbol{\Sigma}_3 = \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix}$. A scatterplot of $n = 1000$ innovations drawn from this mixture is shown in Section 6.4.2. For the non-spherical Gaussian case, as in the last panel of Table 5.1, we set the covariance matrix to $\boldsymbol{\Sigma}_4 = \begin{pmatrix} 5 & 4 \\ 4 & 4.5 \end{pmatrix}$, so that the bivariate innovation exhibits a large positive correlation of 0.843.

All these densities have mean zero. As mentioned at the end of Section 5.2.1, this is not required for center-outward R-estimation, but avoids a preliminary centering of residuals in all other estimation methods.

For each of these innovation densities, we generated $N = 300$ Monte Carlo realizations—larger values of N did not show significant changes— of the stationary solution of (5.5.1), of length $n = 1000$ (n "large": Section 5.5.1) and $n = 300$ (n "small": Section 5.5.2), respectively. For each realization, we computed the QMLE, our R-estimators (sign test, Spearman, and van der Waerden scores), and, for the purpose of comparison, the QMLE based on t_5 likelihood (although

inconsistent, QMLEs based on t-distribution are a popular choice in the time series literature) and the reweighted multivariate least trimmed squares estimator (henceforth, RMLTSE) of Croux and Joossen (2008). The boxplots and tables of bias and mean squared errors below allow for a comparison of the finite-sample performance of our R-estimators and those routinely-applied M-estimators.

Throughout, QMLEs were computed from the `MTS` package in R program, RMLTSEs were computed from the function `varxfit` in the package `rmgarch` in R program, and t_5-QMLEs were obtained by minimizing the negative log-likelihood function using the `optim` function in R program. The R-estimators were obtained via the one-step procedure as in the algorithm described in Section 6.3 of Chapter 6—five iterations for $n = 1000$, ten iterations for $n = 300$.

5.5.1 Large sample results

The averaged bias and MSE of each estimator for $n = 1000$ (factorizing into $n_R n_S = 25 \times 40$) are summarized in Table 5.1, where ratios of the sums (over the four parameters) of the MSEs of the QMLE over those of each of the other estimators are also reported. The corresponding boxplots under the skew-normal (Figure 6.2), skew-t_3 (Figure 6.3)), spherical t_3 (Figure 6.4) and non-spherical Gaussian (Figure 6.5) innovations are provided in Section 6.4.3.1 of Chapter 6.

Inspection of Table 5.1 reveals that under asymmetric innovation densities (mixture, skew-normal and skew-t_3), the vdW and Spearman R-estimators dominate the other three M-estimators, with significant efficiency gains under the mixture and skew-t_3 distributions. One may wonder what happens if asymmetry is removed and only the heavy-tail feature is kept. The MSE ratios under the spherical t_3 distribution answer this question: the R-estimators still outperform the QMLE. Recalling that asymptotic optimality can be achieved by our R-estimators under spherical densities, it would be interesting to investigate their performance under a non-spherical distribution with large correlation. The MSE ratios under the non-spherical Gaussian distribution show that the vdW and Spearman

Table 5.1: The estimated bias, MSE, and overall MSE ratios of the QMLE, t_5-QMLE , RMLTSE, and R-estimators under various innovation densities. The true parameter is $\text{vec}(\boldsymbol{A}) = (0.2, -0.6, 0.3, 1.1)'$. The sample size is $n = 1000$; $N = 300$ replications. (The bias and MSE are multiplied by 1000).

	Bias ($\times 10^3$)				MSE ($\times 10^3$)				MSE ratio
	a_{11}	a_{21}	a_{12}	a_{22}	a_{11}	a_{21}	a_{12}	a_{22}	
(Spherical Normal)									
QMLE	-0.484	-0.054	0.201	-1.571	0.769	0.679	0.173	0.195	
t_5-QMLE	-0.547	-0.132	0.429	-1.582	0.833	0.751	0.190	0.210	0.916
RMLTS	-0.629	-0.992	0.424	-1.334	0.843	0.760	0.193	0.215	0.903
vdW	-0.662	-0.434	0.504	-1.833	0.780	0.688	0.178	0.205	0.982
Spearman	-1.263	-0.979	1.274	-2.134	0.810	0.728	0.189	0.216	0.935
Sign	-0.372	-0.600	1.545	-2.642	1.314	1.141	0.305	0.310	0.592
(Mixture)									
QMLE	-1.318	-0.476	2.907	-0.103	0.839	0.153	0.342	0.056	
t_5-QMLE	-0.852	0.483	4.820	0.248	4.420	0.261	1.641	0.156	0.215
RMLTS	-0.703	0.268	3.166	-0.116	0.876	0.168	0.351	0.069	0.949
vdW	-1.111	-0.465	2.347	-0.883	0.316	0.085	0.149	0.042	2.346
Spearman	-0.841	-0.539	2.338	-0.791	0.291	0.088	0.140	0.041	2.480
Sign	-1.691	0.048	5.256	-1.425	1.332	0.149	0.564	0.074	0.656
(Skew-normal)									
QMLE	-0.992	1.800	0.651	-2.108	0.804	1.039	0.281	0.311	
t_5-QMLE	-0.378	2.588	-0.083	-2.827	1.000	1.294	0.365	0.397	0.797
RMLTS	-0.519	1.515	0.172	-2.383	0.835	1.111	0.295	0.333	0.946
vdW	-1.031	0.990	0.811	-2.520	0.668	0.998	0.214	0.291	1.122
Spearman	-1.295	0.625	0.848	-2.171	0.694	1.032	0.222	0.294	1.086
Sign	-1.608	0.888	1.346	-4.039	1.360	1.673	0.415	0.519	0.614
(Skew-t_3)									
QMLE	-2.242	-2.055	0.763	0.213	1.022	0.856	0.379	0.336	
t_5-QMLE	3.032	1.865	-2.078	-2.134	1.062	0.714	0.707	0.463	0.880
RMLTS	-0.186	0.357	-0.613	-1.373	0.517	0.483	0.278	0.237	1.711
vdW	-1.250	0.170	1.100	-2.014	0.432	0.526	0.151	0.204	1.973
Spearman	-1.022	0.119	1.018	-1.891	0.438	0.537	0.149	0.204	1.952
Sign	-1.515	-0.532	1.065	-3.410	0.966	1.095	0.333	0.501	0.895
(t_3)									
QMLE	-3.558	-0.210	2.092	-0.967	0.844	0.671	0.205	0.185	
t_5-QMLE	-2.185	-0.433	1.332	-0.613	0.386	0.349	0.098	0.095	2.052
RMLTS	-2.473	-0.510	1.313	-0.691	0.438	0.384	0.108	0.106	1.836
vdW	-2.680	-1.937	2.393	-1.053	0.602	0.557	0.143	0.135	1.325
Spearman	-2.880	-2.014	2.663	-1.033	0.640	0.589	0.150	0.142	1.253
Sign	-2.204	-3.916	1.996	0.104	0.784	0.681	0.201	0.179	1.032
(Non-spherical)									
QMLE	0.513	2.682	-0.572	-2.756	1.962	1.705	1.314	1.115	
t_5-QMLE	0.992	3.953	-0.154	-3.008	3.105	2.618	2.013	1.696	0.646
RMLTS	0.077	2.834	0.043	-2.473	2.118	1.886	1.400	1.181	0.926
vdW	-0.335	3.327	0.156	-4.017	2.597	2.273	1.386	1.212	0.816
Spearman	-0.373	3.361	0.487	-3.853	2.562	2.268	1.411	1.222	0.817
Sign	4.157	8.485	-4.713	-8.645	6.717	5.955	3.300	2.582	0.329

R-estimators lose only little efficiency with respect to the QMLE: as we have observed in the motivating example of Section 5.1.2, the good performance of the R-estimators under asymmetric distributions is not obtained at the expense of a loss of accuracy under the symmetric ones.

5.5.2 Small sample results

A major advantage of R-estimation over other semiparametric procedures (based on tangent space projections) is the fact that it does not require any kernel density estimation, which allows for applying our method also in relatively small samples. To gain understanding on that aspect, we consider the same setting as in Section 5.5.1, but with sample size $n = 300$ (an order of magnitude which is quite common in real-data applications: see e.g. Section 5.6) factorizing into $n_R n_S = 15 \times 20$. The results are shown in Section 6.4.3.2 of Chapter 6, where Table 6.1 provides the averaged bias, MSE and overall MSE ratios of all estimators under various innovation densities; all results are in line with those in Table 5.1. The corresponding boxplots are displayed (still in Section 6.4.3.2 of Chapter 6) in Figures 6.6-6.10 and confirm the superiority over the QMLE, also in small samples, of our R-estimators under non-elliptical innovations: even in small samples (with n_R and n_S as small as 15 and 20), our R-estimators outperform the QMLE under non-Gaussian innovations, while performing equally well under Gaussian conditions.

5.5.3 Resistance to outliers

We also investigated the robustness properties of our estimators and, more particularly, their resistance to additive outliers (AO). Following Maronna et al. (2019), we first generated Gaussian VAR(1) realizations $\{\boldsymbol{X}_t\}$ of (5.5.1) ($n = 300$); then, adding the AOs, obtained the contaminated observations $\{\boldsymbol{X}_t^* = \boldsymbol{X}_t + I(t = h)\boldsymbol{\xi}\}$, where h and $\boldsymbol{\xi}$ denote the location and size of the AOs, respectively. We set h in order to have 5% equally spaced AOs and put $\boldsymbol{\xi} = (4, 4)'$. The parameter $\boldsymbol{\theta}$ remains the same as in the previous settings. The contaminated observations are

demeaned prior to estimation procedures. Figure 5.2 provides the boxplots of our three R-estimators (sign, Spearman, vdW) along with the boxplots of the QMLE, t_5-QMLE, and RMLTSE. Comparing those boxplots and the figures shown at the bottom of Table 6.1 (Section 6.4.3.2) with the uncontaminated ones of Figure 5.1 reveals that AOs have a severe impact on the QMLE but a much less significant one on the R-estimators. For a_{12} and a_{22}, the bias and variance of the R-estimates are comparable to those of the RMLTSE, with the latter displaying a much larger bias for a_{11} and a_{21}. Overall, we remark that for the estimation of all parameters, vdW and Spearman R-estimators feature less variability than the t_5-QMLE and RMLTSE.

To gauge the trade-off between robustness and efficiency, we compare the MSE ratio of RMLTSE to the MSE ratios of the R-estimators under Gaussian innovation density, as displayed in Table 5.1 (see top panel)—see also Table 6.1 in Section 6.4. The vdW and Spearman R-estimators exhibit MSE ratios equal to 0.982 and 0.935, respectively, which corresponds to a smaller efficiency loss than for the RMLTSE (MSE ratio equal to 0.903)—suggesting that the trade-off between robustness and efficiency is more favorable for vdW and Spearman R-estimators than for the RMLTSE.

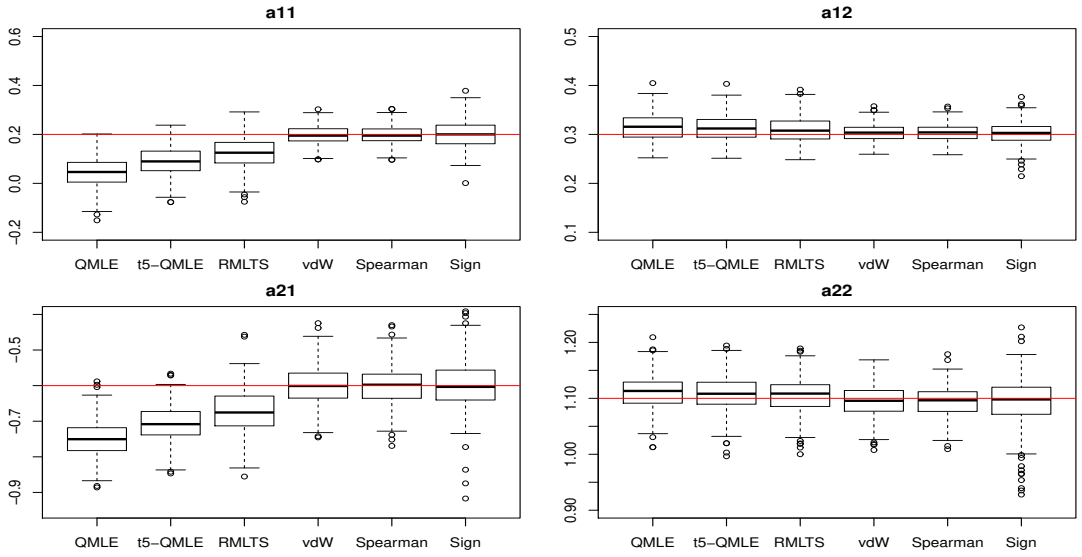
5.5.4 Further simulation results

We also considered a trivariate ($d = 3$) VAR(1), with parameter of interest $\boldsymbol{\theta} \in \mathbb{R}^9$, Gaussian and Gaussian mixture innovations, and sample size $n = 1000$. The results, which confirm the bivariate ones, can be found in Section 6.4.4 (Figures 6.11-6.12), along with details about the simulation design.

5.6 A real-data example

To conclude, we illustrate the applicability and good performance of our R-estimators in a real-data macroeconomic example. We consider the seasonally adjusted series

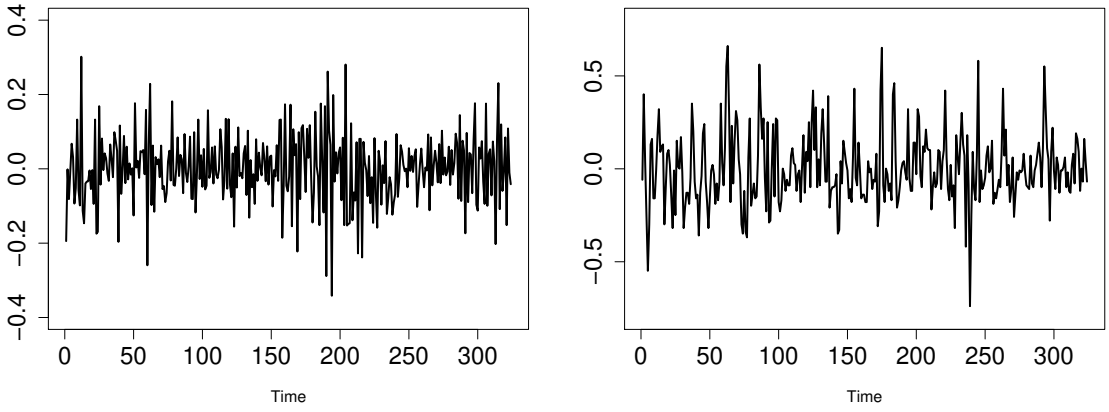
Figure 5.2: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under Gaussian innovations in the presence of additive outliers (sample size $n = 300$; $N = 300$ replications). The horizontal red line represents the actual parameter value.



of monthly housing starts (`Hstarts`) and the 30-year conventional mortgage rate (`Mortg`—no need for seasonal adjustment) in the US from January 1989 to January 2016, with sample size $n = 325$ each (both series are freely available on the Federal Reserve Bank of Saint Louis website, to which we refer for details). The same time series were studied by Tsay (2014, Section 3.15.2). Following Tsay, we analyze the differenced series; Figure 5.3 displays plots of their demeaned differences. While the `Mortg` series seems to be driven by skew innovations (with large positive values more likely than the negative ones), the `Hstarts` series looks more symmetric about zero. Visual inspection suggests the presence of significant auto- and cross-correlations, as expected from macroeconomic theory.

The AIC criterion selects a VARMA$(3, 1)$ model, the parameters of which we estimated using the benchmark QMLE (see e.g. Tsay (2014), Chapter 3) and our R-estimators (sign, Spearman, and van der Waerden). The QMLE-based multivariate Ljung-Box test does not reject the model at nominal level 1%. We report the estimates (along with their standard error, SE, in parentheses) in Table 5.2. Spotting the differences in Table 5.2 is all but simple, even though some look quite significant (see, for instance, the QMLE and R-estimates of $\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{22}$) and

Figure 5.3: Plots of demeaned differences of the monthly housing starts (measured in thousands of units; left panel) and the 30-year conventional mortgage rate (in percentage; right panel) in the US, from January 1989 through January 2016.



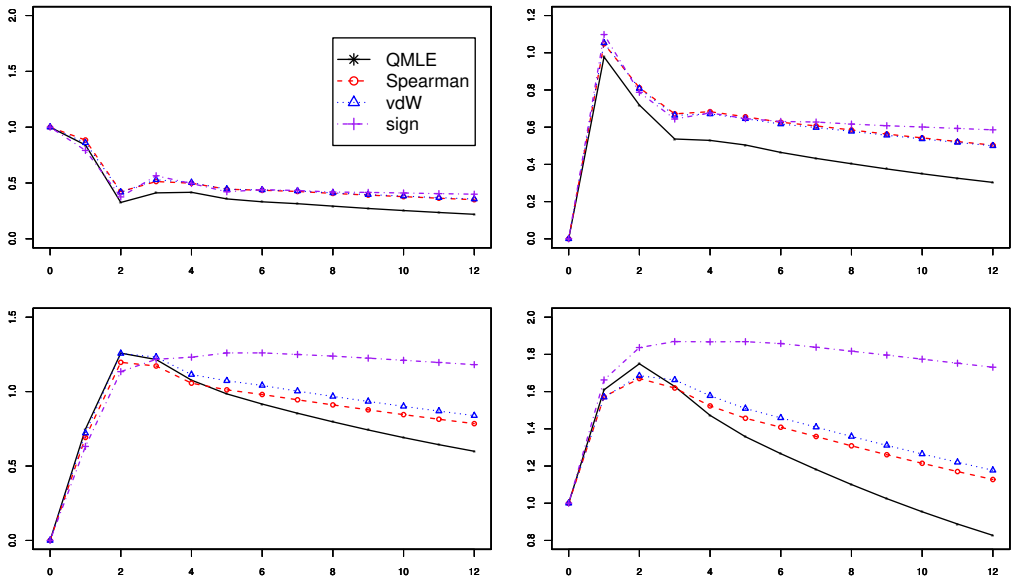
analyzing them is even more difficult.

Impulse response functions (IRFs) are easier to read and interpret; they are widely applied in macroeconometrics—see e.g. Tsay (2014) for a book-length description. Intuitively the IRFs express the effect of changes in one variable on another variable in multivariate time series analysis. In the VARMA case, the IRF is obtained using a MA representation: see Tsay (2014, Section 3.15.2) and Section 6.5 for mathematical details. In Figures 5.4, we plot the estimated IRFs resulting from the QMLE and R-estimators. The top plots show the response of `Hstarts` to its own shocks (left panel) and to the shocks of `Mortg`; the bottom panels show the response of `Mortg` to its own shocks (right panel) and to the shocks of `Hstarts`. Looking at the plots, we see that all IRFs have similar patterns. For instance, for all estimators, the top left panel illustrates that the IRF of the `Hstarts` to its own shocks have two consecutive increases after two initial drops. However, the decay of the QMLE-based IRF is uniformly faster than the R-estimator-based ones. Also, the other plots exhibit a more pronounced decay in the QMLE-based IRFs. Thus, R-estimators suggest a more persistent impact of the shocks: decision makers should be aware of this inferential aspect in the implementation of their economic policy.

Table 5.2: The QMLE and R-estimates of $\boldsymbol{\theta}$ in the VARMA$(3,1)$ fitting of the econometric data (demeaned differenced Hstarts and Mortg series); standard errors are shown in parentheses. The datasets are demeaned differenced Hstarts and Mortg series.

	$\boldsymbol{A}_1$		$\boldsymbol{A}_2$		$\boldsymbol{A}_3$		$\boldsymbol{B}_1$	
QMLE	0.137	0.487	-0.154	-0.199	0.032	0.056	-0.703	-0.490
	(0.265)	(0.353)	(0.284)	(0.130)	(0.171)	(0.072)	(0.258)	(0.350)
	0.596	0.974	0.030	-0.400	0.070	0.110	-0.152	-0.636
	(0.327)	(0.537)	(0.436)	(0.189)	(0.285)	(0.077)	(0.282)	(0.533)
vdW	0.155	0.526	-0.096	-0.181	0.017	0.038	-0.705	-0.527
	(0.141)	(0.088)	(0.122)	(0.079)	(0.133)	(0.062)	(0.088)	(0.071)
	0.561	0.943	0.094	-0.386	0.011	0.128	-0.161	-0.627
	(0.148)	(0.079)	(0.133)	(0.100)	(0.098)	(0.040)	(0.081)	(0.015)
Sign	0.087	0.536	-0.032	-0.198	0.075	-0.044	-0.705	-0.562
	(0.148)	(0.079)	(0.133)	(0.100)	(0.098)	(0.040)	(0.081)	(0.015)
	0.471	1.036	0.107	-0.403	0.035	0.148	-0.161	-0.627
	(0.178)	(0.084)	(0.165)	(0.073)	(0.138)	(0.061)	$(< 10^{-3})$	$(< 10^{-3})$
Spearman	0.180	0.511	-0.090	-0.180	0.030	0.049	-0.705	-0.537
	(0.066)	(0.033)	(0.092)	(0.046)	(0.113)	(0.049)	$(< 10^{-3})$	(0.014)
	0.531	0.946	0.072	-0.374	0.011	0.121	-0.161	-0.627
	(0.124)	(0.054)	(0.115)	(0.075)	(0.112)	(0.042)	$(< 10^{-3})$	$(< 10^{-3})$

Figure 5.4: Plots of estimated impulse response functions of the VARMA$(3,1)$ model for the differenced `Hstarts` (top panels) and `Mortg` (bottom panels) data, based on the QMLE and the R-estimators.



5.7 Conclusions and perspectives

We define a class of R-estimators based on the novel concept of center-outward ranks and signs, itself closely related to the theory of optimal measure transportation. Monte Carlo experiments show that these estimators significantly outperform the classical QMLE under skew multivariate innovations, even when the validity conditions for the latter are satisfied. In a companion paper, we study the performance of the corresponding rank-based tests for VAR models, and, more particularly, propose a center-outward Durbin-Watson test for multiple-output regression and a test of VAR(p_0) against VAR($p_0 + 1$) dependence. Our methodology is not limited to the VARMA case, though; its extension to nonlinear multivariate models, like the dynamic conditional correlation model of Engle (2002), is the subject of ongoing research.

Chapter 6

Supplementary material for Chapter 5

This chapter, as the supplementary material for Chapter 5, collects all proofs, computational aspects, and further numerical results of Chapter 5.

6.1 Technical material: algebraic preparation

Denote by $\boldsymbol{G}_u$ and $\boldsymbol{H}_u$, $u \in \mathbb{Z}$ the *Green's matrices* associated with the linear difference operators $\boldsymbol{A}(L)$ and $\boldsymbol{B}(L)$ in Section 5.2.1: those matrices are defined as the solutions of the homogeneous linear recursions

$$\boldsymbol{A}(L)\boldsymbol{G}_u = \boldsymbol{G}_u - \sum_{i=1}^{p} \boldsymbol{A}_i \boldsymbol{G}_{u-i} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{B}(L)\boldsymbol{H}_u = \sum_{i=0}^{q} \boldsymbol{B}_i \boldsymbol{H}_{u-i} = \boldsymbol{0}, \quad u \in \mathbb{Z}$$

with initial values $\boldsymbol{I}_d, \boldsymbol{0}, \ldots, \boldsymbol{0}$ at $u = 0, -1, \ldots, -p+1$ and $u = 0, -1, \ldots, -q+1$, respectively. Then, the residual process $\{\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}); 1 \leq t \leq n\}$ has the representa-

tion

$$\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}) = \sum_{i=0}^{t-1}\sum_{j=0}^{p} \boldsymbol{H}_i \boldsymbol{A}_j \boldsymbol{X}_{t-i-j}^{(n)} + \begin{bmatrix} \boldsymbol{H}_{t+q-1} & \cdots & \boldsymbol{H}_t \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_d & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{B}_1 & \boldsymbol{I}_d & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_{q-1} & \boldsymbol{B}_{q-2} & \cdots & \boldsymbol{I}_d \end{bmatrix} \begin{bmatrix} \boldsymbol{\epsilon}_{-q+1} \\ \vdots \\ \boldsymbol{\epsilon}_0 \end{bmatrix} \tag{6.1.1}$$

(see Hallin (1986), Garel and Hallin (1995), or Hallin and Paindaveine (2004)).

Assumption (A1) in Section 5.2.2 ensures the exponential decrease of $\{\|\boldsymbol{H}_u\|, u \in \boldsymbol{N}\}$ as $u \to \infty$. Specifically, there exists some $\varepsilon > 0$ such that $\|\boldsymbol{H}_u\|(1+\varepsilon)^u$ converges to 0 as $u \to \infty$. This also holds for the Green matrices $\boldsymbol{G}_u$ associated with the operator $\boldsymbol{A}(L)$. It follows that the initial values $\{\boldsymbol{\epsilon}_{-q+1}, \ldots, \boldsymbol{\epsilon}_0\}$ and $\{\boldsymbol{X}_{-p+1}, \ldots, \boldsymbol{X}_0\}$ in (6.1.1), which are typically unobservable, have no asymptotic influence on the residuals nor any asymptotic results. Therefore, they all can safely be set to zero in the sequel. This allows us to invert the AR and MA polynomials, and to define the Green matrices $\boldsymbol{G}_u$ and $\boldsymbol{H}_u$ as the matrix coefficients of the inverted operators $(\boldsymbol{A}(L))^{-1}$ and $(\boldsymbol{B}(L))^{-1}$:

$$\sum_{u=0}^{\infty} \boldsymbol{G}_u z^u := \left(\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i z^i\right)^{-1} \quad \text{and} \quad \sum_{u=0}^{\infty} \boldsymbol{H}_u z^u := \left(\sum_{i=0}^{q} \boldsymbol{B}_i z^i\right)^{-1}, \quad z \in \mathbb{C}, |z| < 1.$$

Associated with an arbitrary d-dimensional linear difference operator $\boldsymbol{C}(L) := \sum_{i=0}^{\infty} \boldsymbol{C}_i L^i$ (this of course includes operators of finite order s), define, for any

integers u and v, the $d^2u \times d^2v$ matrices

$$\boldsymbol{C}_{u,v}^{(l)} := \begin{bmatrix} \boldsymbol{C}_0 \otimes \boldsymbol{I}_d & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{C}_1 \otimes \boldsymbol{I}_d & \boldsymbol{C}_0 \otimes \boldsymbol{I}_d & \dots & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{C}_{v-1} \otimes \boldsymbol{I}_d & \boldsymbol{C}_{v-2} \otimes \boldsymbol{I}_d & \dots & \boldsymbol{C}_0 \otimes \boldsymbol{I}_d \\ \vdots & & & \vdots \\ \boldsymbol{C}_{u-1} \otimes \boldsymbol{I}_d & \boldsymbol{C}_{u-2} \otimes \boldsymbol{I}_d & \dots & \boldsymbol{C}_{u-v} \otimes \boldsymbol{I}_d \end{bmatrix}$$

and

$$\boldsymbol{C}_{u,v}^{(r)} := \begin{bmatrix} \boldsymbol{I}_d \otimes \boldsymbol{C}_0 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{I}_d \otimes \boldsymbol{C}_1 & \boldsymbol{I}_d \otimes \boldsymbol{C}_0 & \dots & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{I}_d \otimes \boldsymbol{C}_{v-1} & \boldsymbol{I}_d \otimes \boldsymbol{C}_{v-2} & \dots & \boldsymbol{I}_d \otimes \boldsymbol{C}_0 \\ \vdots & & & \vdots \\ \boldsymbol{I}_d \otimes \boldsymbol{C}_{u-1} & \boldsymbol{I}_d \otimes \boldsymbol{C}_{u-2} & \dots & \boldsymbol{I}_d \otimes \boldsymbol{C}_{u-v} \end{bmatrix}.$$

Write $\boldsymbol{C}_u^{(l)}$ for $\boldsymbol{C}_{u,u}^{(l)}$ and $\boldsymbol{C}_u^{(r)}$ for $\boldsymbol{C}_{u,u}^{(r)}$. With this notation, note that $\boldsymbol{G}_u^{(l)}, \boldsymbol{G}_u^{(r)}, \boldsymbol{H}_u^{(l)}$, and $\boldsymbol{H}_u^{(r)}$ are the inverses of $\boldsymbol{A}_u^{(l)}, \boldsymbol{A}_u^{(r)}, \boldsymbol{B}_u^{(l)}$ and $\boldsymbol{B}_u^{(r)}$, respectively. Denoting by $\boldsymbol{C}_{u,v}^{\prime(l)}$ and $\boldsymbol{C}_{u,v}^{\prime(r)}$ the matrices associated with the transposed operator $\boldsymbol{C}'(L) := \sum_{i=0}^{\infty} \boldsymbol{C}_i' L^i$, we have $\boldsymbol{G}_u'^{(l)} = (\boldsymbol{A}_u'^{(l)})^{-1}$, $\boldsymbol{H}_u'^{(l)} = (\boldsymbol{B}_u'^{(l)})^{-1}$, and so on. Define the $d^2(p+q) \times d^2(p+q)$ matrix

$$\boldsymbol{M}_{\boldsymbol{\theta}} := (\boldsymbol{G}_{p+q,p}'^{(l)} \vdots \boldsymbol{H}_{p+q,q}'^{(l)}) : \tag{6.1.2}$$

under Assumption (A1) in Section 5.2.2, $\boldsymbol{M}_{\boldsymbol{\theta}}$ is of full rank.

Also, consider the operator $\boldsymbol{D}(L) := \boldsymbol{I}_d + \sum_{i=1}^{p+q} \boldsymbol{D}_i L^i$ (note that $\boldsymbol{D}(L)$ and most quantities defined below depends on $\boldsymbol{\theta}$; for simplicity, however, we are dropping

this reference to $\boldsymbol{\theta}$), where

$$\begin{bmatrix} \boldsymbol{D}'_1 \\ \vdots \\ \boldsymbol{D}'_{p+q} \end{bmatrix} := - \begin{bmatrix} \boldsymbol{G}_q & \boldsymbol{G}_{q-1} & \dots & \boldsymbol{G}_{-p+1} \\ \boldsymbol{G}_{q+1} & \boldsymbol{G}_q & \dots & \boldsymbol{G}_{-p+2} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{G}_{p+q-1} & \boldsymbol{G}_{p+q-2} & \dots & \boldsymbol{G}_0 \\ \boldsymbol{H}_p & \boldsymbol{H}_{p-1} & \dots & \boldsymbol{H}_{-q+1} \\ \boldsymbol{H}_{p+1} & \boldsymbol{H}_p & \dots & \boldsymbol{H}_{-q+2} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{H}_{p+q-1} & \boldsymbol{H}_{p+q-2} & \dots & \boldsymbol{H}_0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{G}_{q+1} \\ \vdots \\ \boldsymbol{G}_{p+q} \\ \boldsymbol{H}_{p+1} \\ \vdots \\ \boldsymbol{H}_{p+q} \end{bmatrix}$$

(recall that $\boldsymbol{G}_{-1} = \boldsymbol{G}_{-2} = \cdots = \boldsymbol{G}_{-p+1} = \boldsymbol{0}$ and $\boldsymbol{H}_{-1} = \boldsymbol{H}_{-2} = \cdots = \boldsymbol{H}_{-q+1} = \boldsymbol{0}$). Let $\{\boldsymbol{\psi}_t^{(1)}, \dots, \boldsymbol{\psi}_t^{(p+q)}\}$ be a set of $d \times d$ matrices forming a fundamental system of solutions of the homogeneous linear difference equation associated with $\boldsymbol{D}(L)$. Such a system can be obtained from the Green matrices of the operator $\boldsymbol{D}(L)$ (see, e.g., Hallin 1986). Defining

$$\bar{\boldsymbol{\psi}}_m(\boldsymbol{\theta}) := \begin{bmatrix} \boldsymbol{\psi}_1^{(1)} & \dots & \boldsymbol{\psi}_1^{(p+q)} \\ \boldsymbol{\psi}_2^{(1)} & \dots & \boldsymbol{\psi}_2^{(p+q)} \\ \vdots & & \vdots \\ \boldsymbol{\psi}_m^{(1)} & \dots & \boldsymbol{\psi}_m^{(p+q)} \end{bmatrix} \otimes \boldsymbol{I}_d,$$

the *Casorati matrix* $\mathbf{C}_{\boldsymbol{\psi}}$ associated with $\boldsymbol{D}(L)$ is $\bar{\boldsymbol{\psi}}_{p+q}$. Finally, let

$$\boldsymbol{P}_{\boldsymbol{\theta}} := \mathbf{C}_{\boldsymbol{\psi}}^{-1} \quad \text{and} \quad \boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)} := \boldsymbol{H}_{n-1}^{(r)} \boldsymbol{B}_{n-1}^{\prime(l)} \bar{\boldsymbol{\psi}}_{n-1}. \tag{6.1.3}$$

6.2 Proofs

This section gathers the proofs of all mathematical results of Chapter 5. Throughout, we consider $f \in \mathcal{F}_d$ (the family of densities introduced in Section 5.2.2) and

assume that, for all $c \in \mathbb{R}^+$, there exist $b_{c;f}$ and $a_{c;f}$ in $\mathbb{R}$ such that $0 < b_{c;f} \leq a_{c;f} < \infty$ and $b_{c;f} \leq f(\boldsymbol{x}) \leq a_{c;f}$ for $\|\boldsymbol{x}\| \leq c$.

Proof of Proposition 5.2.1.

The LAN result is essentially the same as in Garel and Hallin (1995, (LAN 2) in their Proposition 3.1) and, moving along the same lines as in the proof of Proposition 1 in Hallin and Paindaveine (2004), we obtain the form (5.2.3) of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$. The form of the asymptotic covariance matrix $\boldsymbol{\Lambda}_f(\boldsymbol{\theta})$ and its finiteness easily follow from applying Lemma 4.12 in Garel and Hallin (1995). Details are left to the reader. □

To prove Propositions 5.4.1 and 5.4.2, we first need to establish the asymptotic normality, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ and $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$, of the rank-based $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$. As in the univariate case, due to the fact that the ranks are not mutually independent, the asymptotic normality of a rank statistic does not follow from classical central-limit theorems. The approach we are adopting here is inspired from Hájek, and consists in establishing an asymptotic representation result for the rank-based statistic under study—namely, its asymptotic equivalence with a sum of independent variable which are no longer rank-based—then proving the asymptotic normality of the latter. This is achieved here in a series of lemmas: Lemma 6.2.1 deals with the asymptotic normality of $(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))$, a corollary of which is the asymptotic normality of the truncated versions $\bar{\boldsymbol{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta})$ of $\bar{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$; Lemma 6.2.3 provides the asymptotic representation of $\mathrm{vec}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))$ by $\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))$; the asymptotic representation of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$ by $\bar{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$ and their asymptotic normality are obtained in Lemma 6.2.4. The proofs of Propositions 5.4.1 and 5.4.2 follow.

Let us start with the asymptotic normality of $(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))$.

Lemma 6.2.1. *Let Assumptions (A1), (A2), and (A3) in Chapter 5 hold. Then, for any positive integer i, the vector $(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))$ in (5.4.7) is asymptotically normal with mean $\mathbf{0}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, mean $\boldsymbol{K}_{J_1,J_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$, and covariance $d^{-2}\sigma^2_{J_1}\sigma^2_{J_2}\boldsymbol{I}_{d^2}$ under both.*

Proof. Since $L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\theta};f} = \boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau} + o_{\mathrm{P}}(1)$, the joint asymptotic normality of $(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))$ and $L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\theta};f}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ follows, via the classical Wold-Cramér argument, from the asymptotic normality of

$$N^{(n)}_{\boldsymbol{\alpha},\beta} := (n-i)^{1/2}\boldsymbol{\alpha}'\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})) + \beta\boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$$

for arbitrary $\boldsymbol{\alpha} \in \mathbb{R}^{d^2}$ and $\beta \in \mathbb{R}$. Since $\boldsymbol{Z}^{(n)}_1, \ldots, \boldsymbol{Z}^{(n)}_n$ are i.i.d. and $\boldsymbol{F}_{\pm,t} := \boldsymbol{F}_{\pm}(\boldsymbol{Z}^{(n)}_t)$ is uniform over the unit ball, $N^{(n)}_{\boldsymbol{\alpha},\beta}$ is a sum of martingale differences. If it is uniformly square-integrable, with finite variance $C^{(n)}_{\boldsymbol{\alpha},\beta}$, say, such that $\lim_{n\to\infty} C^{(n)}_{\boldsymbol{\alpha},\beta} =: C_{\boldsymbol{\alpha},\beta}$ exists and is finite, the martingale central limit theorem applies, and $N^{(n)}_{\boldsymbol{\alpha},\beta}$ is asymptotically normal with mean 0 and variance $C_{\boldsymbol{\alpha},\beta}$. Now, the variance of $N^{(n)}_{\boldsymbol{\alpha},\beta}$ takes the form

$$\begin{aligned} C^{(n)}_{\boldsymbol{\alpha},\beta} = {} & (n-i)\boldsymbol{\alpha}'\mathrm{Var}\big(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))\big)\boldsymbol{\alpha} \\ & + 2\beta\boldsymbol{\alpha}'(n-i)^{1/2}\mathrm{Cov}\big(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})), \boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})\big) \\ & + \beta^2\boldsymbol{\tau}'\mathrm{Var}\big(\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})\big)\boldsymbol{\tau}. \end{aligned}$$

The entries of each $\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})$ are uniformly square-integrable. As for $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$, it follows from Lemma 2.2 in Hallin and Werker (2003) that, for any LAN family, a uniformly pth-order integrable version of the central sequence exists: without loss of generality, let us assume that $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$, for $p = 2$, is one of them. The sequence $N^{(n)}_{\boldsymbol{\alpha},\beta}$ thus has a limiting $\mathcal{N}(0, C_{\boldsymbol{\alpha},\beta})$ distribution provided that $\lim_{n\to\infty} C^{(n)}_{\boldsymbol{\alpha},\beta} =: C_{\boldsymbol{\alpha},\beta}$ exists and is finite.

Due to the independence between the signs $\boldsymbol{S}_{\pm,t} := \boldsymbol{F}_{\pm,t}/\|\boldsymbol{F}_{\pm,t}\|$ and the moduli $\|\boldsymbol{F}_{\pm,t}\|$ (which follows from the fact that $\boldsymbol{F}_{\pm,t} \sim \mathrm{U}_d$), and due to the fact that

$\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$ are i.i.d.,

$$
\begin{aligned}
&\lim_{n\to\infty}(n-i)\mathrm{Var}\big(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))\big)\\
&=\lim_{n\to\infty}\mathrm{E}\left\{(n-i)\mathrm{vec}\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})(\mathrm{vec}\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))'\right\}\\
&=\lim_{n\to\infty}(n-i)^{-1}\mathrm{E}\left\{\left[\sum_{t=i+1}^{n}J_1(\|\boldsymbol{F}_{\pm,t}\|)J_2(\|\boldsymbol{F}_{\pm,t-i}\|)\mathrm{vec}(\boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i})\right]\right.\\
&\qquad\qquad\left.\times\left[\sum_{t=i+1}^{n}J_1(\|\boldsymbol{F}_{\pm,t}\|)J_2(\|\boldsymbol{F}_{\pm,t-i}\|)\mathrm{vec}(\boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i})\right]'\right\}\\
&=\frac{1}{d^2}\sigma^2_{J_1}\sigma^2_{J_2}\boldsymbol{I}_{d^2},
\end{aligned}
\tag{6.2.1}
$$

where the last equation follows from the uniform distribution of $\boldsymbol{S}_{\pm,t}$ over $\mathcal{S}_{d-1}$. Next, the uniform square-integrability of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ and its asymptotic normality in Proposition 5.2.1 yield

$$
\begin{aligned}
&\lim_{n\to\infty}(n-i)^{1/2}\mathrm{Cov}\big(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})),\boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})\big)\\
&=\lim_{n\to\infty}\mathrm{E}\left[(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))\boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})\right]\\
&=\lim_{n\to\infty}\mathrm{E}\left[(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))\boldsymbol{\Gamma}_f^{(n)\prime}(\boldsymbol{\theta})\right]\boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau},
\end{aligned}
\tag{6.2.2}
$$

where the last equality follows from (5.2.3). Due to the independence of $\boldsymbol{Z}_i^{(n)}$ and $\boldsymbol{Z}_j^{(n)}$ for $i\neq j$, only $\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta})$ in $\boldsymbol{\Gamma}_f^{(n)}(\boldsymbol{\theta})$ is contributing to (6.2.2). Therefore, using the block matrix form of $\boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)}=\big(\boldsymbol{Q}'_{1,\boldsymbol{\theta}}\ldots\boldsymbol{Q}'_{n-1,\boldsymbol{\theta}}\big)'$, the expression in (6.2.2) reduces to

$$
\lim_{n\to\infty}(n-i)\mathrm{E}\left[\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))(\mathrm{vec}(\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta})))'\right]\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}. \tag{6.2.3}
$$

From (5.2.5), we have

$$
\begin{aligned}
&(n-i)\mathrm{E}\left[\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))(\mathrm{vec}(\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta})))'\right] \\
&= (n-i)^{-1}\mathrm{E}\left\{\left[\sum_{t=i+1}^{n} J_1(\|\boldsymbol{F}_{\pm,t}\|)J_2(\|\boldsymbol{F}_{\pm,t-i}\|)\mathrm{vec}(\boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i})\right]\left[\sum_{t=i+1}^{n}\mathrm{vec}(\boldsymbol{\varphi}_f(\boldsymbol{Z}^{(n)}_t))\boldsymbol{Z}'_{t-i})\right]'\right\} \\
&= \mathrm{E}\big[J_1(\|\boldsymbol{F}_{\pm,t}\|)J_2(\|\boldsymbol{F}_{\pm,t-i}\|)(\boldsymbol{I}_d\otimes\boldsymbol{S}_{\pm,t})\boldsymbol{S}_{\pm,t-i}\boldsymbol{Z}'_{t-i}(\boldsymbol{I}_d\otimes\boldsymbol{\varphi}'_f(\boldsymbol{Z}^{(n)}_t))\big] \qquad (6.2.4)
\end{aligned}
$$

where the last two equalities follow from the independence of $\boldsymbol{Z}^{(n)}_1,\ldots,\boldsymbol{Z}^{(n)}_n$ and the uniform distribution of $\boldsymbol{F}_{\pm,t}\sim \mathrm{U}_d$. In view of (5.2.6), (6.2.2), (6.2.3) and (6.2.4), we thus obtain

$$
\lim_{n\to\infty}(n-i)^{1/2}\mathrm{Cov}\big(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})),\boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})\big)=\boldsymbol{K}_{J_1,J_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}. \quad (6.2.5)
$$

Combining (6.2.1), (6.2.5) and the asymptotic normality of $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$ in Proposition 5.2.1 yields, for arbitrary $\boldsymbol{\alpha}$ and β,

$$
\lim_{n\to\infty}C^{(n)}_{\boldsymbol{\alpha},\beta}=\boldsymbol{\alpha}'\boldsymbol{\alpha}d^{-2}\sigma^2_{J_1}\sigma^2_{J_2}+2\beta\boldsymbol{\alpha}'\boldsymbol{K}_{J_1,J_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}+\beta^2\boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau}. \quad (6.2.6)
$$

It follows that $\big((n-i)^{1/2}\mathrm{vec}'(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})),\ L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\theta};f}\big)'$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, is asymptotically jointly normal, with mean $\big(\mathbf{0}',-\frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau}\big)'$ and covariance

$$
\begin{bmatrix} d^{-2}\sigma^2_{J_1}\sigma^2_{J_2}\boldsymbol{I}_{d^2} & \boldsymbol{K}_{J_1,J_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau} \\ (\boldsymbol{K}_{J_1,J_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau})' & \boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau} \end{bmatrix}. \qquad (6.2.7)
$$

The desired result then readily follows from applying Le Cam's third Lemma. □

Recall that $\boldsymbol{T}^{(n)}_{\boldsymbol{\theta}}=\boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}}\boldsymbol{Q}^{(n)\prime}_{\boldsymbol{\theta}}$. For any positive integer $m\le n-1$, let

$$
\bar{\boldsymbol{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta}):=\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}\bar{\boldsymbol{\Gamma}}^{(m,n)}_{J_1,J_2}(\boldsymbol{\theta}), \qquad (6.2.8)
$$

where

$$\bar{\mathbf{\Gamma}}^{(m,n)}_{J_1,J_2}(\boldsymbol{\theta}) := \big((n-1)^{1/2}(\mathrm{vec}\bar{\mathbf{\Gamma}}^{(n)}_{1,J_1,J_2}(\boldsymbol{\theta}))', \ldots, (n-m)^{1/2}(\mathrm{vec}\bar{\mathbf{\Gamma}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta})\big)')' :$$

clearly, $\bar{\mathbf{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta})$, it is the truncated version of $\bar{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$ defined in Section 5.4.1. The asymptotic normality of $\bar{\mathbf{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta})$ follows from Lemma 6.2.1 as a corollary.

Corollary 6.2.1.1. *Let Assumptions (A1), (A2), and (A3) in Chapter 5 hold. Then, for any positive integer m, the vector $\bar{\mathbf{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta})$ in (6.2.8) is asymptotically normal, with mean $\mathbf{0}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, mean*

$$\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}(\boldsymbol{I}_m \otimes \boldsymbol{K}_{J_1,J_2,f})\boldsymbol{T}^{(m+1)'}_{\boldsymbol{\theta}}\boldsymbol{\tau} \tag{6.2.9}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$, and covariance $d^{-2}\sigma^2_{J_1}\sigma^2_{J_2}\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}\boldsymbol{T}^{(m+1)'}_{\boldsymbol{\theta}}$ under both.

The following auxiliary lemma, which follows along the same lines as Lemma 4 in Hallin and Paindaveine (2002b) and Lemma 5 in Hallin and Paindaveine (2004), will be useful in subsequent proofs.

Lemma 6.2.2. *Let $i \in \{1, \ldots, n-1\}$ and $t, t' \in \{i+1, \ldots, n\}$ be such that $t \neq t'$. Assume that $g : \mathbb{R}^{nd} = \mathbb{R}^d \times \cdots \times \mathbb{R}^d \to \mathbb{R}$ is even in all its arguments, and such that the expectation in* (6.2.10) *below exists. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$,*

$$\mathrm{E}\big[g(\boldsymbol{Z}^{(n)}_1, \ldots, \boldsymbol{Z}^{(n)}_n)(\boldsymbol{P}'_t\boldsymbol{Q}_t)(\boldsymbol{R}'_{t-i}\boldsymbol{S}_{t'-i})\big] = 0, \tag{6.2.10}$$

where $\boldsymbol{P}_t, \boldsymbol{Q}_t, \boldsymbol{R}_t$ and $\boldsymbol{S}_t$ are any four random vectors among $\boldsymbol{S}^{(n)}_{\pm,t}$ and $\boldsymbol{S}^{(n)}_{\pm,t} - \boldsymbol{S}_{\pm,t}$.

The next lemma establishes an asymptotic representation result for the rank-based cross-covariance matrices $\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})$ defined in (5.3.12) by showing their asymptotic equivalence with $\bar{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})$ defined in (5.4.7). LAN implies that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$ and $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ are mutually contiguous; (6.2.11) therefore holds under both. This asymptotic representation in the Hájek style of a center-outward serial rank statistic extends to a multivariate setting a univariate result first established by Hallin et al. (1985).

Lemma 6.2.3. *Let Assumptions (A1), (A2), and (A3) in Chapter 5 hold. Then, for any positive integer i,*

$$\operatorname{vec}\left(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}) - \bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})\right) = o_{\mathrm{P}}(n^{-1/2}) \tag{6.2.11}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ *and* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$, *as* $n \to \infty$.

Proof. Note that $(n-i)^{1/2}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}) - \bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})) = \boldsymbol{\delta}^{(n)}_1 + \boldsymbol{\delta}^{(n)}_2$ where

$$\boldsymbol{\delta}^{(n)}_1 := (n-i)^{-1/2} \sum_{t=i+1}^{n} \left(J_1(\frac{R^{(n)}_{\pm,t}}{n_R+1}) J_2(\frac{R^{(n)}_{\pm,t-i}}{n_R+1}) - J_1(\|\boldsymbol{F}_{\pm,t}\|) J_2(\|\boldsymbol{F}_{\pm,t-i}\|) \right) \boldsymbol{S}^{(n)}_{\pm,t} \boldsymbol{S}^{(n)\prime}_{\pm,t-i}$$

and

$$\boldsymbol{\delta}^{(n)}_2 := (n-i)^{-1/2} \sum_{t=i+1}^{n} J_1(\|\boldsymbol{F}_{\pm,t}\|) J_2(\|\boldsymbol{F}_{\pm,t-i}\|) \left(\boldsymbol{S}^{(n)}_{\pm,t} \boldsymbol{S}^{(n)\prime}_{\pm,t-i} - \boldsymbol{S}_{\pm,t} \boldsymbol{S}'_{\pm,t-i} \right).$$

It suffices to show that $\operatorname{vec}(\boldsymbol{\delta}^{(n)}_1)$ and $\operatorname{vec}(\boldsymbol{\delta}^{(n)}_2)$ both converge in quadratic mean to zero as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$.

Let $\|\cdot\|_{L^2}$ denote the l_2-norm. For $\boldsymbol{\delta}^{(n)}_1$, we make use of Lemma 6.2.2, and we exploit the independence of the ranks $\{R^{(n)}_{\pm,t}; t = 1, \ldots, n\}$ and the signs $\{\boldsymbol{S}^{(n)}_{\pm,t}; t = 1, \ldots, n\}$ (see Hallin (2017)). Given that $(\operatorname{vec}\boldsymbol{A})'(\operatorname{vec}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}'\boldsymbol{B})$, we have

$$\left\|\operatorname{vec}(\boldsymbol{\delta}^{(n)}_1)\right\|^2_{L^2} = (n-i)^{-1} \sum_{t=i+1}^{n} \mathrm{E}\left[\left(J_1(\frac{R^{(n)}_{\pm,t}}{n_R+1}) J_2(\frac{R^{(n)}_{\pm,t-i}}{n_R+1}) - J_1(\|\boldsymbol{F}_{\pm,t}\|) J_2(\|\boldsymbol{F}_{\pm,t-i}\|) \right)^2 \right].$$

The Glivenko-Cantelli result in Hallin (2017, Proposition 5.1) entails

$$\max_{1\le t\le n} \left| R^{(n)}_{\pm,t}/(n_R+1) - \|\boldsymbol{F}_{\pm,t}\| \right| \to 0 \quad a.s. \quad \text{as} \quad n \to \infty. \tag{6.2.12}$$

In view of the assumptions made on J_1 and J_2, Lemma 6.1.6.1 of Hájek et al. (1999) yields

$$\|\operatorname{vec}(\boldsymbol{\delta}^{(n)}_1)\|^2_{L^2} \to 0 \quad \text{as } n \to \infty. \tag{6.2.13}$$

For $\boldsymbol{\delta}_2^{(n)}$, we have

$$\boldsymbol{\delta}_2^{(n)} = (n-i)^{-1/2} \sum_{t=i+1}^{n} J_1(\|\boldsymbol{F}_{\pm,t}\|) J_2(\|\boldsymbol{F}_{\pm,t-i}\|) \left[\left(\boldsymbol{S}_{\pm,t}^{(n)} - \boldsymbol{S}_{\pm,t} \right) \boldsymbol{S}_{\pm,t-i}^{(n)\prime} + \boldsymbol{S}_{\pm,t} \left(\boldsymbol{S}_{\pm,t-i}^{(n)\prime} - \boldsymbol{S}_{\pm,t-i}^{\prime} \right) \right].$$

Similar to the arguments used for $\boldsymbol{\delta}_1^{(n)}$, Lemma 6.2.2 and $(\text{vec}\boldsymbol{A})'(\text{vec}\boldsymbol{B}) = \text{tr}(\boldsymbol{A}'\boldsymbol{B})$ imply

$$\|\text{vec}(\boldsymbol{\delta}_2^{(n)})\|_{L^2}^2 \leq 2(n-i)^{-1} \sum_{t=i+1}^{n} \text{E}\left[\left(J_1(\|\boldsymbol{F}_{\pm,t}\|) J_2(\|\boldsymbol{F}_{\pm,t-i}\|) \right)^2 \|\boldsymbol{S}_{\pm,t}^{(n)} - \boldsymbol{S}_{\pm,t}\|^2 \right] \tag{6.2.14}$$

$$+ 2(n-i)^{-1} \sum_{t=i+1}^{n} \text{E}\left[\left(J_1(\|\boldsymbol{F}_{\pm,t}\|) J_2(\|\boldsymbol{F}_{\pm,t-i}\|) \right)^2 \|\boldsymbol{S}_{\pm,t-i}^{(n)} - \boldsymbol{S}_{\pm,t-i}\|^2 \right]. \tag{6.2.15}$$

Still in view of Proposition 5.1 in Hallin (2017), $\max_{1\leq t\leq n} \|\boldsymbol{S}_{\pm,t}^{(n)} - \boldsymbol{S}_{\pm,t}\| \to 0$ a.s. as $n \to \infty$. Since J_1 and J_2 are square-integrable and $\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$ are independent, both (6.2.14) and (6.2.15) converge to 0. The result follows. □

We now can extend the above asymptotic representation and asymptotic normality results from the rank-based cross-covariance matrices $\underset{\sim}{\boldsymbol{\Gamma}}_{i,J_1,J_2}^{(n)}(\boldsymbol{\theta})$ to the rank-based central sequence $\underset{\sim}{\boldsymbol{\Delta}}_{J_1,J_2}^{(n)}(\boldsymbol{\theta})$.

Lemma 6.2.4. *Let Assumptions (A1), (A2), and (A3) in Chapter 5 hold. Then,*

$$\underset{\sim}{\boldsymbol{\Delta}}_{J_1,J_2}^{(n)}(\boldsymbol{\theta}) - \bar{\underset{\sim}{\boldsymbol{\Delta}}}_{J_1,J_2}^{(n)}(\boldsymbol{\theta}) = o_{\text{P}}(1) \quad \textit{as } n \to \infty \tag{6.2.16}$$

both under $\text{P}_{\boldsymbol{\theta};f}^{(n)}$ *and* $\text{P}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}^{(n)}$. *Moreover,* $\underset{\sim}{\boldsymbol{\Delta}}_{J_1,J_2}^{(n)}(\boldsymbol{\theta})$ *is asymptotically normal, with mean* $\mathbf{0}$ *under* $\text{P}_{\boldsymbol{\theta};f}^{(n)}$, *mean*

$$\lim_{n\to\infty} \left\{ \boldsymbol{T}_{\boldsymbol{\theta}}^{(n)} (\boldsymbol{I}_{n-1} \otimes \boldsymbol{K}_{J_1,J_2,f}) \boldsymbol{T}_{\boldsymbol{\theta}}^{(n)\prime} \right\} \boldsymbol{\tau} \tag{6.2.17}$$

under $\text{P}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}^{(n)}$, *and covariance* $d^{-2}\sigma_{J_1}^2\sigma_{J_2}^2 \lim_{n\to\infty} \left\{ \boldsymbol{T}_{\boldsymbol{\theta}}^{(n)} \boldsymbol{T}_{\boldsymbol{\theta}}^{(n)\prime} \right\}$ *under both.*

Note that the limits appearing in the above asymptotic means and covariances

exist due to Assumption (A1) on the characteristic roots of the VARMA operators involved.

Proof. For (6.2.16), due to Lemma 6.2.3 and contiguity, it is sufficient to prove that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, for $m = m(n) \leq n-1$ and provided that $m(n) \to \infty$ as $n \to \infty$,

$$\limsup_{n\to\infty} \|\bar{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta}) - \bar{\boldsymbol{\Delta}}^{(n)}_{m(n),J_1,J_2}(\boldsymbol{\theta})\| = o_{\mathrm{P}}(1) \tag{6.2.18}$$

and

$$\limsup_{n\to\infty} \|\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{m(n),J_1,J_2}(\boldsymbol{\theta})\| = o_{\mathrm{P}}(1). \tag{6.2.19}$$

For $m = n-1$, the left-hand sides in (6.2.18) and (6.2.19) are exactly zero. Therefore, we only need to consider $m \leq n-2$. It follows from Proposition 3.1 (LAN2) in Garel and Hallin (1995) that

$$\bar{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta}) - \bar{\boldsymbol{\Delta}}^{(n)}_{m(n),J_1,J_2}(\boldsymbol{\theta})$$
$$= \begin{bmatrix} \sum_{i=m+1}^{n-1}\sum_{j=0}^{i-1}\sum_{k=0}^{\min(q,i-j-1)}[(\boldsymbol{G}_{i-j-k-1}\boldsymbol{B}_k)\otimes\boldsymbol{H}'_j](n-i)^{1/2}(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))) \\ \vdots \\ \sum_{i=m+1}^{n-1}\sum_{j=0}^{i-p}\sum_{k=0}^{\min(q,i-j-p)}[(\boldsymbol{G}_{i-j-k-p}\boldsymbol{B}_k)\otimes\boldsymbol{H}'_j](n-i)^{1/2}(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))) \\ \sum_{i=m+1}^{n-1}(\boldsymbol{I}_d\otimes\boldsymbol{H}'_{i-1}(n-i)^{1/2}(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})) \\ \vdots \\ \sum_{i=m+1}^{n-1}(\boldsymbol{I}_d\otimes\boldsymbol{H}'_{i-q}(n-i)^{1/2}(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})) \end{bmatrix}$$

for any $p \leq m \leq n-2$,. Due to the square-integrability of J_1, J_2 and the fact that $\boldsymbol{Z}^{(n)}_1, \ldots, \boldsymbol{Z}^{(n)}_n$ are i.i.d., it follows from $(\mathrm{vec}\boldsymbol{A})'(\mathrm{vec}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}'\boldsymbol{B})$ that

$$\|(n-i)^{1/2}(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})))\|^2_{L^2} = (n-i)^{-1}\sum_{t=i+1}^{n}\mathrm{E}\left[J_1^2(\|\boldsymbol{F}_{\pm,t}\|)\right]\mathrm{E}\left[J_2^2(\|\boldsymbol{F}_{\pm,t-i}\|)\right] = \sigma^2_{J_1}\sigma^2_{J_2} < \infty.$$

Recall that, under Assumption (A1), the Green matrices $\boldsymbol{G}_u$ and $\boldsymbol{H}_u$ decrease exponentially fast (see Section 6.1). Using the fact that $\|\mathbf{A}\mathbf{x}\|_{L^2} \leq \|\mathbf{A}\|\,\|\mathbf{x}\|_{L^2}$ (where $\|\mathbf{A}\|$ denotes the operator norm of $\mathbf{A}$) and the triangular inequality, we

thus obtain

$$\limsup_{n\to\infty} \|\bar{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta}) - \bar{\mathbf{\Delta}}^{(n)}_{m(n),J_1,J_2}(\boldsymbol{\theta})\|_{L^2} = 0.$$

The result (6.2.18) follows. Turning to (6.2.19), we have, in view of (6.2.13), (6.2.14) and (6.2.15),

$$\max_{1\le i\le n-1} \|(n-i)^{1/2}[\mathrm{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta})) - \mathrm{vec}(\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,J_1,J_2}(\boldsymbol{\theta}))]\|^2_{L^2} = o(1)$$

as $n \to \infty$. Hence, (6.2.19) follows along the same lines as (6.2.18). The asymptotic normality of $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$ then follows from (6.2.16) and the asymptotic normality of $\bar{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$, itself implied by (6.2.18) and Lemma 6.2.1.1. The asymptotic mean and variance are the limits as $m = m(n)$ and n tend to infinity, of the asymptotic mean and variance of $\bar{\mathbf{\Delta}}^{(n)}_{m(n),J_1,J_2}(\boldsymbol{\theta})$ and do not depend on the way m grows with n. □

Proof of Proposition 5.4.1.

Proposition 5.4.1 readily follows from (6.2.19) and the asymptotic linearity of the truncated $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{m,J_1,J_2}(\boldsymbol{\theta})$ implied by Assumption (A4). □

Proof of Proposition 5.4.2.

From the definition of $\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)}$ in (5.4.8), the asymptotic linearity in Proposition 5.4.1, the consistency of $\hat{\mathbf{\Upsilon}}^{(n)}_{J_1,J_2}$, the convergence of $\mathbf{\Upsilon}^{(n)}_{J_1,J_2,f}$ to $\mathbf{\Upsilon}_{J_1,J_2,f}$, and the asymptotic discreteness of $\hat{\boldsymbol{\theta}}^{(n)}$ (which allows us to treat $n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})$ as if it were a bounded constant: see Lemma 4.4 in Kreiss (1987)), we have

$$\begin{aligned}
&n^{1/2}(\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)} - \boldsymbol{\theta}) \\
&= n^{1/2}\left\{\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\left[\left(\hat{\mathbf{\Upsilon}}^{(n)}_{J_1,J_2}\right)^{-1} \underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\hat{\boldsymbol{\theta}}^{(n)})\right] - \boldsymbol{\theta}\right\} \\
&= n^{1/2}\left\{\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\left[\mathbf{\Upsilon}^{-1}_{J_1,J_2,f}\left(\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta}) - \mathbf{\Upsilon}^{(n)}_{J_1,J_2,f} n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})\right)\right] - \boldsymbol{\theta}\right\} + o_{\mathrm{P}}(1) \\
&= \mathbf{\Upsilon}^{-1}_{J_1,J_2,f}\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1).
\end{aligned}$$

This, in view of the asymptotic normality of $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{J_1,J_2}(\boldsymbol{\theta})$ in Lemma 6.2.4, completes

the proof of Proposition 5.4.2. □

6.3 Computational issues

6.3.1 Implementation details

In this section, we briefly discuss some computational aspects related to the implementation of our methodology.

(i) Consistency requires that both n_R and n_S tend to infinity. In practice, we factorize n into $n_R n_S + n_0$ in such a way that both n_R and n_S are large. Typically, n_R is of order $n^{1/d}$ and n_S is of order $n^{(d-1)/d}$, whilst $0 \leq n_0 < \min(n_S, n_R)$ has to be small as possible—its value, however, is entirely determined by the values of n_R and n_S. Generating "regular grids" of n_S points over the unit sphere $\mathcal{S}_{d-1}$ as described in Section 5.3 is easy for $d = 2$, where perfect regularity can be achieved by dividing the unit circle into n_S arcs of equal length $2\pi/n_S$. For $d \geq 3$, "perfect regularity" is no longer possible. A random array of n_S independent and uniformly distributed unit vectors does satisfy (almost surely) the requirement for weak convergence (to U_d). More regular deterministic arrays (with faster convergence) can be constructed, though, such as the *low-discrepancy sequences* (see, e.g., Niederreiter (1992), Judd (1998), Dick and Pillichshammer (2014), or Santner et al. (2003)) considered in numerical integration and the design of computer experiments; we suggest the use of the function `UnitSphere` in R package `mvmesh`.

(ii) The empirical center-outward distribution function $\boldsymbol{F}_{\pm}^{(n)}$ is obtained as the solution of an optimal coupling problem. Many efficient algorithms have been proposed in the measure transportation literature (see, e.g., Peyré and Cuturi (2019)). We followed Hallin et al. (2020a), using a Hungarian algorithm (see the `clue` R package).

(iii) The computation of the one-step R-estimator in (5.4.8) involves two basic ingredients: a preliminary root n-consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$ and an estimator of the cross-information matrix $\boldsymbol{\Upsilon}_{J_1,J_2,f}$. For the preliminary $\hat{\boldsymbol{\theta}}^{(n)}$, robust M-estimators

such as the reweighted multivariate least trimmed squares estimator (RMLTSE) proposed by Croux and Joossens (2008) for VAR models are obvious candidates; provided that fourth-order moments finite, the QMLE still constitutes a reasonable choice, though. Different preliminary estimators may lead to different one-step R-estimators. Differences, however, gradually wane on iterating (for fixed n) the one-step procedure and the asymptotic impact (as $n \to \infty$) of the choice of $\hat{\boldsymbol{\theta}}^{(n)}$ is nil. Turning to the estimation of $\boldsymbol{\Upsilon}_{J_1,J_2,f}$, the issue is that this matrix depends on the unknown actual density f. A simple consistent estimator is obtained by letting $\boldsymbol{\tau} = \mathbf{e}_i$, $i = 1, \ldots, (p+q)d^2$ in (5.4.5) where $\mathbf{e}_i$ denotes the ith vector of the canonical basis in the parameter space $\mathbb{R}^{(p+q)d^2}$: the difference $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\mathbf{e}_i) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\hat{\boldsymbol{\theta}}^{(n)})$ then provides a consistent estimator of the i-th column of $-\boldsymbol{\Upsilon}_{J_1,J_2,f}(\boldsymbol{\theta})$. See Hallin et al. (2006) or Cassart et al. (2010) for more sophisticated estimation methods.

6.3.2 Algorithm

We give here a detailed description of the estimation algorithm. Due to the exponential decay, under Assumption (A1) in Section 5.2.2, of the coefficients of the MA(∞) representation of VARMA(p, q) models, there is no need to bother about the truncation of the central sequence, which safely can be set to $m = n - 1$ or $m = \lfloor (1-\gamma)n \rfloor$ with $0 < \gamma < 1$. Then, the implementation of our R-estimation method then goes along the lines of the following algorithm.

Algorithm 1: Center-outward R-estimation for semiparametric VARMA models

Input: a d-dimensional sample $\{\boldsymbol{X}_t; 1 \leq t \leq n\}$, orders p and q of the VARMA process, number k of iterations in the one-step procedure; truncation lag m.

Output: R-estimator $\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)}$

1. Factorize n into $n_R n_S + n_0$ and generate (see *(i)* of Section 6.3.1), a "regular grid" of $n_R n_S$ points over the unit ball $\mathbb{S}_d$.
2. Compute a preliminary root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$.
3. Set the initial values $\boldsymbol{\epsilon}_{-q+1}, \ldots, \boldsymbol{\epsilon}_0$ and $\boldsymbol{X}_{-p+1}, \ldots, \boldsymbol{X}_0$ all equal to zero, and compute residuals $\boldsymbol{Z}_1^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}), \ldots, \boldsymbol{Z}_n^{(n)}(\hat{\boldsymbol{\theta}}^{(n)})$ recursively or from (6.1.1).
4. Create a $n \times n$ matrix $\mathbf{D}$ with (i, j) entry the squared Euclidean distance between $\mathbf{Z}_i^{(n)}$ and the j-th gridpoint. Based on that matrix, compute $\{\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}); t = 1, \ldots, n\}$ solving the optimal pairing problem in (5.3.1), using e.g. the Hungarian algorithm.
5. From $\boldsymbol{F}_{\pm}^{(n)}$, compute the center-outward ranks (5.3.2) and signs (5.3.3).
6. Specify the scores J_1 and J_2 (e.g., the standard scores proposed in Section 5.4.2) and compute $\boldsymbol{M}_{\hat{\boldsymbol{\theta}}^{(n)}}$, $\boldsymbol{P}_{\hat{\boldsymbol{\theta}}^{(n)}}$, and $\boldsymbol{Q}^{(n)}_{\hat{\boldsymbol{\theta}}^{(n)}}$, as defined in Section 6.1, then $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,J_1,J_2}(\hat{\boldsymbol{\theta}}^{(n)})$ (use e.g. one of the expressions available in Section 5.4.2). Finally, combine these expressions into $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\hat{\boldsymbol{\theta}}^{(n)})$.
7. For some chosen $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_{(p+q)d^2}$, compute $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\tau})$, then, via (5.4.5), $\hat{\boldsymbol{\Upsilon}}^{(n)}_{J_1,J_2}$.
8. Set $\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)} = \hat{\boldsymbol{\theta}}^{(n)}$.
9. **for** $i \leftarrow 1$ ***to*** k **do**

$$\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)} \leftarrow \hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)} + n^{-1/2} \left(\hat{\boldsymbol{\Upsilon}}^{(n)}_{J_1,J_2}\right)^{-1} \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{J_1,J_2}(\hat{\underset{\sim}{\boldsymbol{\theta}}}^{(n)}).$$

end

6.4 Supplementary material for Section 5.5

6.4.1 Center-outward quantile contours, with a graphical illustration

We provide here some additional concepts from Hallin (2017) and Hallin et al. (2020a). Recall that an *order statistic* $\boldsymbol{Z}^{(n)}_{(\cdot)}$ of the un-ordered n-tuple $\boldsymbol{Z}^{(n)}$ is an arbitrarily ordered version of the same—for instance, $\boldsymbol{Z}^{(n)}_{(\cdot)} = \left(\boldsymbol{Z}^{(n)}_{(1)}, ..., \boldsymbol{Z}^{(n)}_{(n)}\right)$, where $\boldsymbol{Z}^{(n)}_{(i)}$ is such that its first component is the ith order statistic of the n-tuple of first components.

The *center-outward quantile contours* are defined as

$$\mathcal{C}^{(n)}_{\pm;\boldsymbol{Z}^{(n)}_{(\cdot)}}\left(\frac{j}{n_R+1}\right) := \left\{\boldsymbol{Z}^{(n)}_t | R^{(n)}_{\pm,t} = j\right\}, \tag{6.4.1}$$

where $j/(n_R+1)$, $j = 0, 1, ..., n_R$ is an empirical probability content, to be interpreted as a quantile order. Figure 6.1 provides a graphical illustration of this concept: $n = 1000$ (with $n_R = 25$ and $n_S = 40$) bivariate observations were drawn from the Gaussian mixture (5.5.2), the skew-normal and skew-t_3 described in Section 6.4.2, and, for a comparison, from a spherical multivariate normal. The plots show that the center-outward quantile contours nicely conform to the shape of the underlying distribution in both symmetric and asymmetric cases.

6.4.2 Skew-normal, skew-t, and Gaussian mixture innovation densities

The skew-normal distribution considered in Section 5.5 has density (with $\phi(\cdot; \boldsymbol{\Sigma})$ standing for the $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ density, Φ for the univariate standard normal distribution function)

$$f_{\boldsymbol{\epsilon}}(\boldsymbol{z}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) := 2\phi(\boldsymbol{z} - \boldsymbol{\xi}; \boldsymbol{\Sigma})\Phi(\boldsymbol{\alpha}'\boldsymbol{w}^{-1}(\boldsymbol{z} - \boldsymbol{\xi})), \quad \boldsymbol{z} \in \mathbb{R}^d, \tag{6.4.2}$$

where $\boldsymbol{\xi} \in \mathbb{R}^d$, $\boldsymbol{\alpha} \in \mathbb{R}^d$, and $\boldsymbol{w} = \mathrm{diag}(w_1, \ldots, w_d) > 0$ are location, shape, and scale parameters, respectively. The skew-t_ν distribution has density

$$f_{\boldsymbol{\epsilon}}(\boldsymbol{z}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu) := 2\det(\boldsymbol{w})^{-1} t_d(\boldsymbol{x}; \boldsymbol{\Sigma}, \nu) T\left(\boldsymbol{\alpha}'\boldsymbol{x}\big((\nu+d)/(\nu+\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})\big)^{1/2}; \nu+d\right), \quad \boldsymbol{z} \in \mathbb{R}^d, \tag{6.4.3}$$

where $\boldsymbol{x} = \boldsymbol{w}^{-1}(\boldsymbol{z} - \boldsymbol{\xi})$, $T(y; \nu)$ denotes the univariate t_ν distribution function and

$$t_d(\boldsymbol{x}; \boldsymbol{\Sigma}, \nu) := \frac{\Gamma((\nu+d)/2)}{(\nu\pi)^{d/2}\Gamma(\nu/2)\det(\boldsymbol{\Sigma})^{1/2}} \left(1 + \frac{\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}{\nu}\right)^{-(\nu+d)/2}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

We refer to Azzalini and Dalla Valle (1996), Azzalini and Capitanio (2003) for details.

Our samples for skew-normal and skew-t_3 were simulated from the function `rmst` in the R Package `sn` by setting $\boldsymbol{\xi} = \mathbf{0}, \boldsymbol{\alpha} = (5, 2)', \boldsymbol{\Sigma} = \begin{pmatrix} 7 & 4 \\ 4 & 5 \end{pmatrix}$. In order to satisfy the classical conditions for M-estimation, we centered the simulated innovations about their mean, a centering which does not affect our R-estimators.

Figure 6.1 provides scatterplots of samples of size $n = 1000$ from the spherical normal, the skew-normal, the skew-t_3, and the Gaussian mixture described in Section 5.5.

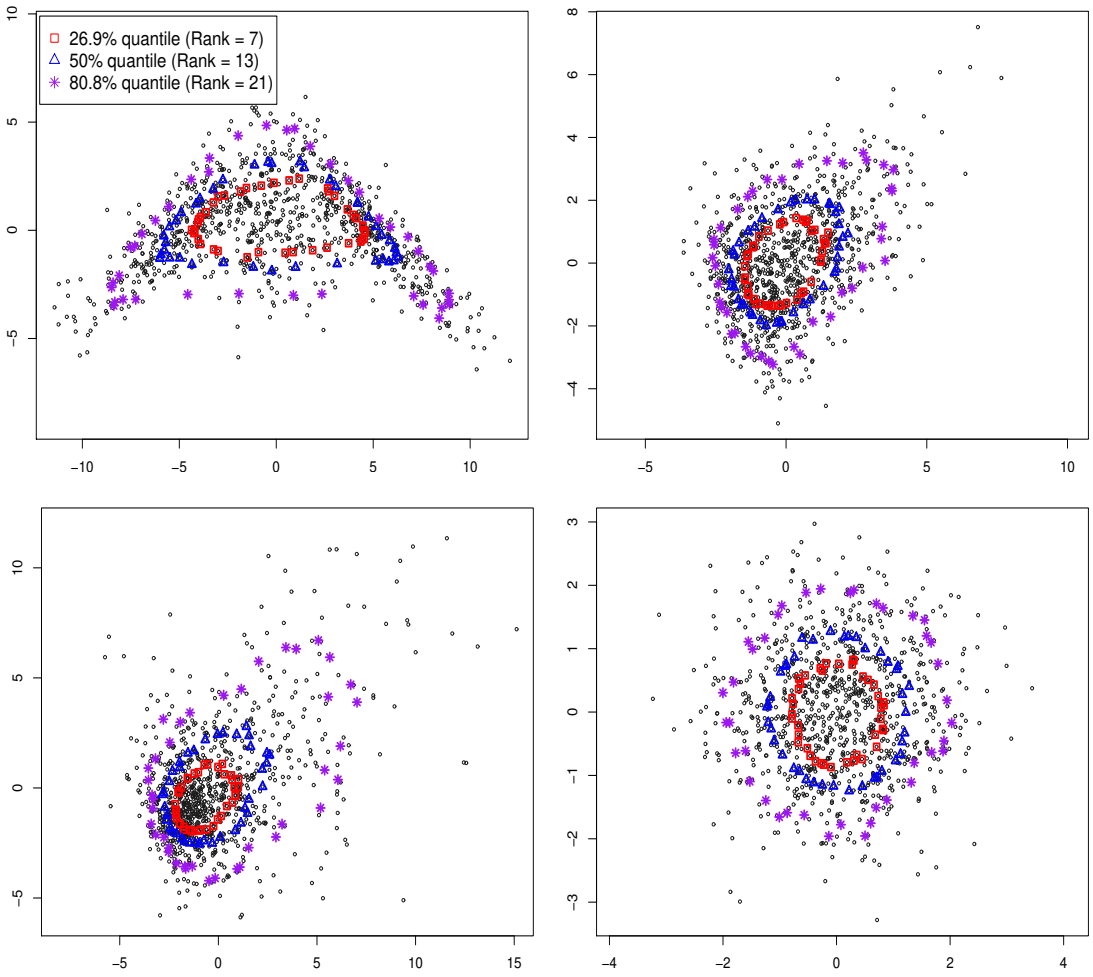
6.4.3 Additional numerical results

6.4.3.1 Large sample

As a complement to Section 5.5.1, we provide here, for sample size $n = 1000$, boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under skew-normal, skew-t_3, t_3 and non-spherical Gaussian innovations with covariance

$$\boldsymbol{\Sigma}_4 = \begin{pmatrix} 5 & 4 \\ 4 & 4.5 \end{pmatrix};$$

Figure 6.1: Empirical center-outward quantile contours (probability contents 26.9%, 50 %, and 80%, respectively) computed from $n = 1000$ points drawn from the Gaussian mixture (5.5.2) (top left), the skew-normal and skew-t_3 described in Section 6.4.2 (top right and bottom left) and, for a comparison, from a standard multivariate normal (bottom right).



See Figure 6.2, 6.3, 6.4 and 6.5, respectively.

Under skew-normal (Figure 6.2) and skew-t_3 (Figure 6.3) innovations, the vdW and Spearman R-estimators are less dispersed than other M-estimators, showing that they are more resistant to skewness. Under spherical t_3 innovations (Figure 6.4), outlying observations are relatively frequent and the QMLE is no longer root-n consistent. The RMLTSE does its job as a robustified estimator and slightly outperforms the R-estimators (the weakest of which is the sign-test score one). The non-spherical Gaussian boxplots (Figure 6.5) show that the vdW and Spearman R-estimators are quite similar to the QMLE.

6.4.3.2 Small sample and outliers

For sample size $n = 300$, we display here, in Figures 6.6, 6.7, 6.8, 6.9, and 6.10, the boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under the Gaussian mixture (5.5.2), spherical Gaussian, skew-normal, skew-t_3, and t_3, respectively. These pictures complement the boxplots available in Section 5.5.3, for the additive outlier case.

All boxplots, as well as Table 6.1 confirm the fact that, while doing equally well under spherical and Gaussian-tailed innovations, as the common practice QMLE, R-estimation is resisting skewness, heavy tails, non-elliptical contours, and the presence of additive outliers, sometimes better even than the robust RMLTSE.

Figure 6.2: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden) under skew-normal innovations (6.4.2); sample size $n = 1000$; $N = 300$ replications. The horizontal red line represents the actual parameter value.

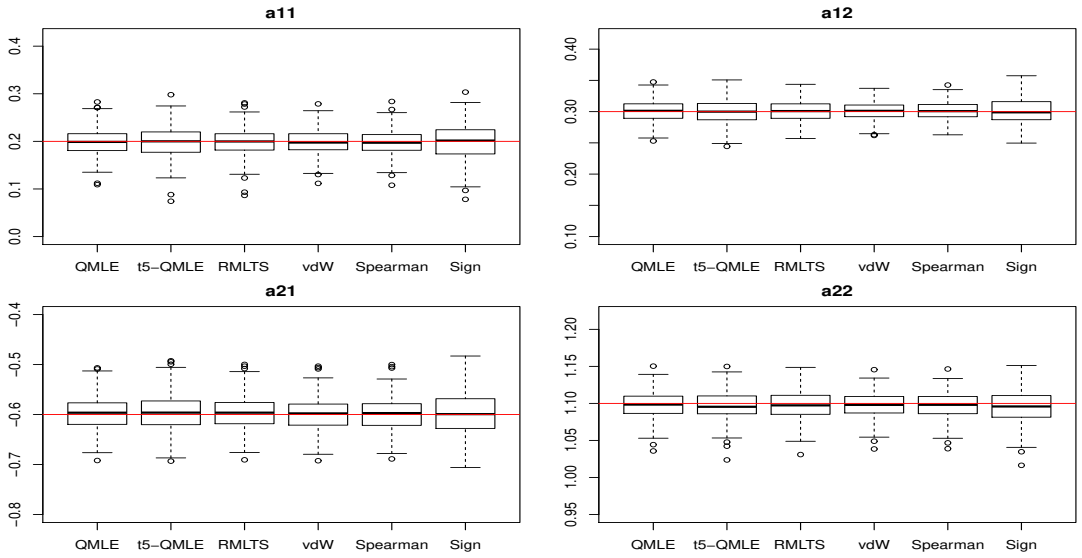


Figure 6.3: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden) under skew-t_3 innovations (6.4.3); sample size $n = 1000$; $N = 300$ replications. The horizontal red line represents the actual parameter value.

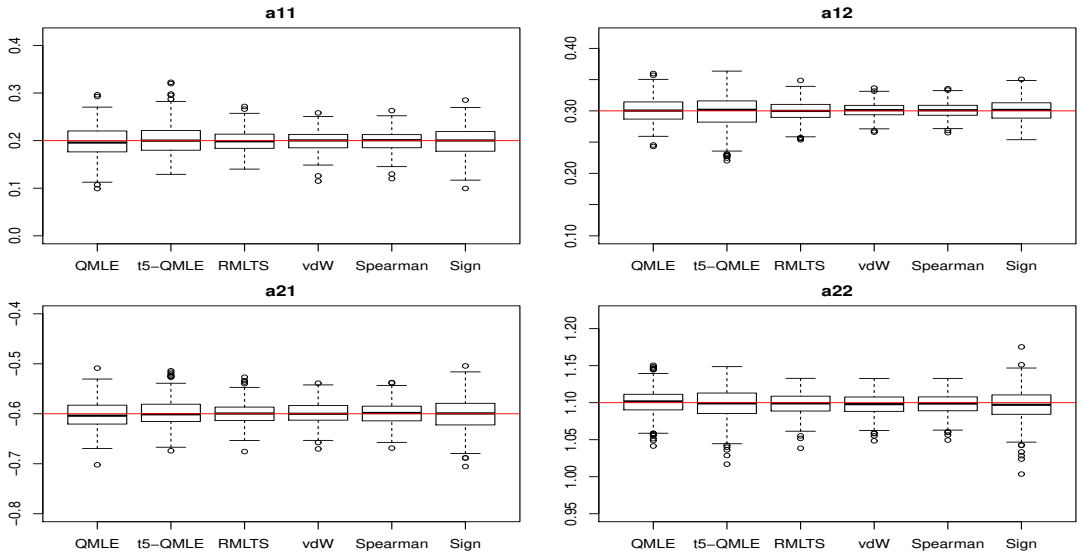


Figure 6.4: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under t_3 innovations; sample size $n = 1000$; $N = 300$ replications. The horizontal red line represents the actual parameter value.

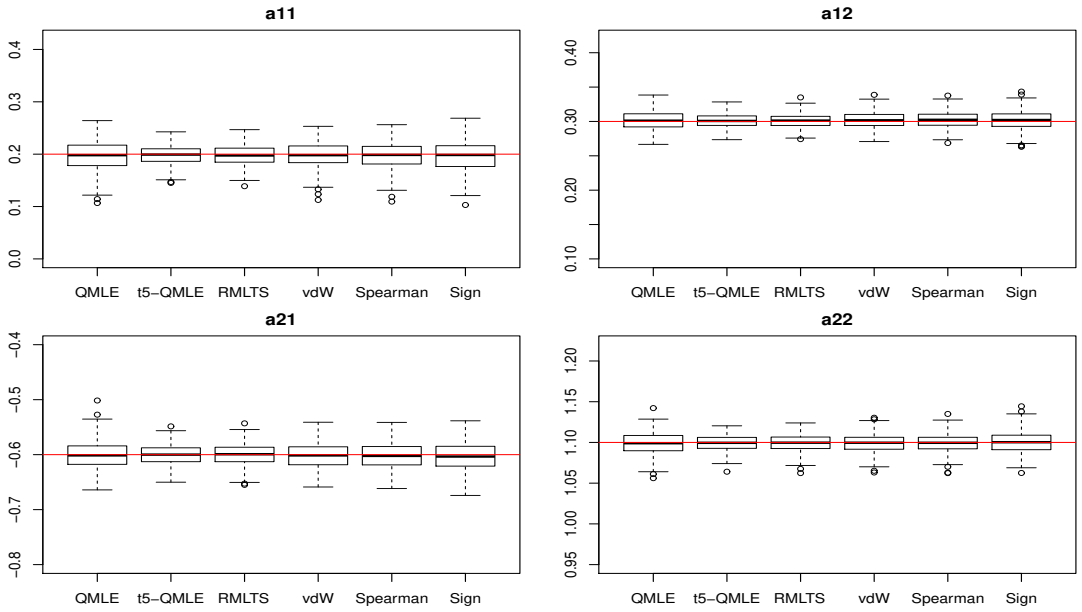


Figure 6.5: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under non-spherical Gaussian innovations; sample size $n = 1000$; $N = 300$ replications. The horizontal red line represents the actual parameter value.

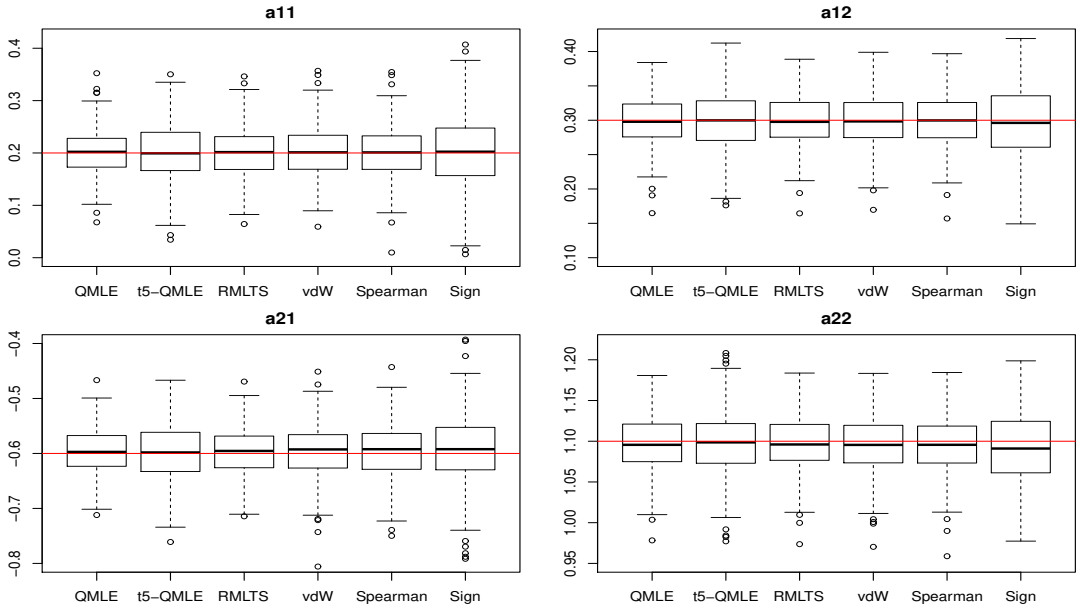


Table 6.1: The estimated bias ($\times 10^3$), MSE ($\times 10^3$), and overall MSE ratios of the QMLE, t_5-QMLE, RMLTSE, and R-estimators under various innovation densities. The sample size is $n = 300$; $N = 300$ replications.

	Bias ($\times 10^3$)				MSE ($\times 10^3$)				MSE ratio
	a_{11}	a_{21}	a_{12}	a_{22}	a_{11}	a_{21}	a_{12}	a_{22}	
(Spherical Normal)									
QMLE	-7.208	-2.006	2.639	-0.870	2.624	2.715	0.543	0.733	
t_5-QMLE	-8.352	-2.065	3.701	-1.071	2.783	2.796	0.580	0.751	0.957
RMLTS	-8.374	-2.423	3.481	-0.706	3.014	2.818	0.607	0.714	0.925
vdW	4.247	-3.994	-2.337	1.076	1.486	1.003	0.985	1.000	1.478
Spearman	5.041	-6.119	-3.395	3.332	1.661	1.204	1.165	1.292	1.243
Sign	6.124	-6.672	-4.254	4.294	2.586	1.839	1.487	0.992	0.958
(Mixture)									
QMLE	-3.430	-0.123	4.399	-1.814	2.751	0.550	1.000	0.213	
t_5-MLE	-1.593	0.240	5.467	-1.277	12.295	0.918	4.129	0.461	0.254
RMLTS	-2.459	-0.397	3.997	-1.392	2.707	0.578	1.025	0.220	0.997
vdW	-2.484	-0.007	5.065	1.348	1.427	0.368	0.733	0.379	1.554
Spearman	-2.632	0.742	5.160	1.104	1.329	0.379	0.694	0.332	1.652
Sign	-3.152	-0.066	10.017	1.164	4.313	0.745	2.283	0.566	0.571
(Skew-normal)									
QMLE	-9.045	-7.223	5.870	-2.116	3.564	3.308	1.087	1.022	
t_5-QMLE	-7.788	-7.028	6.400	-1.115	4.581	3.992	1.518	1.327	0.787
RMLTS	-9.558	-6.833	5.186	-1.844	3.988	3.574	1.200	1.140	0.907
vdW	-7.086	-1.523	7.358	-5.660	1.879	3.052	0.442	0.706	1.477
Spearman	-6.960	-1.101	7.198	-5.676	1.911	3.109	0.448	0.721	1.451
Sign	-12.525	0.748	10.592	-6.080	3.989	5.962	1.014	1.180	0.740
(Skew-t_3)									
QMLE	-11.108	-4.201	3.932	-1.327	3.148	2.710	1.446	1.209	
t_5-QMLE	1.801	5.000	3.371	-1.652	3.796	2.771	2.269	1.417	0.830
RMLTS	-3.378	0.428	4.358	-1.058	1.918	1.780	1.129	0.833	1.504
vdW	-7.152	0.232	6.544	-3.750	1.718	2.320	0.634	1.240	1.440
Spearman	-5.594	-1.927	6.402	-2.279	1.719	2.388	0.625	1.365	1.396
Sign	-3.380	-1.968	6.469	-0.033	4.816	4.863	1.900	2.054	0.624
(t_3)									
QMLE	0.168	-0.844	2.047	-1.063	2.279	2.593	0.647	0.658	
t_5-QMLE	-2.189	0.647	1.176	-1.347	1.160	1.215	0.339	0.343	2.021
RMLTS	-3.538	2.340	0.680	-1.734	1.343	1.377	0.379	0.358	1.787
vdW	-3.426	-0.037	3.681	-6.190	1.435	2.896	0.309	0.816	1.132
Spearman	-2.715	0.208	3.737	-5.768	1.387	2.930	0.306	0.788	1.141
Sign	-2.552	1.297	2.626	-6.454	2.842	5.634	0.564	2.045	0.557
(Additive outliers)									
QMLE	-154.990	-149.720	15.327	10.173	27.667	24.982	1.021	1.080	
t_5-QMLE	-110.645	-105.918	12.836	7.714	15.310	13.590	0.859	1.049	1.777
RMLTS	-76.970	-71.918	9.792	4.743	9.931	8.795	0.853	1.042	2.655
vdW	-3.426	-0.037	3.681	-6.190	1.435	2.896	0.309	0.816	10.034
Spearman	-2.715	0.208	3.737	-5.768	1.387	2.930	0.306	0.788	10.118
Sign	-2.552	1.297	2.626	-6.454	2.842	5.634	0.564	2.045	4.939

Figure 6.6: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under Gaussian mixture (sample size $n = 300$; $N = 300$ replications). The horizontal red line represents the actual parameter value.

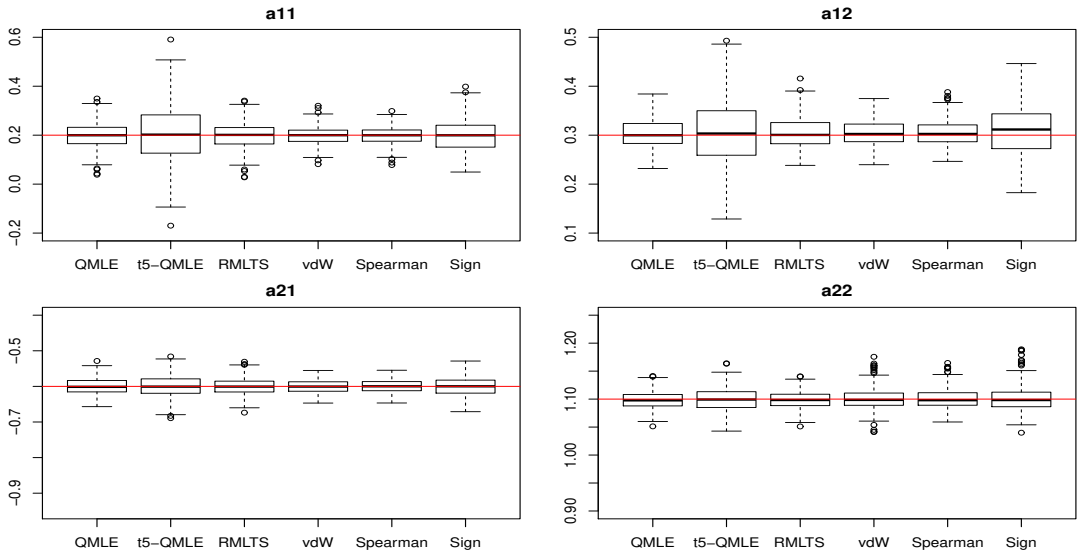


Figure 6.7: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under spherical Gaussian innovations; sample size $n = 300$; $N = 300$ replications. The horizontal red line represents the actual parameter value.

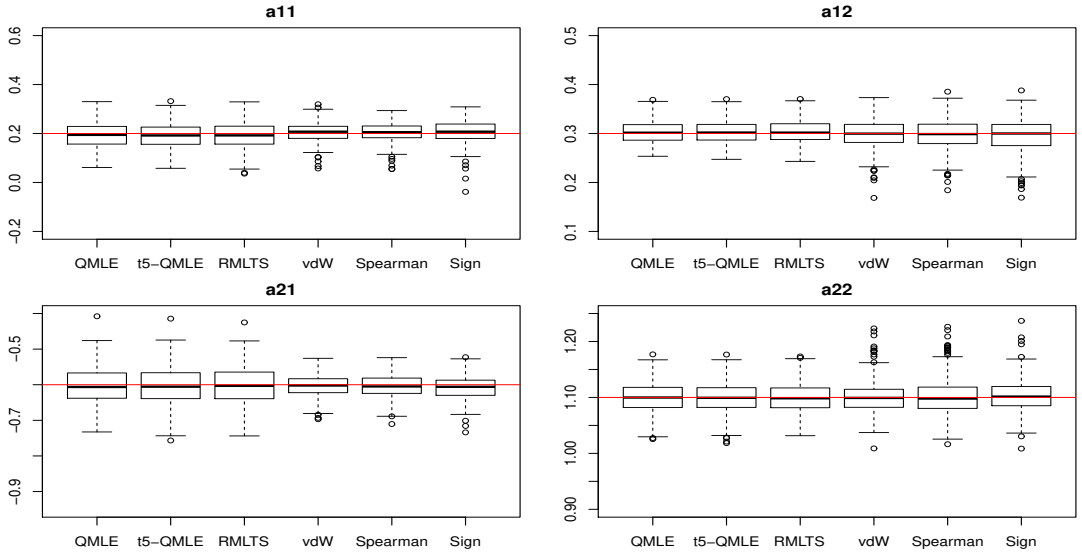


Figure 6.8: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under skew-normal innovations (6.4.2); sample size $n = 300$; $N = 300$ replications. The horizontal red line represents the actual parameter value.

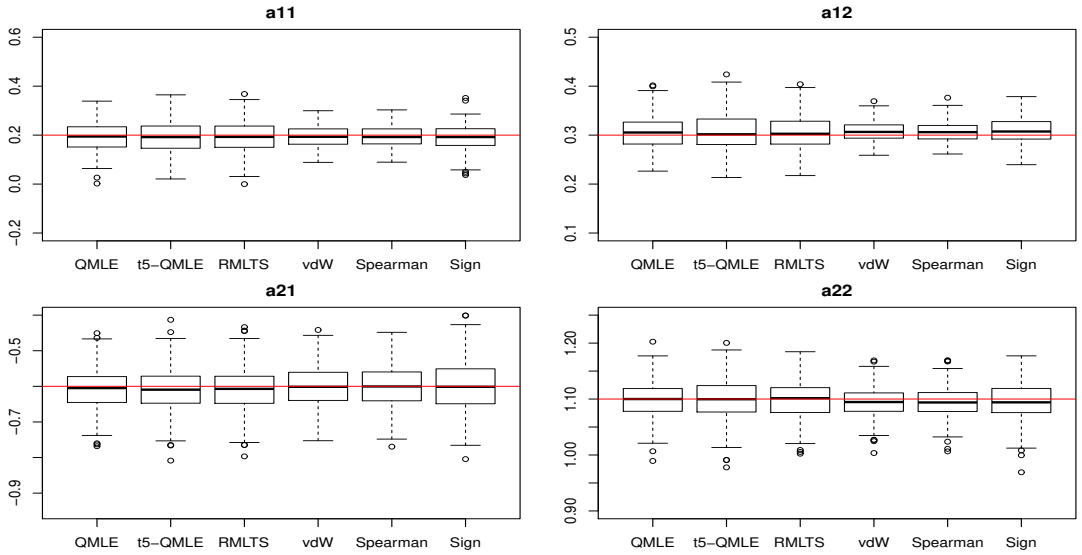


Figure 6.9: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under skew-t_3 innovations (6.4.3); sample size $n = 300$; $N = 300$ replications. The horizontal red line represents the actual parameter value.

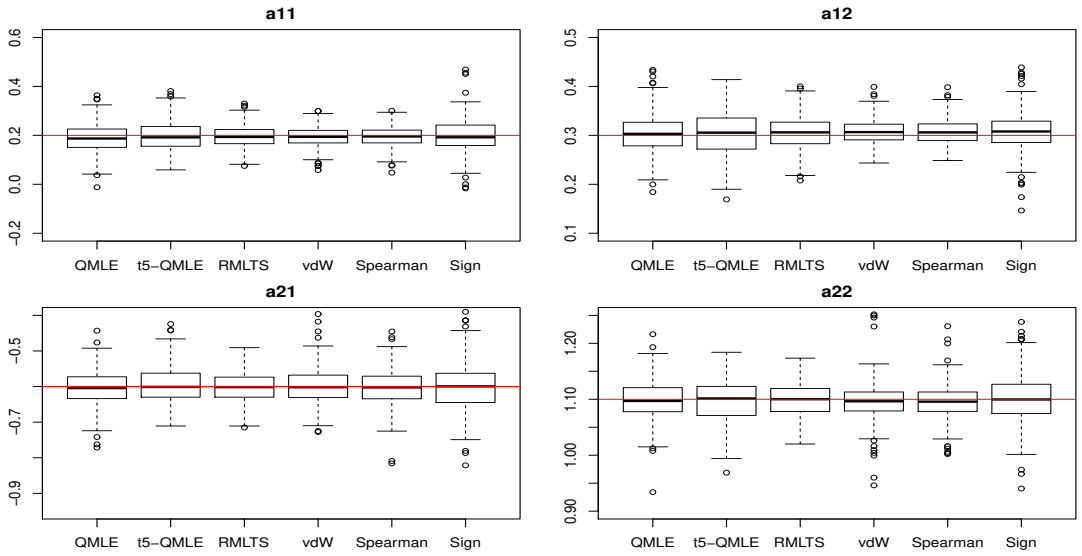
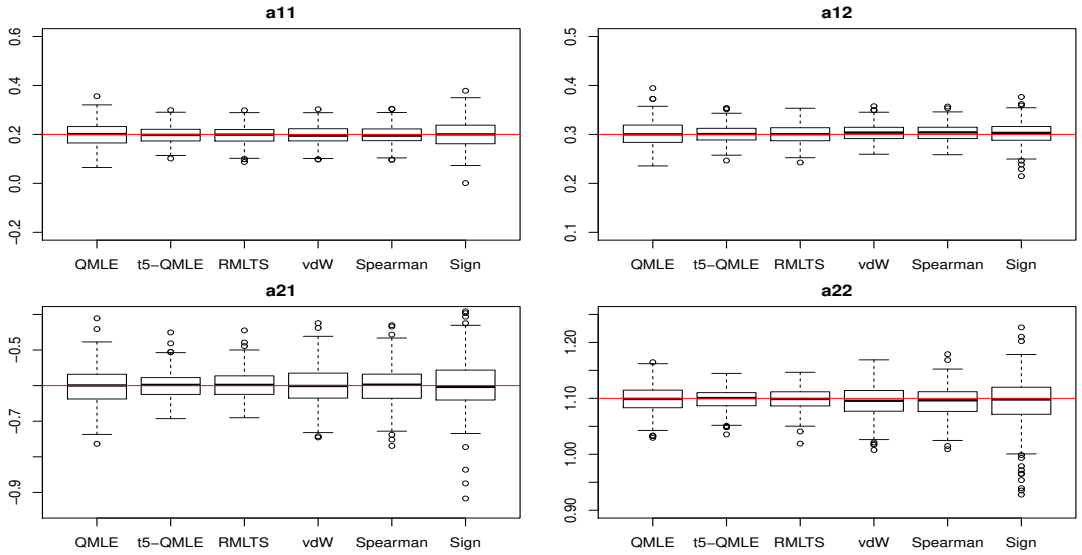


Figure 6.10: Boxplots of the QMLE, t_5-QMLE, RMLTSE, and R-estimators (sign test, Spearman, and van der Waerden scores) under spherical t_3 innovations; sample size $n = 300$; $N = 300$ replications. The horizontal red line represents the actual parameter value.



6.4.4 Higher dimension

Due to the rapid growth of their number of parameters, VARMA models are not meant for the analysis of high-dimensional time series (where different approaches are in order—see, e.g., Hallin et al. (2020c)). One may wonder, however, whether the attractive properties of R-estimators extend beyond the bivariate context. We therefore provide here some numerical results in dimension $d = 3$.

Consider the three-dimensional VAR(1) model

$$(\boldsymbol{I}_3 - \boldsymbol{A}L)\,\boldsymbol{X}_t = \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z},$$

with $\boldsymbol{\theta}' := \mathrm{vec}'(\boldsymbol{A}) = (0.55, 0.2, 0.13, -0.2, 0.5, -0.1, 0.1, 0.11, 0.6)$ satisfying Assumption (A1) of Section 5.2.2. We are limiting our investigation to two selected innovation densities: the spherical three-dimensional Gaussian and the Gaussian mixture

$$\frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + \frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) + \frac{1}{4}\mathcal{N}(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3), \tag{6.4.4}$$

with

$$\boldsymbol{\mu}_1 = (-5, -5, 0)', \ \boldsymbol{\mu}_2 = (5, 5, 2)', \ \boldsymbol{\mu}_3 = (0, 0, -3)'$$

and

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 7 & 3 & 5 \\ 3 & 6 & 1 \\ 5 & 1 & 7 \end{bmatrix}, \ \boldsymbol{\Sigma}_2 = \begin{bmatrix} 7 & -5 & -3 \\ -5 & 7 & 4 \\ -3 & 4 & 5 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma}_3 = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For the computation of the center-outward ranks and signs, we used the algorithm described in Section 6.3.2 with $n_R = 15, n_S = 66, n_0 = 10$. For numerical implementation, we generated regular grids on the sphere via the routine `UnitSphere` in R package `mvmesh`, where we refer to for details. The boxplots for the Gaussian mixture and spherical Gaussian innovations are displayed in Figures 6.11 and 6.12, respectively. Inspection of Figures 6.11 and 6.12 yields the same conclusions as in the bivariate motivating example (Figures 5.1).

6.5 Impulse response function: a compendium

As explained in Section 5.6, impulse response functions provide a convenient way of exploring the relation between the components of multiple time series. In particular, it is used to study the impact of changes in one variable on its own future values and those of other time series. For the d-dimensional VARMA(p, q) model in (5.2.1), the impulse response function can be obtained as follows.

Write (5.2.1) under the corresponding VMA(∞) form

$$\boldsymbol{X}_t = \boldsymbol{W}(L)\boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z},$$

where

$$\boldsymbol{W}(L) := \sum_{l=0}^{\infty} \boldsymbol{W}_l L^l = \left(\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i L^i \right)^{-1} \left(\boldsymbol{I}_d + \sum_{j=1}^{q} \boldsymbol{B}_j L^j \right) \boldsymbol{\epsilon}_t$$

with $\boldsymbol{W}_l$ being the coefficient at lag l.

Now, suppose that we are interested in studying the impact on $\boldsymbol{X}_{t+h}$, $h \geq 0$ of increasing the value at time t of the kth series X_{kt}, $1 \leq k \leq d$ by one unit. Without

Figure 6.11: Boxplots of the QMLE and R-estimator (van der Waerden scores) under the Gaussian mixture innovation density (6.4.4) for $d = 3$; sample size $n = 1000$; $N = 300$ replications. In each panel, the MSE ratio of the QMLE with respect to the R-estimator is reported. The horizontal red line represents the actual parameter value.

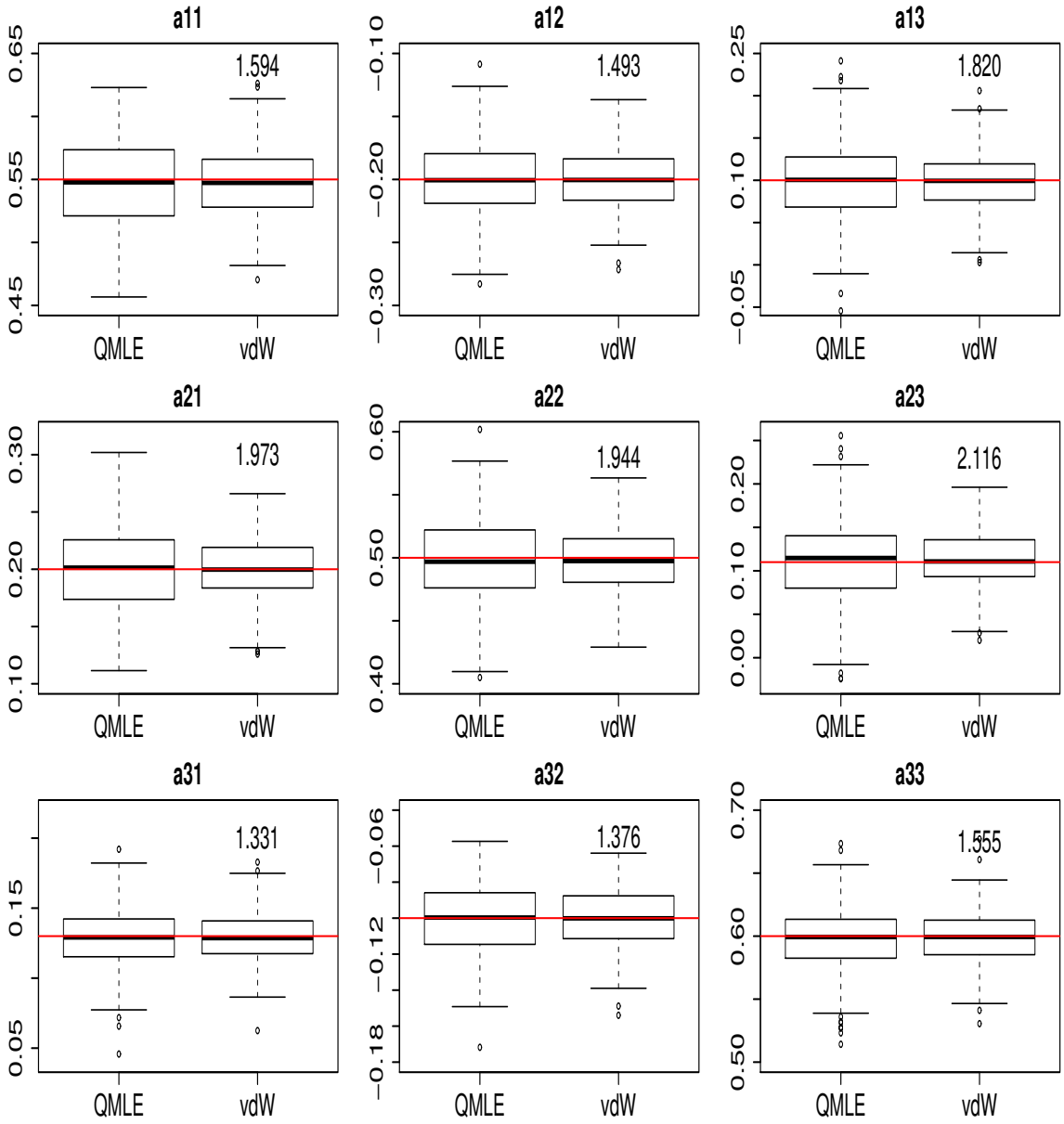
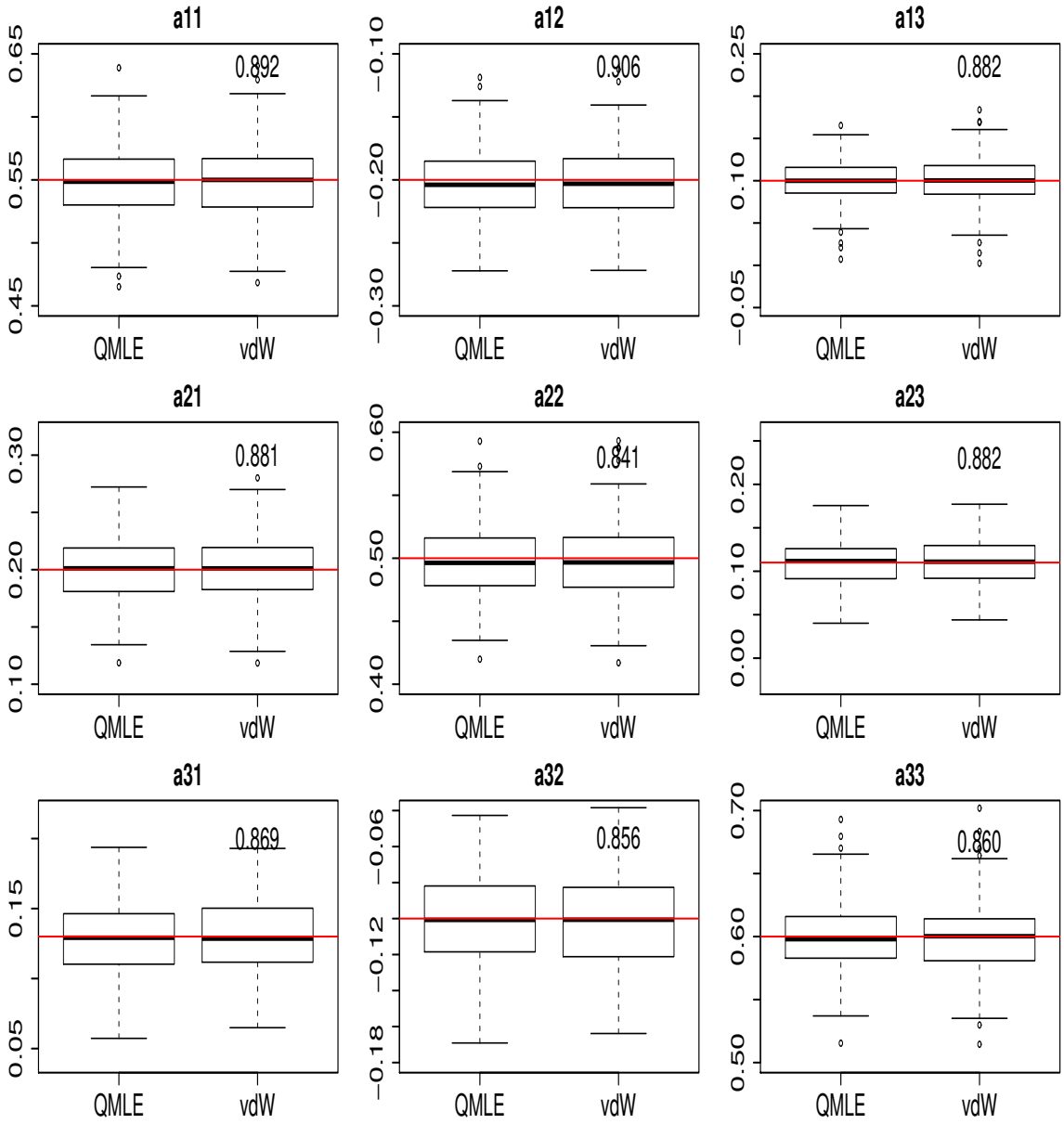


Figure 6.12: Boxplots of the QMLE and R-estimator (van der Waerden scores) under spherical Gaussian for $d = 3$; sample size $n = 1000$; $N = 300$ replications. In each panel, the MSE ratio of the QMLE with respect to the R-estimator is reported. The horizontal red line represents the actual parameter value.



loss of generality, we can assume $t = 0$. Setting $\boldsymbol{X}_t = \boldsymbol{0}$ for $t \leq 0$, $\boldsymbol{\epsilon}_0 = \boldsymbol{e}_k$ and $\boldsymbol{\epsilon}_t = \boldsymbol{0}$ for $t > 0$, where $\boldsymbol{e}_k$ denotes the kth unit vector in the canonical basis of $\mathbb{R}^d$, we then have

$$\boldsymbol{X}_0 = \boldsymbol{\epsilon}_0 = \boldsymbol{e}_k, \quad \boldsymbol{X}_1 = \boldsymbol{W}_1\boldsymbol{\epsilon}_0 = \boldsymbol{W}_{1,k}, \quad \boldsymbol{X}_2 = \boldsymbol{W}_2\boldsymbol{\epsilon}_0 = \boldsymbol{W}_{2,k}, \quad ...,$$

where $\boldsymbol{W}_{l,k}$ denotes the kth column of $\boldsymbol{W}_l$. Therefore, the impact under study is reflected in the kth column of the coefficient matrix $\boldsymbol{W}_h$. For this reason, the coefficient matrices $\{\boldsymbol{W}_{h,k}; h \geq 0\}$ are referred to as the coefficients of impulse response functions; see Tsay (2014, Chapter 2 and 3) for further details.

Chapter 7

Conclusions

This thesis aims at developing robust estimation procedures for both univariate (GARCH) and multivariate (VARMA) time series models. Specifically, for GARCH models, which are widely-used to analyse financial time series, two types of estimators—M- and R-estimators are considered and a generalized weighted bootstrap technique is employed to approximate their distributions. For semi-parametric VARMA models (with the innovation density playing the role of nuisance parameter), which have wide applications in e.g. economics and biology, thanks to the novel concepts of center-outward ranks and signs recently proposed by Hallin (2017), we propose a class of R-estimators, which successfully extend R-estimation to the setting of non-elliptical distributions. Conclusions of this thesis are listed as follows.

- Both simulation and real data analysis show that compared to the routinely-applied QMLE, the M-estimators for GARCH models, which are $\sqrt{n}$-consistent under a mild moment assumption, are more robust under heavy-tailed distributions. For instance, the μ- and Cauchy's estimators achieve much lower MSE than the QMLE under the $t(3)$ and $t(2.2)$ distributions, where the fourth moment of the innovation term is infinite. Under a finite sample size, we show via simulations that the weighted bootstrap, which is easy-to-implement, provides better approximation of distributions of the M-

estimators than the one based on the asymptotic normality.

- The R-estimators for GARCH models are asymptotic normal under the existence of a more than second moment of the errors. Thus, similar to the M-estimators, they are more robust than the QMLE, which is asymptotic normal under the finite fourth moment assumption. Moreover, the R-estimators are highly efficient. For example, the van der Waerden or normal score is shown to achieve almost the same efficiency as the QMLE even under the normal distribution for a finite sample, and it is more efficient than the QMLE under heavy-tailed distributions. In terms of the weighted bootstrap of the R-estimators, our extensive simulations reveal excellent coverage rates of the approximations.

- The center-outward R-estimators for VARMA models based on Le Cam one-step approach are proved to be $\sqrt{n}$-consistent and asymptotic normal under a broad class of innovation densities including, e.g., multimodal mixtures of Gaussians and multivariate skew-t distributions. Simulation study under large and small sample sizes shows that our R-estimators significantly outperform the QMLE under non-elliptical innovations, and they are more robust against additive outliers than the QMLE. Also, similar to univariate models, the good performance of the R-estimation is not obtained at the cost of poor performance under normal innovations.

This thesis also discusses computational aspects and algorithms related to the implementation of the methodology. In terms of the M- and R-estimators for GARCH models, the estimators can easily be obtained by using (3.4.1) and (4.2.10), respectively; their bootstrap estimators can also easily be computed by using (3.4.2) and (4.4.1), respectively. Algorithm of computing the R-estimators for VARMA models is given in Algorithm 1, which is more computational costly than those of the robust estimators for GARCH models since it involves solving the optimization problem in (5.3.1). This thesis uses a Hungarian algorithm, which is available in

the `clue` R package, to solve the problem. To reduce the computation cost, one may consider more efficient algorithms available in the measure transportation literature; see, e.g., Peyré and Cuturi (2019).

For future work, there are many related topics that may worth exploring and here we only name a few of them.

- This thesis shows advantage of the M- and R-estimators over the QMLE for GARCH models under heavy-tailed error distributions via extensive simulation. However, only symmetric distributions are considered, and whether this advantage still holds under skew distributions needs to be investigated.

- This thesis investigates the bootstrap approximation of the distributions of the R-estimators for GARCH models through simulation study. It remains to derive the asymptotic results of the bootstrap to examine whether it is consistent or not. The proof, however, could be a mathematically challenging problem since ranks are integer-valued discontinuous functions.

- The R-estimators for GARCH models are not distribution-free since the objective functions involve both the ranks and the observations. Another definition of the R-estimators is possible by following the approach in Chapter 5—using only the ranks and two score functions. Could these R-estimators outperform our R-estimators in Chapter 4 under heavy-tailed or skew distributions would be an interesting problem to explore.

- The technique used to prove the asymptotic normality of the R-estimators for GARCH models is based on some results for empirical processes in Section 4.7. Can we simplify the proofs and relax the moment assumption on the innovation term by using the local asymptotic normality of GARCH models and Le Cam's asymptotic theory as in Chapter 5?

- When it comes to R-estimation in the multivariate settings, to our best knowledge, this thesis is the first successful attempt to employ the center-

outward ranks and signs. It would be interesting to extend the center-outward R-estimation to other models, such as dynamic conditional correlation models of Engle (2002) and nonlinear structural vector autoregressive of Shen et al. (2019). To achieve these goals, a major step is to derive the local asymptotic normality results of these models. For a complex model, this is expected to be a challenging problem.

- Hallin (2017) derived various important properties, such as distribution-freeness and the Glivenko-Cantelli theorem, of the center-outward ranks and signs in dimension $d \geq 2$. However, whether these properties will hold or not in high dimension remains a open problem. Moreover, can we also define ranks and signs for directional data based on the measure transportation theory?

- We remark that applications of the center-outward ranks and signs are not only limited to R-estimation. There are broad possible applications in statistics and econometrics, and some of them, as listed in Hallin (2017), include goodness-of-fit tests, estimating multivariate value-at-risk and multivariate expected shortfall by using center-outward quantile contours, and so on.

Bibliography

[1] Allal, J., Kaaouachi, A., and Paindaveine, D. (2001) R-estimation for ARMA models, *Journal of Nonparametric Statistics*, 13, 815–831.

[2] Andreou, E. and Werker, B. (2015). Residual-based rank specification tests for AR-GARCH type models. *Journal of Econometrics*, 185, 305–331.

[3] Andrews, B. (2008). Rank-based estimation for autoregressive moving average time series models. *Journal of Time Series Analysis*, 29, 51–73.

[4] Andrews, B. (2012). Rank-based estimation for GARCH process. *Econometric Theory*, 28, 1037–1064.

[5] Angelidis, T., Benos, A. and Degiannakis, S. (2004). The use of GARCH models in VaR estimation. *Statistical Methodology*, 1, 105–128.

[6] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B*, 65, 367–389.

[7] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715–726

[8] Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.*, 37, 577–580.

[9] Baillie, R.T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73, 5–59.

[10] Bauwens, L., Hafner, C. and Laurent, S. (2012). *Handbook of Volatility Models and Their Applications*, John Wiley and Sons.

[11] Berkes, I., Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation, *Bernoulli*, 9, 201–227.

[12] Berndt, E., Hall, B., Hall, R. and Hausman, J. (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement*, 3, 653–665.

[13] Bickel, P.J. (1965). On some robust estimates of location. *Annals of Mathematical Statistics*, 36, 847–858.

[14] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore, MD.

[15] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. Wiley.

[16] Boeckel, M., Spokoiny, V., and Suvorikova, A. (2018). Multivariate Brenier cumulative distribution functions and their application to nonparametric testing, *arXiv:1809.04090*.

[17] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 100, A307–A327.

[18] Bose, A. and Mukherjee K. (2009). Bootstrapping a weighted linear estimator of the ARCH parameters. *Journal of Time Series Analysis*, 30, 315–331.

[19] Bougerol, P. and Picard, N. (1992a). Strict stationarity of generalized autoregressive processes. *Annals of Probability*, 20, 1714–1730.

[20] Bougerol, P. and Picard, N. (1992b). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52, 115–127.

[21] Carlier, G., Chernozhukov, V., and Galichon, A. (2016). Vector quantile regression, *Annals of Statistics*, 44, 1165–1192.

[22] Cassart, D., Hallin, M., and Paindaveine, D. (2010). On the estimation of cross-information quantities in R-estimation. In J. Antoch, M. Hušková and P.K. Sen, Eds: *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková*, I.M.S., 35–45.

[23] Chatterjee, S. and Bose, A. (2005). Generalized bootstrap for estimating equations. *Annals of Statistics*, 33, 414–436.

[24] Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks, and signs. *Annals of Statistics*, 45, 223–256.

[25] Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, 4, 841–862.

[26] Christoffersen, P. (2012). *Elements of Financial Risk Management.* Academic Press; 2nd edition.

[27] Christoffersen, P. and Goncalves, S. (2005). Estimation risk in financial risk management. *The Journal of Risk* 7, 1–28.

[28] Corradi, V. and Iglesias, E.M. (2008). Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Journal of Econometrics*, 144, 500–510.

[29] Croux, C. and Joossens, K. (2008). Robust estimation of the vector autoregressive model by a least trimmed squares procedure. *COMPSTAT 2008*, 489–501.

[30] Cuesta-Albertos, J.A. and Matrán, C. (1989). Notes on the Wasserstein metric in Hilbert spaces, *Annals of Probability*, 17, 1264–1276.

[31] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*, Cambridge, Cambridge University Press.

[32] Deb, N. and Sen, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *arXiv preprint arXiv:1909.08733*.

[33] Dick, J. and Pillichshammer, F. (2014). Discrepancy theory and quasi-Monte Carlo integration, in W. Chen, A. Sirvastava and G. Travaglini, Eds, *A Panorama of Discrepancy Theory*, 539–620, Springer, New York.

[34] Duan, J.-C. (1995). The GARCH option pricing model. *Mathematical Finance*, 5, 13–32.

[35] Duan, J.-C. and Simonato, J.-G. (2001). American option pricing under GARCH by a Markov chain approximation. *Journal of Economic Dynamics and Control*, 25, 1689–1718.

[36] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1–26.

[37] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia, Pa.: Society for Industrial and Applied Mathematics.

[38] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica (pre-1986)*, 50, 987–1007.

[39] Engle, R.F. (2002). Dynamic Conditional Correlation. *Journal of Business & Economic Statistics*, 20, 339–350.

[40] Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5, 1–50.

[41] Engle, R.F. and Mustafa, C. (1992). Implied ARCH models from option prices. *Journal of Econometrics*, 52, 289–311.

[42] Figalli, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Analysis*, 177, part B, 413–421.

[43] Francq, C. and Zakoïan, J. (2009). Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association*, 104, 313–324.

[44] Francq, C. and Zakoïan, J. (2010). *GARCH models: structure, statistical inference and financial applications*, Chichester, John Wiley and Sons.

[45] Fujita, A., Sato, J.R., Garay-Malpartida, H.M., Yamaguchi, R., Miyano, S., Sogayar, M.C. and Ferreira, C.E. (2007). Modeling gene expression regulatory networks with the sparse vector autoregressive model. *BMC systems biology*, 1, 39.

[46] Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics*, 3, 551–579.

[47] Ghosal, P. and Sen, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. *arXiv preprint arXiv:1905.05340*.

[48] Ghosh, J.K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.*, 42, 1957–1961.

[49] Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779–1801.

[50] Hájek, J., Šidák, Z. and Sen, P.K. (1999). *Theory of Rank Tests*, 2nd edition. San Diego: Academic Press.

[51] Hall, P. and Yao, Q. (2003). Inference in arch and garch models with heavy–tailed errors. *Econometrica*, 71, 285–317.

[52] Hallin, M. (1986). Non-stationary q-dependent processes and time-varying moving-average models: invertibility properties and the forecasting problem. *Advances in Applied Probability*, 18, 170–210.

[53] Hallin, M. (2017). On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$. ECARES WP. Available at `https://ideas.repec.org/p/eca/wpaper/2013-258262.html`.

[54] Hallin, M., del Barrio, E., Cuesta-Albertos, J., and Matrán, C. (2020a). Center-outward distribution and quantile functions, ranks, and signs in dimension d: a measure transportation approach, *Annals of Statistics*, in press.

[55] Hallin, M., Hlubinka, D., and Hudecová, Š. (2020b). Efficient center-outward rank tests for multiple-output regression and MANOVA. Available at `http://arxiv.org/abs/2007.15496`

[56] Hallin, M., Ingenbleek, J-Fr., and Puri, M.L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Annals of Statistics,* 13, 1156–1181.

[57] Hallin, M. and La Vecchia, D. (2017). R-estimation in semiparametric dynamic location-scale models. *Journal of Econometrics,* 2, 233–247.

[58] Hallin, M. and La Vecchia, D. (2020). A simple R-estimation method for semiparametric duration models. *Journal of Econometrics,* 218, 736–749.

[59] Hallin, M., Barigozzi, M., Forni M., Zaffaroni P., and Lippi M. (2020c). *Time Series In High Dimension: The General Dynamic Factor Model*. World Scientific Publishing Company.

[60] Hallin, M., La Vecchia, D. and Liu, H. (2020). Center-Outward R-Estimation for Semiparametric VARMA Models. *Journal of the American Statistical Association*, 1–14.

[61] Hallin, M., Oja, H., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape: II Optimal R-estimation of shape. *Annals of Statistics,* 34, 2757–2789.

[62] Hallin, M. and Paindaveine, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Annals of Statistics*, 30, 1103–1133.

[63] Hallin, M. and Paindaveine, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli*, 8, 787–815.

[64] Hallin, M. and Paindaveine, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Annals of Statistics*, 6, 2642–2678.

[65] Hallin, M. and Paindaveine, D. (2005). Asymptotic linearity of serial and nonserial multivariate signed rank statistics. *Journal of Statistical Planning and Inference*, 136, 1–32.

[66] Hallin, M. and Puri, M.L. (1994). Aligned rank tests for linear models with autocorrelated errors. *Journal of Multivariate Analysis,* 50, 175–237.

[67] Hallin, M., van den Akker, R., and Werker, B. (2015). On quadratic expansions of log-likelihoods and a general asymptotic linearity result. In

M. Hallin, D. Mason, D. Pfeifer, and J. Steinebach Eds, *Mathematical Statistics and Limit Theorems, Festschrift in Honor of Paul Deheuvels*, Springer, 147–166.

[68] Hallin, M. and Werker, B.J.M. (2003). Semiparametric efficiency, distribution-freeness and invariance. *Bernoulli*, 9 , 137–165.

[69] Heston, S.L. and Nandi, S. (2000). A closed form GARCH option pricing model. *Review of Financial Studies*, 13, 585–625.

[70] Hettmansperger, T.P. and McKean, J.W. (2010). *Robust Nonparametric Statistical Methods*, Boca Raton, London and New York, Taylor and Francis Group.

[71] Hodges, J. and Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t-test. *Annals of Mathematical Statistics*, 34, 324–335.

[72] Hodges, J. and Lehmann, E.L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34, 598–611.

[73] Horvath, L. and Liese, F. (2004). L_p-estimators in ARCH models. *Journal of Statistical Planning and Inference*, 119, 277–309.

[74] Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73–101.

[75] Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Annals of Statistics*, 5, 799–821.

[76] Huber, P.J. and Ronchetti, E.M. (2009). *Robust statistics*. New Jersey, John Wiley and Sons, 2nd edition.

[77] Iqbal, F. and Mukherjee, K. (2010). M-estimators of Some GARCH-type Models; Computation and Application. *Statistics and Computing*, 4, 435–445.

[78] Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics*, 43, 1449–1458.

[79] Judd, K.L. (1998). *Numerical Methods in Economics*, MIT Press, Cambridge, MA.

[80] Jurečková, J. (1971). Nonparametric estimate of regression coefficients. *Annals of Mathematical Statistics*, 42, 1328–1338.

[81] Jurečková, J. and Sen, P.K. (1996) *Robust Statistical Procedures: Asymptotics and Interrelations*, New York, Chichester, Brisbane, Toronto and Singapore, John Wiley and Sons.

[82] Jeong, M. (2017). Residual-based garch bootstrap and second order asymptotic refinement. *Econometric Theory*, 3, 779–790.

[83] Koul, H.L. (1971). Asymptotic behavior of a class of confidence regions based on ranks in regression. *Annals of Mathematical Statistics*, 42, 466–476.

[84] Koul, H.L. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics*, 22, 540–562.

[85] Koul, H.L. and Saleh, A.M.E. (1993). R-estimation of the parameters of autoregressive AR(p) models. *Annals of Statistics*, 21, 534–551.

[86] Kreiss, J.-P. (1987). On adaptative estimation in stationary ARMA processes, *Annals of Statistics,* 15, 112–133.

[87] Kunsch, H.R. (1989). The Jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217–1241.

[88] Kupiec, P.H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, 2, 73–84.

[89] Lahiri, S.N. (2010). *Resampling Methods for Dependent Data*, New York, Springer.

[90] Le Cam, L. and Yang, G.L. (2000). *Asymptotics in Statistics : Some basic concepts (2nd ed.)*. New York: Springer.

[91] Lee, S.W. and Hansen, B.E. (1994). Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator, *Econometric Theory*, 10, 29–52.

[92] Linton, O., Pan, J. and Wang, H. (2010). Estimation for a nonstationary semi-strong garch(1,1) model with heavy–tailed errors. *Econometric Theory*, 26, 1–28.

[93] Liu, H. (2016). M-estimation and bootstrap in GARCH models. *Dissertation for the MSc in Quantitative Finance*, September, 2016, Lancaster University.

[94] Liu, R.Y. (1992). Data depth and multivariate rank tests, in Y. Dodge, Ed., L^1 *Statistics and Related Methods*. North-Holland, Amsterdam, 279–294.

[95] Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests, *Journal of the American Statistical Association,* 88, 257–260.

[96] Lumsdaine, R.L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometrica*, 64, 575–596.

[97] Mancini, L. and Trojani, F. (2006). Robust semiparametric bootstrap methods for Value at Risk prediction under GARCH-type volatility process. SSRN Working Paper, University of St. Gallen.

[98] Maronna, R., Martin D., Yohai V., and Salibian-Barrera M. (2019). *Robust Statistics: Theory and Methods (with R)*, Wiley.

[99] Mukherjee, K. (2007). Generalized R-estimators under conditional heteroscedasticity. *Journal of Econometrics*, 141, 383–415.

[100] Mukherjee, K. (2008). M-estimation in GARCH models. *Econometric Theory*, 24, 1530–1553.

[101] Mukherjee, K. (2020). Bootstrapping M-estimators in GARCH models. *Biometrika*, 107, 753–760.

[102] Mukherjee, K. and Bai, Z. (2002). R-estimation in autoregression with square-integrable score function. *Journal of Multivariate Analysis*, 81, 167–186.

[103] Mukherjee, K. and Wang, Y. (2014). On the computation of R-estimators. In Lahiri, S., Schick, A., SenGupta, A. and Sriram, T., Eds: *Contemporary developments in statistical theory: A Festschrift for Hira Lal Koul*, Springer, 279–288.

[104] Muler, N. and Yohai, V.J. (2008). Robust estimates for GARCH models. *Journal of Statistical Planning and Inference*, 138, 2918–2940.

[105] Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory*, 6, 318–334.

[106] Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347–370.

[107] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63, SIAM, Philadelphia, PA.

[108] Oja, H. (2010). *Multivariate Nonparametric Methods with R: an approach based on spatial signs and ranks*. Springer, New York.

[109] Orhan, M. and Koksal, B. (2012). A comparison of GARCH models for VaR estimation. *Expert Systems with Applications*, 3, 3582–3592.

[110] Panaretos, V. M. and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Application*, 6, 405–431.

[111] Peyré, G. and Cuturi, M. (2019). Computational optimal transport with applications to Data Science. *Foundations and Trends® in Machine Learning*, 1, 355–607.

[112] Peng L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, 90, 967–975.

[113] Puri, M.L. and Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis*. John Wiley & Sons, New York.

[114] Rachev, S.T. and Rüschendorf, L. (1998). *Mass Transportation Problems* I and II, Springer, New York.

[115] Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York.

[116] Shen, Y., Giannakis, G.B. and Baingana, B. (2019). Nonlinear structural vector autoregressive models with application to directed brain networks. *IEEE Transactions on Signal Processing*, 67, 5325–5339.

[117] Shi, H., Drton, M. and Han, F. (2020). Distribution-free consistent independence tests via center-outward ranks and signs. *Journal of the American Statistical Association*, (just-accepted), 1–34.

[118] So, M.K.P. and Yu, P.L.H. (2006). Empirical analysis of GARCH models in value at risk estimation. *Journal of International Financial Markets, Institutions and Money*, 16, 180–197.

[119] Taylor, S. (2005). *Asset Price Dynamics, Volatility, and Prediction*. Princeton, N.J.: Princeton University Press.

[120] Terpstra, J.T., McKean, J.W., and Naranjo, J.D. (2001). GR-estimates for an autoregressive time series. *Statistics & Probability Letters*, 51, 165–172.

[121] Tsay, R.S. (2010). *Analysis of Financial Time Series,* Hoboken, N.J.: Wiley. 3rd edition.

[122] Tsay, R.S. (2014). *Multivariate Time Series Analysis: With R and Financial Applications.* Hoboken, New Jersey: Wiley.

[123] van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press.

[124] van Eeden, C. (1972). An analogue, for signed rank statistics, of Jurečková's asymptotic linearity theorem for rank statistics. *Annals of Mathematical Statistics*, 43, 791–802.

[125] Weiss, A.A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory*, 2, 107–131.

[126] Zakoïan, J.M. (1994). Threshold heteroscedastic models. *Journal of Economic Dynamics and Control*, 18, 931–955.

[127] Zhou, W.X., Bose, K., Fan, J. and Liu, H. (2018). A new perspective on robust M-estimation: Finite sample theory and applications to dependence-adjusted multiple testing. *Annals of Statistics*, 46, 1904–1931.