Robust Statistical Procedures for Location and Scale Dynamic Models with Applications to Risk Management

Loriano Mancini

Submitted for the degree of Ph.D. in Economics

University of Lugano, Switzerland

Accepted on the recommendation of

Prof. G. Barone-Adesi, University of Lugano

Prof. F. Peracchi, Tor Vergata University

Prof. E. Ronchetti, advisor, University of Geneva

and University of Lugano

Prof. F. Trojani, advisor, University of Lugano

March 2004

A Claudia.

Acknowledgments

First of all I wish to thank my advisors Professor Fabio Trojani and Professor Elvezio Ronchetti for their guidance, encouragement and support during my Ph.D. studies. To them I owe many intellectual debts. I wish also to thank Professor Fabio Trojani for the long time spent with me to discuss problems and ideas.

It is a pleasure to thank Professor Giovanni Barone-Adesi for many illuminating discussions on several financial topics and for accepting to be a member of the thesis committee.

I wish to thank Professor Franco Peracchi for being co-examiner of this work.

I am indebted to Claudio Ortelli for his help in introducing me to the field of robust statistics.

Thanks go to my friends Francesca Bellini and Ettore Croci for the beautiful time spent in Lugano. Thanks go also to my sister Tiziana and my old friends from the 'Università degli Studi di Perugia' for their unconditional and folk support.

Especially, I wish to thank Claudia for her love and support.

Contents

List of Tables

List of Figures

Introduction

The statistical properties of asset returns have been extensively investigated in the empirical financial literature; cf., for instance, Bollerslev, Chou and Kroner (1992), Foresi and Peracchi (1995), Pagan (1996), Cont (2001) and references therein. Some well-known empirical characteristics of speculative asset returns are time changing volatilities, volatility clustering, leverage effects and heavy tailed distributions[1]. Most of these stylized facts can be reproduced by nonlinear time series model for asset returns, like for instance models of the (G)ARCH family; cf. Engle (1982) and Bollerslev (1986). Therefore, such models are of crucial importance both for financial modelling and for empirical applications.

Several authors have applied nonlinear time series models to describe the dynamics of volatility in financial applications. In dynamic portfolio selection problems, dynamic volatility based strategies have been shown to outperform static asset allocation models which do not take into account volatility changes; cf. Ferson and Siegel (2001) and Fleming, Kirby and Ostdiek (2001, 2003). Similarly, Busse (1999) provided evidence that mutual fund managers tend to reduce their market exposure during periods of high expected volatilities suggesting that many active portfolio managers behave like volatility timers.

[1]These basic facts were already documented in the sixties by Mandelbrot (1963) and Fama (1965).

In the option pricing framework, the Black and Scholes (1973) and Merton (1973) option pricing model has been formulated under the assumption of i.i.d. normally distributed log-returns. In order to account for volatility smiles/smirks, term structure patterns and further stylized facts observed in option prices, the basic model has been extended to allow also for stochastic volatilities of asset returns. This introduces nonlinearities in asset prices and/or volatility dynamics; cf., for instance, Hull and White (1987), Scott (1987), Wiggins (1987), Stein and Stein (1991), Heston and Nandi (2000), Barone-Adesi, Engle and Mancini (2004).

In the risk management area, an accurate model for portfolio returns is necessary in order to obtain reliable risk measure estimates, such as the VaR or the expected shortfall of a portfolio return. Some broadly used VaR prediction methods in the practice assume i.i.d. asset returns or normally distributed innovations; cf., for instance, Riskmetrics (1995). These methods cannot account for changing volatilities or heavy tailed distributions of asset returns and tend therefore often to underestimate risk figures and 'required capitals' of trading portfolios. In such cases, financial institutions can suffer unexpected large losses precisely when the reserve amounts are inadequate; cf., for instance, Barone-Adesi, Bourgoin and Giannopoulos (1998), Barone-Adesi, Giannopoulos and Vosper (1999, 2000), McNeil and Frey (2000) and Pritsker (2001).

Asset returns are also characterized by several 'irregularities' due for instance to liquidity patterns or market crashes that can hardly be included/reflected in a time series model. Therefore, it can be often more realistic to expect that parametric time series models can describe only some 'structural' part of asset returns, which is not affected by such irregularities. This viewpoint can be relevant for many financial questions, as for instance in portfolio selection problems where model misspecification/uncertainty is explicitly included in agents' intertemporal preferences; cf., for instance, Trojani and Vanini (2002). From a more statistically oriented viewpoint, time series models can be affected by several forms of misspecification. A pragmatic and realistic approach in this respect consists in

assuming that the underlying data generating process can locally deviate from the structural or the stochastic part of some parametric (reference) model. A general theoretical framework to handle this last type of model misspecification is provided by the Theory of Robust Statistics, which aims to provide inference results for the structural part of the data generating process. At the same time, robust statistics attempts at quantifying the maximal bias of a statistic under local model misspecifications and at detecting observations that are influential for a given problem[2]. The seminal ideas on robust inference traces back to Huber (1964), who introduced M-estimators as a generalization of Maximum Likelihood estimators. Then, Hampel (1968, 1974) introduced a fundamental tool in robust statistics, i.e. the influence function, to investigate the robustness and the asymptotic bias properties of a statistic. In the robust testing framework, Ronchetti (1982) and Heritier and Ronchetti (1994) formalized bounded-influence tests for general hypotheses. In the time series context, Künsch (1984) and Martin and Yohai (1986) introduced a time series influence function while Ronchetti and Trojani (2001) and Ortelli and Trojani (2004) derived robust GMM and robust EMM estimators and tests for a large class of models.

This thesis contributes to this research area by proposing and applying some feasible optimal robust estimators and tests for general nonlinear location and scale time series models. More specifically, we propose a general framework to handle local misspecifications of nonlinear second order parametric models for time series and we provide an application of such a methodology to risk management. We show how robust inference procedures can improve the statistical analysis of nonlinear time series models for asset returns in terms of parameter estimates and inference results, both on simulated and real data. Such statistical improvements should translate into financial and economic improvements. In our risk management application we show that robust procedures provide more accurate and stable

[2]See for some overviews on the Theory of Robust Statistics Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986).

VaR profiles over time. In such cases, financial institutions can adapt more smoothly and more efficiently risk exposures, thereby obtaining a comparative market advantage.

Outline. The thesis consists of two main chapters organized as independent articles. Chapter 1 proposes a new robust estimator for the parameters of a general conditional location and scale model. It also provides optimal robust versions of likelihood-type tests for such models. The efficiency and the robustness properties of the proposed statistics are studied in some Monte Carlo simulations and in an application to exchange rate data.

Chapter 2 presents an application of the new robust methodologies to risk management. We estimate a model of the (G)ARCH family for historical portfolio returns with the robust estimator proposed in Chapter 1. Then, resampling procedures are applied on standardized residuals to estimate VaR risk measures. Finally, the method is backtested on four stock price series and compared to competing approaches based on non robust volatility estimation procedures for the volatility process.

Chapter 1

Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models

1.1 Abstract

This paper studies the local robustness of estimation and testing procedures of the conditional location and scale parameters in a strictly stationary time series model. We first derive optimal bounded-influence estimators for the parameters of conditional location and scale models under a conditionally Gaussian reference model. Based on these results, optimal bounded-influence versions of the classical

likelihood-based tests for parametric hypotheses are then obtained. We propose a feasible and efficient algorithm for the computation of our robust estimators, which makes use of some analytical Laplace approximations to estimate the auxiliary recentering vectors ensuring Fisher consistency in robust estimation. This strongly reduces the necessary computation time by avoiding the simulation of multidimensional integrals, a task that has typically to be addressed in the robust estimation of nonlinear models for time series. In some Monte Carlo simulations of an AR(1)-ARCH(1) process we show that our robust estimators and tests maintain a very high efficiency under ideal model conditions and at the same time perform very satisfactorily under several forms of departures from a conditionally normal AR(1)-ARCH(1) process. On the contrary, classical Pseudo Maximum Likelihood inference procedures are found to be highly inefficient under such local model misspecifications. These patterns are confirmed by an application to robust testing for ARCH.

1.2 Introduction

This paper studies the local robustness properties of estimation and testing procedures for the conditional location and scale parameters of a strictly stationary time series model; see (1.1) below.

First, it characterizes the local robustness of inference procedures within such a model class by considering explicitly the features of the given time series setting. Second, it derives optimal robust estimation and inference procedures starting from the class of conditionally unbiased estimators for the parameters of the conditional location and scale equations. Finally, it offers easily computable numerical procedures for robust estimation in the given setting, which are only slightly more demanding than those required by a classical Pseudo Maximum Likelihood (PML, Gourieroux, Monfort and Trognon (1984)) estimation. This avoids the simulation of multidimensional integrals—a typical task in the robust estimation of nonlinear models for time series—thereby largely reducing the computa-

tion time.

The class of conditional location and scale time series models is quite broad and includes several well-known dynamic models largely applied in economics and empirical finance, such as pure conditional scale models (like ARCH and GARCH models; Engle (1982) and Bollerslev (1986)) or models that jointly parameterize the conditional location and the scale of the given time series (like for instance ARCH in mean models; Engle, Lilien and Robins (1987)). Typically, classical (non robust) estimation of the parameters of such models is obtained by means of a PML approach based on the *nominal* assumption of a conditionally Gaussian log-likelihood; see, for instance, Bollerslev and Wooldridge (1992).

Such PML estimators are based on an unbounded conditional score function, implying—as we show below—an unbounded times series influence function (IF, Künsch (1984) and Hampel (1974)) for the implied asymptotic functional estimator. As a consequence, PML estimators for conditional location and scale models are not robust with respect to local departures from conditional normality. In this paper we derive inference procedures for the parameters of conditional location and scale models which are robust to local nonparametric misspecifications of a parametric, conditionally Gaussian, location and scale reference model. More specifically, we consider the class of robust, conditionally unbiased, M-estimators for the parameters of conditional location and scale models[1] and derive the optimal (i.e. the most efficient) robust estimator within this class. Based on such estimators, several Maximum Likelihood (ML)-type bounded-influence tests for parametric hypotheses on the parameters of the conditional location and scale equations are then obtained following the approach proposed in Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001).

The need for robust procedures both in estimation and testing has been stressed by many authors

[1] As we show below, this class of estimators is convenient to develop robust inference procedures for the model setting considered in the paper.

and is now widely recognized both in the statistical and the econometric literature; cf., for instance, Hampel (1974), Koenker and Bassett (1978), Huber (1981), Koenker (1982), Hampel et al. (1986), Peracchi (1990), and more recently Markatou and Ronchetti (1997), Krishnakumar and Ronchetti (1997), Ronchetti and Trojani (2001), Ortelli and Trojani (2004). However, the problem of robust estimation for the parameters of conditional location and scale models has been considered so far by a few authors and only from the specific perspective of high breakdown estimation. Even less attention has been devoted to robust inference within conditional location and scale models. High breakdown estimators resistant to large amount of contamination in the sample data have been proposed by Sakata and White (1998) and Muler and Yohai (1999). These estimators are very useful at the exploratory stage and estimation stage. Here we focus at the inference stage, where we typically have an approximate model and we can expect small deviations from the model. Alternatives to high breakdown estimators are also needed because these estimators are computationally intensive and cannot be applied to estimate the parameters of a class of broadly applied models—such as threshold ARCH or ARCH in mean models[2].

This paper derives optimal bounded-influence estimation and testing procedures for a general conditional location and scale time series model, which are computationally only slightly more demanding than the ones required by a classical PML estimation of such models. The more specific contributions to the current literature are the following.

First, we characterize the robustness of conditionally unbiased M-estimators for nonlinear conditional location and scale models by computing the time series IF for the implied asymptotic functional

[2]More precisely, Muler and Yohai (1999) proposed a class of high breakdown estimators for pure ARCH processes (having zero conditional mean). Sakata and White (1998) proposed a class of high breakdown estimators for models where the parameter space can be partitioned according to the parameters arising in the conditional location and scale equations. For these estimators, they also computed formally the corresponding (time series) breakdown point.

estimator. This has been defined by Künsch (1984) who applied this concept to AR(p) processes. We extend this result to our general model (1.1); see below. This is a first necessary step which allows us to construct robust statistical procedures which can control for (i) the local asymptotic bias on the parameter estimates and (ii) the local asymptotic distortion on the level and the power of ML-type tests.

Second, we derive the optimal bounded-influence estimator for the parameters of conditional location and scale models under a conditionally Gaussian reference model. This extends the optimality result in Künsch (1984) (obtained for AR(p) models) and the application of optimal conditionally unbiased M-estimators in Künsch, Stefanski and Carroll (1989) (obtained for generalized linear models) to general nonlinear second order dynamic models. Based on these results, optimal bounded-influence versions of the classical Wald, score and likelihood ratio tests are derived along the general lines proposed in Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001).

Third, we propose a feasible algorithm for the computation of our optimal robust estimators, which can be easily implemented in standard packages, such as Matlab. This procedure is based on a truncating procedure which uses a set of Huber's weights to downweight the impact of influential observations. Fisher consistency at the model is preserved by means of some auxiliary recentering vectors, which in a time series setting have generally to be computed by simulations—as for instance in a Robust Generalized Method of Moments (RGMM, Ronchetti and Trojani (2001)) setting. Using the conditional unbiasedness of our estimator we provide analytical Laplace approximations for such vectors which strongly reduce the necessary computation time by avoiding the simulation of multidimensional integrals.

Fourth, we study by Monte Carlo simulation the efficiency and the robustness properties of our estimator. We estimate a simple AR(1)-ARCH(1) process under several models of local contamination of a conditionally Gaussian process. Under the Gaussian reference model the classical ML estimator

and our robust estimator have essentially the same efficiency. On the contrary, under local deviations from conditional normality classical PML estimators, tests and confidence intervals are found to be highly inefficient, while robust procedures perform very satisfactorily.

Finally, we present an application to robust testing for ARCH where robust procedures help to identify ARCH structures which could not be detected using the classical inference approach.

The structure of the paper is as follows. Section 1.3 introduces the conditional location-scale model and the classical PML estimation procedure. Section 1.4 computes the time series IF for conditionally unbiased M-estimators. The asymptotic bias on the parameter estimates induced by local deviations from the conditional Gaussian reference model is then approximated by means of the IF. In a second step, the optimal robust estimator is derived and the optimality of robust inference procedures based on such an estimator is discussed. The section is concluded by deriving an analytic approximation for the auxiliary recentering vectors in our robust estimation and by presenting the algorithm that can be used to compute our robust estimator in applications. Section 1.5 discusses robust testing procedures based on our robust estimators and gives some advice for applications. Section 1.6 presents the Monte Carlo experiments where the performance of our robust estimation and inference approach is evaluated in the setting of an AR(1)-ARCH(1) model. The empirical application to testing for ARCH is presented in Section 1.7. Section 1.8 summarizes and concludes.

1.3 The Standard Setting

Let $\mathcal{Y} := (y_t)_{t\in\mathbb{Z}}$ be a real valued strictly stationary random sequence on the probability space $(\mathbb{R}^\infty, \mathcal{F}, \mathbb{P}_*)$. Assume that $\mathbb{P}_*$ belongs to some parametric model $\mathcal{P} := \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$, i.e. $\mathbb{P}_* = \mathbb{P}_{\theta_0}$

for some[3] $\theta_0 \in \Theta$. Assume that y_t satisfies the second order regression model

$$\begin{aligned} y_t &= \mu_t(\theta_0) + \varepsilon_t(\theta_0), \\ \varepsilon_t^2(\theta_0) &= \sigma_t^2(\theta_0) + \nu_t(\theta_0), \end{aligned} \tag{1.1}$$

where $\mu_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ parameterize, respectively, the conditional mean and the conditional variance of y_t given the information $\mathcal{F}_{t-1}$ up to time[4] $t-1$. Therefore, under $\mathbb{P}_{\theta_0}$

$$E[\varepsilon_t(\theta_0)|\mathcal{F}_{t-1}] = E[\nu_t(\theta_0)|\mathcal{F}_{t-1}] = 0, \tag{1.2}$$

for all $t \in \mathbb{Z}$. Denote by $y_1^m := (y_1, \ldots, y_m)$ the finite random sequence of $\mathcal{Y}$ and by $\mathbb{P}_*^m$ the restriction of $\mathbb{P}_*$ on the σ-algebra generated by m process coordinates, i.e. the m-dimensional marginal distribution of y_1^m induced by $\mathbb{P}_*$. Similarly, define $y_{1+t}^{m+t} := (y_{1+t}, \ldots, y_{m+t})$ for all $t \in \mathbb{Z}$ and denote by $y_{-\infty}^1 := (\ldots, y_{-1}, y_0, y_1)$ the infinite random sequence of $\mathcal{Y}$ up to time 1.

Our aim is to develop robust inference procedures for model (1.1) when the true distribution $\mathbb{P}_*$ belongs to some nonparametric neighborhood of a parametric reference model $\mathbb{P}_{\theta_0}$. Then model (1.1) is regarded as an "approximate" description of the true data generating process $\mathbb{P}_*$.

Model (1.1) covers a broad class of well-known parametric models for time series. Some examples are presented in the sequel.

Example 1.1 AR(1) models assume

$$\begin{aligned} \mu_t(\theta_0) &= \rho\, y_{t-1}, \quad |\rho| < 1, \\ \sigma_t^2(\theta_0) &= \sigma^2. \end{aligned} \tag{1.3}$$

Robust inference procedures for linear auto-regressive processes have been well studied in robust statistics. Künsch (1984) defined a time series influence function (IF) in this context and derived

[3]This correct specification assumption is relaxed in Section 1.4.

[4]This implies that the distribution of $(\varepsilon_t(\theta_0), \nu_t(\theta_0))|\mathcal{F}_{t-1}$ is also parameterized by θ_0.

an optimally bounded-influence estimator for the parameters of a general AR(p) model. Martin and Yohai (1986) provided bounded-influence estimators for AR and MA models and studied the asymptotic bias implied by additive outliers. Finally, Bustos and Yohai (1986) proposed robust estimators for the parameters of an ARMA model using robust estimators of the residuals autocovariances.

Example 1.2 AR(1)-ARCH(1) models (cf. Engle (1982)) assume[5]

$$\begin{aligned} \mu_t(\theta_0) &= \rho_0 + \rho_1\, y_{t-1}, \\ \sigma_t^2(\theta_0) &= \alpha_0 + \alpha_1 (y_{t-1} - \rho_0 - \rho_1\, y_{t-2})^2 \end{aligned} \tag{1.4}$$

for $\rho_0 \in \mathbb{R}$, $|\rho_1| < 1$, $\alpha_0 > 0$ and $0 \leq \alpha_1 < 1$. Bounded-influence estimators for such models are available in the class of robust GMM (RGMM) or robust EMM (REMM) estimators; cf. Ronchetti and Trojani (2001) and Ortelli and Trojani (2004), respectively. Sakata and White (1998) developed high breakdown estimators for conditional location and scale models that include (1.4) as a special case. Muler and Yohai (1999) considered the pure ARCH setting.

Example 1.3 Double threshold AR(1)-ARCH(1) models with volatility asymmetries (see for instance Li and Li (1996) and Glosten, Jagannathan and Runkle (1993)) assume

$$\begin{aligned} \mu_t(\theta_0) &= \beta_0 + (\beta_1 + \beta_2\, d_{1,t-1})\, y_{t-1}, \\ \sigma_t^2(\theta_0) &= \alpha_0 + \alpha_1\, d_{1,t-1} + (\alpha_2 + \gamma_0\, d_{2,t-1})(y_{t-1} - \beta_0 - (\beta_1 + \beta_2\, d_{1,t-2})\, y_{t-2})^2 \end{aligned} \tag{1.5}$$

with the dummy variable $d_{1,t-1} = 1$ if $\beta_0 + \beta_1\, y_{t-1} > 0$ and zero otherwise, $d_{2,t-1} = 1$ if $\varepsilon_{t-1}(\theta_0) < 0$ and zero otherwise[6]. To our knowledge, so far no robust estimators have been applied to estimate such

[5]This class is very popular in financial econometrics because it accounts for two important stylized facts of financial time series, namely, heavy tails and volatility clustering of asset returns; cf., for instance, Mandelbrot (1963) and Fama (1965).

[6]These models are broadly used in financial applications because they account for the leverage effect, i.e. the stronger impact on volatility of bad news ($\varepsilon_{t-1}(\theta_0) < 0$) than good news ($\varepsilon_{t-1}(\theta_0) \geq 0$) when $\gamma_0 > 0$.

models. In principle robust EMM procedures could be applied to this model class. However, they would be highly computationally intensive. By contrast, the high breakdown estimators of Sakata and White (1998) cannot be applied directly to threshold ARCH models of the above form because they assume that the parameter space Θ can be partitioned as $\Theta = \Theta_1 \times \Theta_2$ in order to imply $\mu_t(\theta_0) = \mu_t(\theta_1)$, for $\theta_1 \in \Theta_1$ and all $t \in \mathbb{Z}$.

Example 1.4 GARCH(1,1) models (cf. Bollerslev (1986)) assume

$$\begin{aligned} \mu_t(\theta_0) &= 0, \\ \sigma_t^2(\theta_0) &= \alpha_0 + \alpha_1 y_{t-1}^2 + \delta\, \sigma_{t-1}^2(\theta_0) \\ &= \alpha_0/(1-\delta) + \alpha_1 \textstyle\sum_{j=0}^{\infty} \delta^j y_{t-1-j}^2, \end{aligned} \tag{1.6}$$

where α_0, α_1, $\delta > 0$ and $\alpha_1 + \delta < 1$. The last equality for $\sigma_t^2(\theta_0)$ highlights that a GARCH model is indeed an ARCH model with an infinite number of lagged y variables. Formally, high breakdown estimators for GARCH models are not available. Sakata and White (1998) conjecture that their estimators should have a relatively high breakdown point also for GARCH models under appropriate assumptions. On the other hand, robust EMM procedures can be applied to obtain bounded-influence estimators for these models.

Our goal is to derive efficient, locally robust estimators and testing procedures for the joint inference on the conditional mean and the conditional variance parameters of model (1.1)–(1.2), that do not require heavy computational methods.

In the sequel we first briefly review the classical Pseudo Maximum Likelihood estimation procedure of model (1.1)–(1.2). In a later section we will derive the optimal robust version (see Hampel (1974) and Stefanski, Carroll and Ruppert (1986)) of such estimators.

Under the assumption that $\mathbb{P}_* = \mathbb{P}_{\theta_0} \in \mathcal{P}$, a Maximum Likelihood (ML) estimator of θ_0 is available. This happens, for instance, when under $\mathbb{P}_{\theta_0}$ the random variable y_t has a conditionally Gaussian

distribution, $y_t|\mathcal{F}_{t-1} \sim \mathcal{N}(\mu_t(\theta_0), \sigma_t^2(\theta_0))$. For many applications, the assumption of normality is not satisfied. Generalizations of ML then only assume that the dynamic equations (1.1) are satisfied by y_t under $\mathbb{P}_*$, i.e. that the model is dynamically correctly specified. Let $\mathbb{P}_0$ denote some probability distribution on $(\mathbb{R}^\infty, \mathcal{F})$ which implies the dynamic model (1.1)–(1.2) for y_t. When $\mathbb{P}_* = \mathbb{P}_0$, ML estimators of θ_0 are no longer available as the distribution of $\varepsilon_t(\theta_0)$ is unknown. Therefore, the conditional location and scale model (1.1)–(1.2) is typically estimated by the Pseudo Maximum Likelihood (PML, Gourieroux, Monfort and Trognon (1984)) method based on a conditionally Gaussian score function; cf. also Bollerslev and Wooldridge (1992) for GARCH-type models.

Specifically, such a PML estimator (PMLE) $\hat{\theta}_n^{pml}$ is the solution of the maximization problem

$$\max_{\theta \in \Theta} \sum_{t=1}^{n} l(\tilde{y}_1^t; \theta),$$

where

$$l(\tilde{y}_1^t; \theta) := -\frac{1}{2} \ln \sigma_t^2(\theta) - \frac{1}{2} \frac{(\tilde{y}_t - \mu_t(\theta))^2}{\sigma_t^2(\theta)}$$

and $\tilde{y}_1^n := (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n)$ are sample observations of the process[7] $\mathcal{Y}$. Under $\mathbb{P}_0$ and by means of standard regularity conditions[8], $\bar{l}_n(\theta) := n^{-1} \sum_{t=1}^{n} l(\tilde{y}_1^t; \theta)$ converges almost surely, uniformly on Θ, to $\bar{l}(\theta) := E_0[l(y; \theta)]$, where $E_0[\cdot]$ is the expectation under $\mathbb{P}_0$ and either $y = y_1^m$ for some $m < \infty$ (as in Example 1.1, 1.2 and 1.3) or $y = y_{-\infty}^1$ (as in Example 1.4). Such a distinction between the cases $m < \infty$ and $m = \infty$ entirely depends on the specification of $\mu_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ in model (1.1).

As the Gaussian pseudo true density is in the class of quadratic exponential densities[9], $\hat{\theta}_n^{pml}$ is consistent under $\mathbb{P}_0$. Precisely, consistency is implied by the unique maximum of $\bar{l}(\theta)$ at θ_0. Then, the

[7] As usual, when $\mu_t(\theta)$ and/or $\sigma_t^2(\theta)$ depend on pre-sample values $y_0, y_{-1}, \ldots$, such values are replaced, for estimation purposes, by 'in-sample estimates'.

[8] See for instance Gourieroux, Monfort and Trognon (1984) Appendix 1, Section 1.

[9] See Gourieroux, Monfort and Trognon (1984), p. 691.

following asymptotic first order condition holds

$$E_0[s(y^1_{-\infty};\theta_0)] = 0, \tag{1.7}$$

where $s(y^1_{-\infty};\theta) := \partial l(y^1_{-\infty};\theta)/\partial\theta$ denotes the pseudo score function based on the Gaussian pseudo true density, i.e.

$$s(y^1_{-\infty};\theta) = \frac{1}{\sigma_1^2(\theta)}\frac{\partial\mu_1(\theta)}{\partial\theta}\varepsilon_1(\theta) + \frac{1}{2\sigma_1^2(\theta)}\frac{\partial\sigma_1^2(\theta)}{\partial\theta}\left(\frac{\varepsilon_1^2(\theta)}{\sigma_1^2(\theta)} - 1\right). \tag{1.8}$$

The asymptotic condition (1.7) is equivalent to equation (1.2) and (1.8). The PMLE of θ_0 then solves the finite sample estimating equations

$$n^{-1}\sum_{t=1}^{n} s(\tilde{y}_1^t;\hat{\theta}_n^{pml}) = 0, \tag{1.9}$$

which are the finite sample version of the asymptotic condition (1.7). Under the model $\mathbb{P}_0$, the usual Taylor expansion of (1.9) implies that $\sqrt{n}(\hat{\theta}_n^{pml} - \theta_0)$ converges in distribution to the Gaussian distribution $\mathcal{N}(0, V(s;\theta_0))$, where $V(s;\theta_0) := J(\theta_0)^{-1}\, I(\theta_0)\, J(\theta_0)^{-1}$ and

$$J(\theta_0) := E_0\left[-\frac{\partial s(y^1_{-\infty};\theta)}{\partial\theta^\top}\right]_{\theta=\theta_0}, \quad I(\theta_0) := E_0\left[s(y^1_{-\infty};\theta_0)\, s(y^1_{-\infty};\theta_0)^\top\right];$$

cf. again Bollerslev and Wooldridge (1992), p. 148.

When $\mathbb{P}_0 = \mathbb{P}_{\theta_0}$ the conditional distribution of $\varepsilon_t(\theta_0)$ is truly Gaussian, $y_t|\mathcal{F}_{t-1} \sim \mathcal{N}(\mu_t(\theta_0), \sigma_t^2(\theta_0))$ and the PMLE is indeed the MLE of θ_0. In this case, $I(\theta_0) = J(\theta_0)$ and $V(s;\theta_0)$ attains the Cramér-Rao lower bound $I(\theta_0)^{-1}$. The Gaussianity of y_t only affects the efficiency of the PMLE and not the functional form of the estimating equation (1.9). Hence, the bias in the estimate of θ_0 induced by contaminated distributions will have the same functional form for the PMLE and the MLE (and will be given by the influence function (1.18)).

It is well-known that PMLE's are in the class of M-estimators since they are defined by the roots of the implicit equation (1.9). This class is very convenient to develop robust estimators; cf. Huber (1981)

and Hampel et al. (1986). The next section proposes a class of robust M-estimators and an optimal robust version of the PMLE for time series models of the form (1.1)–(1.2).

1.4 Locally Robust Estimation

In this section we relax the assumption that the true underlying distribution $\mathbb{P}_*$ is given by the parametric distribution $\mathbb{P}_{\theta_0}$. Specifically, we allow $\mathbb{P}_*$ to be in a nonparametric neighborhood of the reference model specified by the parametric distribution $\mathbb{P}_{\theta_0}$ for $\mathcal{Y}$. For instance, the case $\mathbb{P}_* = \mathbb{P}_0$ falls in this setting, when the distance between the error distribution under $\mathbb{P}_0$ and $\mathbb{P}_{\theta_0}$ is "small", and the PMLE can be used for inference on θ_0. However, as we will show, the PMLE is not robust and if $\mathbb{P}_*$ is indeed "slightly" different from $\mathbb{P}_0$ the arising asymptotic bias can be very large. In general, we allow the model $\mathbb{P}_{\theta_0}$ to be dynamically misspecified with respect to $\mathbb{P}_*$.

Here, we are interested in robust M-estimators which are "resistant" to local deviations of $\mathbb{P}_*$ from the given reference model $\mathbb{P}_{\theta_0}$. Recall that under $\mathbb{P}_{\theta_0}$ one has $\varepsilon_t(\theta_0)|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2(\theta_0))$, where $\varepsilon_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ are defined by (1.1). Hence, the reference distribution of $\mathcal{Y}$ is indirectly specified by the uncorrelated, standard Gaussian, random sequence $(u_t(\theta_0))_{t\in\mathbb{Z}}$, where $u_t(\theta_0) := \varepsilon_t(\theta_0)\,\sigma_t(\theta_0)^{-1}$.

To analyze the local stability properties of M-estimators, we introduce the following functional notation for the asymptotic value of an estimator $a(\cdot)$ (cf., for instance, Martin and Yohai (1986), p. 786),

$$a : dom(a) \subset \mathcal{M}_{stat}^m \longrightarrow \Theta,$$

where $\mathcal{M}_{stat}^m := \{m\text{-dimensional marginals of strictly stationary processes}\}$ and $m \in \mathbb{N}^*$. Then, the M-functional $a(\cdot)$ is implicitly defined by an estimating function $\psi : \mathbb{R}^m \times \Theta \longrightarrow \mathbb{R}^p$ such that

$$E[\psi(y_1, \ldots, y_m; a(\mathbb{P}_*^m))] = 0, \tag{1.10}$$

where $\mathbb{P}_*^m$ is the restriction of $\mathbb{P}_*$ on m process coordinates. As already pointed out, the number of process coordinates m entirely depends on the parametric model (1.1) of interest. For instance, in Examples 1.1, 1.2 and 1.3 we have $m = 2, 3, 3$, respectively, while in Example 1.4 $m = \infty$. In robust statistics, only the finite dimensional case $m < \infty$ has been well studied in the setting of linear AR models; cf. Künsch (1984). For the infinite dimensional case $m = \infty$ Martin and Yohai (1986) proposed a specific influence function to study the asymptotic bias implied by a specific contamination, namely the replacement model, for the estimates of MA models.

In the rest of the paper, we consider the finite dimensional case $m < \infty$. Hence, the asymptotic value estimates of θ_0 is obtained by evaluating $a(\cdot)$ at the m-dimensional distribution $\mathbb{P}_*^m$. The reference model distribution is given by $\mathbb{P}_{\theta_0}^m$. As we focus on local robustness we consider local deviations of $\mathbb{P}_*^m$ from $\mathbb{P}_{\theta_0}^m$. Therefore, we assume that $\mathbb{P}_*^m$ is in the following nonparametric neighborhood $\mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$ of the reference model $\mathbb{P}_{\theta_0}^m$,

$$\mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m) := \{\mathbb{P}_\eta^m = (1-\eta)\mathbb{P}_{\theta_0}^m + \eta\mathbb{G}^m, \quad \eta \leq b,\ b \in [0,1],\ \mathbb{G}^m \in \mathcal{M}_{stat}^m\}. \tag{1.11}$$

The neighborhood defined in (1.11) is a simple way to formalize local perturbations[10] of the model $\mathbb{P}_{\theta_0}^m$.

Recall that the reference model $\mathbb{P}_{\theta_0}^m$ needs not to be the true model for $\mathcal{Y}$ but has rather to be interpreted as an "approximate" model for the true data generating process $\mathbb{P}_*^m$, where the notion of "approximate" model is formalized implicitly in terms of "small" distributional distances.

Remark 1.1 The true underlying model $\mathbb{P}_*^m$ is well approximated by some element of $\mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$ in all cases where we have a ϵ percentage of contaminated sample data as (cf. Künsch (1984), p. 846)

$$\mathbb{P}_*^m = (1 - \epsilon\, c(m))\, \mathbb{P}_{\theta_0}^m + \epsilon\, c(m)\, \mathbb{G}^m + o(\epsilon), \tag{1.12}$$

[10]Notice that $d_k(\mathbb{P}_\eta^m, \mathbb{P}_{\theta_0}^m) \leq \eta$ for all $\mathbb{G}^m \in \mathcal{M}_{stat}^m$, where $d_k(\cdot,\cdot)$ denotes the Kolmogorov distance.

where $c(m) := \partial p(m,\epsilon)/\partial\epsilon|_{\epsilon=0}$ and

$$p(m,\epsilon) = \mathbb{P}\,[\,\text{at least one outlier in a block of length } m\,].$$

However, since $p(m,0) = 0$ for all $m \in \mathbb{N}$ and $\lim_{m\to\infty} p(m,\epsilon) = 1$ for all $\epsilon > 0$, we have $\lim_{m\to\infty} c(m) = \infty$. Therefore, when $m = \infty$, (1.12) cannot be applied[11] to define a neighborhood for the reference distribution $\mathbb{P}_{\theta_0}$ of $\mathcal{Y}$. In general, any mixture distribution $\mathbb{P}_\epsilon = (1-\epsilon)\,\mathbb{P}_{\theta_0} + \epsilon\,\mathbb{G}$, $\epsilon \in [0,1]$, on an infinite dimensional state space does not correspond to any interesting distribution for time series models as it implies that any sample path of the series is generated either by $\mathbb{P}_{\theta_0}$ or by $\mathbb{G}$; cf. Martin and Yohai (1986), p. 791. One possible way to study the local robustness of a statistic in this setting is to focus on a specific contamination model. For instance, additive outlier models have been studied in Martin and Yohai (1986) in the context of (linear) MA models. In this case, the "neighborhood" of the reference model $\mathbb{P}_{\theta_0}$ is implicitly defined by all possible distributions induced by the assumed contamination models.

In the next section we introduce the influence function for time series to characterize the asymptotic bias and the asymptotic variance of an M-functional estimator $a(\cdot)$ defined by (1.10). This will motivate, in a second step, our optimal robust estimator for models of the form (1.1)–(1.2).

1.4.1 Time Series Influence Function

Robust procedures aim at the estimation of the parameter θ_0 in the reference model $\mathbb{P}^m_{\theta_0}$ when local deviations from such a reference model are allowed. Such deviations induce an asymptotic bias on the functional estimator $a(\cdot)$ defined by[12]

$$bias := a(\mathbb{P}^m_*) - a(\mathbb{P}^m_{\theta_0}) = a(\mathbb{P}^m_*) - \theta_0.$$

---

[11] This problem is related to the second 'open question' in Künsch (1984), p. 859.

[12] We assume that $a(\cdot)$ is Fisher consistent.

For the robust inference on θ_0, the standard robustness condition on the corresponding estimator is a bounded asymptotic bias. To describe the linearized asymptotic bias of $a(\cdot)$ induced by some model $\mathbb{P}_\eta^m \in \mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$, one can consider the first order von Mises (1947) expansion of $a(\cdot)$ at $\mathbb{P}_{\theta_0}^m$ (cf., for instance, Fernholz (1983)),

$$a(\mathbb{P}_\eta^m) - a(\mathbb{P}_{\theta_0}^m) = \eta\, a'(\theta_0, \mathbb{G}^m) + o(\|\mathbb{P}_\eta^m - \mathbb{P}_{\theta_0}^m\|), \tag{1.13}$$

where $a'(\theta_0, \mathbb{G}^m)$ is the Gâteaux derivative of $a(\cdot)$ in the direction $\mathbb{G}^m - \mathbb{P}_{\theta_0}^m$, i.e.

$$a'(\theta_0, \mathbb{G}^m) := \lim_{\eta \downarrow 0} \frac{a((1-\eta)\mathbb{P}_{\theta_0}^m + \eta\, \mathbb{G}^m) - a(\mathbb{P}_{\theta_0}^m)}{\eta},$$

provided the limit exists. By contrast with a simple i.i.d. setting, in a time series framework $a'(\theta_0, \mathbb{G}^m)$ is determined by a set of equivalent kernels. To characterize such kernels, we introduce the following concept from the theory of robust statistics; cf. Künsch (1984), p. 847.

Definition 1.1 The influence function of the functional estimator $a(\cdot)$ is an equivalent class of kernels $IF : \mathbb{R}^m \times \Theta \longrightarrow \mathbb{R}^p$ such that

$$a'(\theta, \mathbb{G}^m) = \int_{\mathbb{R}^m} IF(y;\theta)\, d\mathbb{G}^m(y), \quad \text{for all } \mathbb{G}^m \in \mathcal{M}_{stat}^m.$$

We use the short notation $IF(y;\theta)$ for $IF(y; a, \mathbb{P}_\theta)$ as in Künsch (1984). In the one dimensional case, $m = 1$, the set of kernels is a singleton. Hence, in this case the influence function (IF) is unique and can be directly computed by taking functional Gâteaux derivatives of $a(\cdot)$ in the direction $\delta_x - \mathbb{P}_{\theta_0}^1$, where δ_x is the Dirac mass at $x \in \mathbb{R}$; cf. Hampel (1974) and Hampel et al. (1986).

In the m-dimensional case, $m \geq 2$, the IF can no longer be computed just by taking derivatives of $a(\cdot)$ in some suitable singular directions, namely because two different extremal measures in $\mathcal{M}_{stat}^m$ are not mutually singular. Therefore, any differentiation in such directions gives a corresponding kernel[13]

[13]For instance, for the functional estimator $a(\cdot)$ of the AR(1) model in Example 1.1, two different kernels are obtained when differentiating in the directions $(\delta_{x_{(1)},y_{(2)}} + \delta_{y_{(1)},x_{(2)}})/2$ and $(\delta_{x_{(1)},y_{(2)}} + \delta_{y_{(1)},z_{(2)}} + \delta_{z_{(1)},x_{(2)}})/3$, respectively; cf. Künsch (1984), p. 847.

conforming to Definition 1.1; cf. Künsch (1984), p. 847. As a consequence, in this case the kernel is not unique. Moreover, any function

$$IF(y_1, \ldots, y_m; \theta) + g(y_1, \ldots, y_{m-1}; \theta) - g(y_2, \ldots, y_m; \theta), \tag{1.14}$$

where $g : \mathbb{R}^{m-1} \times \Theta \longrightarrow \mathbb{R}^p$ is an integrable function, is again a kernel satisfying Definition 1.1, because

$$E_{\mathbb{G}^m}[g(y_1, \ldots, y_{m-1}; \theta) - g(y_2, \ldots, y_m; \theta)] = 0,$$

by the strictly stationarity of $\mathbb{G}^m$.

Künsch (1984) introduced a natural additional condition on a kernel satisfying Definition 1.1, which determines a unique version of the IF up to additive constants. This condition simply requires that, at the reference model $\mathbb{P}^m_{\theta_0}$, $y_m | y_1, \ldots, y_{m-1}$ has no influence on the asymptotic bias of the estimator.

Definition 1.2 The conditional influence function, $IF^{cond}(y_1^m; \theta_0)$, of the functional estimator $a(\cdot)$ is a kernel satisfying Definition 1.1, such that

$$E_{\theta_0}[IF^{cond}(y_1, \ldots, y_m; \theta_0) | \mathcal{F}_{m-1}] = 0, \tag{1.15}$$

where $E_{\theta_0}[\cdot]$ denotes the expectation with respect to the reference distribution $\mathbb{P}^m_{\theta_0}$.

IF^{cond} is unique[14] (cf. Künsch (1984), Th. 1.3) and has some desirable properties. First, under the reference model $\mathbb{P}^m_{\theta_0}$, it implies the following simple expression for the asymptotic covariance matrix $V(\psi; \theta_0)$ of $a(\cdot)$

$$V(\psi; \theta_0) := E_{\theta_0}[IF^{cond}(y_1^m; \theta_0)\, IF^{cond}(y_1^m; \theta_0)^\top],$$

which in applications does not need to be estimated by a Newey and West (1987) covariance matrix estimator, because of the martingale difference property (1.15). Moreover, as shown in the sequel,

[14]Under condition (1.15) the g-function in (1.14) is unique up to an additive constant.

conditional IF for MLE's has a natural interpretation in terms of the information matrix and a straightforward derivation; cf. Künsch (1984), (1.25).

Since under the model $\mathbb{P}_{\theta_0}$ the process $(IF^{cond}(y_{1+t}^{m+t};\theta_0))_{t\in\mathbb{Z}}$ is a martingale difference sequence and the conditional mean-variance of y_t is correctly specified, it is natural to introduce the class of conditionally unbiased M-estimators for θ_0. These functional estimators $a(\cdot)$ are implicitly defined by some function $\psi : \mathbb{R}^m \times \Theta \longrightarrow \mathbb{R}^p$, such that the conditional moment conditions

$$E_{\theta_0}[\psi(y_1,\ldots,y_m;a(\mathbb{P}_{\theta_0}^m))|\mathcal{F}_{m-1}] = 0 \tag{1.16}$$

hold for a unique $\theta_0 \in \Theta$. Such estimators have several desirable properties that are presented below. By construction $(\psi(y_{1+t},\ldots,y_{m+t};\theta_0))_{t\in\mathbb{Z}}$ is also a martingale difference sequence under the model $\mathbb{P}_{\theta_0}$. Thus, by definition $a(\cdot)$ is conditionally Fisher consistent and the asymptotic estimating equation for θ_0 is

$$E_{\theta_0}[\psi(y_1,\ldots,y_m;a(\mathbb{P}_{\theta_0}^m))] = 0. \tag{1.17}$$

The Gaussian score function (1.8) is an example of a function ψ defining a conditionally unbiased estimator of θ_0.

Example 1.5 The PMLE for the AR(1)-ARCH(1) process in Example 1.2 is defined by the ψ function

$$\psi(y_1^3;\theta_0) = -k_{1,3} + k_{2,3}\frac{y_3-\rho_0-\rho_1 y_2}{\sqrt{\alpha_0+\alpha_1(y_2-\rho_0-\rho_1 y_1)^2}} + k_{1,3}\frac{(y_3-\rho_0-\rho_1 y_2)^2}{\alpha_0+\alpha_1(y_2-\rho_0-\rho_1 y_1)^2}$$

where

$$k_{1,3} := \frac{1}{2\sigma_3^2}\begin{pmatrix} -2\alpha_1(y_2-\rho_0-\rho_1 y_1) \\ -2\alpha_1(y_2-\rho_0-\rho_1 y_1)y_1 \\ 1 \\ (y_2-\rho_0-\rho_1 y_1)^2 \end{pmatrix}, \quad k_{2,3} := \frac{1}{\sigma_3}\begin{pmatrix} 1 \\ y_2 \\ 0 \\ 0 \end{pmatrix}.$$

More generally, the function ψ for the PMLE of model (1.1)–(1.2) has the functional form

$$\psi(y_1^m;\theta_0) = -k_{1,m} + k_{2,m}\, u_m(\theta_0) + k_{1,m}\, u_m^2(\theta_0),$$

where

$$k_{1,m} := \frac{1}{2\sigma_m^2(\theta_0)} \left.\frac{\partial \sigma_m^2(\theta)}{\partial \theta}\right|_{\theta=\theta_0}, \quad k_{2,m} := \frac{1}{\sigma_m(\theta_0)} \left.\frac{\partial \mu_m(\theta)}{\partial \theta}\right|_{\theta=\theta_0}.$$

Therefore, different specifications of $\mu_m(\theta_0)$ and $\sigma_m^2(\theta_0)$ are easily accommodated in the $\mathcal{F}_{m-1}$-measurable random vectors $k_{1,m}$ and $k_{2,m}$. Since $E_{\theta_0}[u_m(\theta_0)|\mathcal{F}_{m-1}] = 0$ and $E_{\theta_0}[u_m(\theta_0)^2|\mathcal{F}_{m-1}] = 1$, ψ defines a conditionally unbiased M-estimator.

The main property of a conditionally unbiased M-estimator defined by (1.16) is that the corresponding conditional IF is computed as in the one dimensional case by calculating the limit

$$IF^{cond}(x_1^m;\theta_0) := \lim_{\eta \downarrow 0} \frac{a((1-\eta)\mathbb{P}_{\theta_0}^m + \eta\, \delta_{x_{(1)},\ldots,x_{(m)}}) - a(\mathbb{P}_{\theta_0}^m)}{\eta},$$

where $\delta_{x_{(1)},\ldots,x_{(m)}}$ is the Dirac mass at $\{(y_1,\ldots,y_m) = (x_{(1)},\ldots,x_{(m)})\}$ and provided the limit exists. Precisely, by defining $\overline{\mathbb{P}}_\eta^m := (1-\eta)\,\mathbb{P}_{\theta_0}^m + \eta\,\delta_{x_{(1)},\ldots,x_{(m)}}$ and implicitly differentiating equation (1.17) in the direction $\delta_{x_{(1)},\ldots,x_{(m)}} - \mathbb{P}_{\theta_0}^m$ yields

$$\begin{aligned} 0 &= \left.\frac{\partial}{\partial \eta} E_{\overline{\mathbb{P}}_\eta^m}[\psi(y; a(\overline{\mathbb{P}}_\eta^m))]\right|_{\eta=0} \\ &= E_{\mathbb{P}_{\theta_0}^m}\left[\frac{\partial \psi(y; a(\mathbb{P}_\theta^m))^\top}{\partial a}\right]_{\theta=\theta_0} \left.\frac{\partial}{\partial \eta} a(\overline{\mathbb{P}}_\eta^m)\right|_{\eta=0} + \int_{\mathbb{R}^m} \psi(y; a(\mathbb{P}_{\theta_0}^m))\, \left.\frac{\partial}{\partial \eta} d\overline{\mathbb{P}}_\eta^m(y)\right|_{\eta=0} \end{aligned}$$

and, as $(\partial/\partial\eta)\overline{\mathbb{P}}_\eta^m\Big|_{\eta=0} = \delta_{x_{(1)},\ldots,x_{(m)}} - \mathbb{P}_{\theta_0}^m$, using (1.17)

$$IF^{cond}(x_1^m; a(\mathbb{P}_{\theta_0}^m)) = D(\psi; a(\mathbb{P}_{\theta_0}^m))^{-1}\, \psi(x_1^m; a(\mathbb{P}_{\theta_0}^m)), \tag{1.18}$$

where

$$D(\psi;\theta_0) := -E_{\mathbb{P}_{\theta_0}^m}\left[\frac{\partial \psi(y;\theta)^\top}{\partial \theta}\right]_{\theta=\theta_0}.$$

As the conditional IF is unique and defines a martingale difference process, equation (1.18) is the only admissible representation. When the dependence of the conditional IF on the corresponding score function ψ has to be emphasized we use in the sequel the notation IF_{ψ}^{cond}.

If the conditional IF exists then one can prove also for our more general model setting that it is unique (up to an additive constant) by using the same arguments as in the first part of the proof of Th. 1.3 in Künsch (1984). Existence of the conditional IF for the case of a bounded score function ψ (which is the relevant case for our robust estimator introduced in Section 1.4.2) can also be proved along the lines of Th. 1.3 in Künsch (1984).

If condition (1.16) does not hold, the conditional IF is not given by (1.18), as the following simple example shows.

Example 1.6 For an AR(1) process in Example 1.1, an unconditionally unbiased M-estimator for $\theta_0 = (\rho\ \sigma^2)^\top$ can be defined by the following ψ-function

$$\psi(y_1, y_2; \theta_0) = \begin{pmatrix} y_2 y_1 - \sigma^2 \rho/(1-\rho^2) \\ \\ y_2^2 - \sigma^2/(1-\rho^2) \end{pmatrix}.$$

However, the conditional IF of such an estimator is not given by (1.18) because ψ is not a martingale difference.

A bounded conditional IF ensures a bounded linearized asymptotic bias induced by any contaminated distribution $\mathbb{P}_\eta^m$ in the neighborhood $\mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$

$$bias := a(\mathbb{P}_\eta^m) - a(\mathbb{P}_{\theta_0}^m) = \eta \int_{\mathbb{R}^m} IF^{cond}(y;\theta_0)\ \frac{\partial}{\partial \eta} \mathbb{P}_\eta^m(dy)\Big|_{\eta=0} + o(\|\mathbb{P}_\eta^m - \mathbb{P}_{\theta_0}^m\|)$$

and the derivative on the right hand side is uniformly bounded for any $\mathbb{P}_\eta^m \in \mathcal{U}^\eta(\mathbb{P}_{\theta_0}^m)$ when $m < \infty$. Moreover, since the conditional IF is linearly related to the ψ-function of the asymptotic estimating equation (1.17), it is bounded if and only if the ψ-function is bounded. As the Gaussian score function (1.8) is unbounded (at least) in $\varepsilon_1(\theta)$, PMLE's based on such a score function are not robust.

For MLE's the conditional IF is

$$IF^{cond}(y_1^m;\theta_0) = I(\theta_0)^{-1} s(y_1^m;\theta_0),$$

where $I(\theta_0)$ is the information matrix. For instance, the conditional IF of the MLE in Example 1.1 is

$$IF^{cond}(x_1,x_2;\theta_0) = \begin{bmatrix} (1-\rho^2)^{-1} & 0 \\ 0 & 0.5\sigma^{-4} \end{bmatrix}^{-1} \begin{pmatrix} \sigma^{-2}(x_2-\rho x_1)x_1 \\ -0.5\sigma^{-2} + 0.5\sigma^{-4}(x_2-\rho x_1)^2 \end{pmatrix}.$$

For a contaminated distribution $\mathbb{P}_\eta^2 = (1-\eta)\mathbb{P}_{\theta_0}^2 + \eta\,\mathbb{G}^2$ we can then compute the implied linearized asymptotic bias as

$$a(\mathbb{P}_\eta^2) - a(\mathbb{P}_{\theta_0}) \approx \eta \begin{bmatrix} (1-\rho^2)^{-1} & 0 \\ 0 & 0.5\sigma^{-4} \end{bmatrix}^{-1} \begin{pmatrix} \sigma^{-2}(\zeta^{(1,2)} - \rho\zeta^{(2)}) \\ -0.5\sigma^{-2} + 0.5\sigma^{-4}\left((1+\rho^2)\zeta^{(2)} - 2\rho\zeta^{(1,2)}\right) \end{pmatrix},$$

where

$$\zeta^{(1,2)} := E_{\mathbb{G}^2}[y_1 y_2] \;\text{ and }\; \zeta^{(2)} := E_{\mathbb{G}^2}[y_1^2] = E_{\mathbb{G}^2}[y_2^2] \;\text{ because } \mathbb{G}^2 \in \mathcal{M}_{stat}^2.$$

Hence, the asymptotic bias can be arbitrarily large, depending on the values of $\zeta^{(1,2)}$ and $\zeta^{(2)}$ on the neighborhood $\mathcal{U}^\eta(\mathbb{P}_{\theta_0}^2)$. In Figure 1.1 we plot $\|V(s;\theta_0)^{-1/2}\, IF^{cond}(x_1,x_2;\theta_0)\|$, that is the normed self-standardized IF of the MLE for θ_0 under $\mathbb{P}_{\theta_0}^2$ (cf. also Section 1.4.2 below), as a function of x_1 and x_2 for the parameter choice $\rho = 0.8$ and $\sigma^2 = 2$. In the left and right region of the grid (where, respectively, x_1 is 'low', x_2 is 'high' and vice versa), the normed self-standardized IF is clearly unbounded denoting high sensitivity of the MLE to possible "jumps" in the data[15].

As already mentioned, a bounded conditional IF ensures a bounded asymptotic bias for any distribution in the neighborhood $\mathcal{U}(\mathbb{P}_{\theta_0}^m)$ of the reference model $\mathbb{P}_{\theta_0}^m$. In the next section we define such a robust estimator for the parameters of model (1.1).

---

[15] In this example the most robust estimator has a self-standardized sensitivity no larger than $\sqrt{2}$; cf. Hampel et al. (1986), p. 228.

1.4.2 Optimal Conditionally Unbiased Robust Estimators

In the following we derive the optimal bounded-influence estimator (see Proposition 1.1 and Corollary 1.1) for models of the form (1.1)–(1.2), under the conditionally Gaussian reference model $\mathbb{P}_{\theta_0}^m$, in the class of conditionally unbiased M-estimators for θ_0.

The most common approach to derive bounded-influence estimators is to impose a bound on the 'self-standardized sensitivity' γ of the estimator (cf., for instance, Krasker and Welsch (1982)) defined by

$$\gamma(\psi) := \sup_{z \in \mathbb{R}^m} \|V(\psi;\theta_0)^{-1/2} IF_{\psi}^{cond}(z;\theta_0)\|,$$

where $V(\psi;\theta_0) = E_{\theta_0}[IF_{\psi}^{cond}(z;\theta_0)\, IF_{\psi}^{cond}(z;\theta_0)^\top]$. This sensitivity measure has some desirable properties. Firstly, the bound on γ does not depend on the scaling of observations, a feature that improves the algorithm convergence. Secondly, the maximal bias for level and power of several ML-type tests can be controlled by imposing a bound on γ; cf. Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001). By definition, non robust estimators have $\gamma = \infty$ while bounded influence estimators have $\gamma \leq c < \infty$, for some positive constant $c \geq \sqrt{p}$; cf. Hampel et al. (1986), p. 228.

Optimality Results

Under a conditionally Gaussian reference model $\mathbb{P}_{\theta_0}^m$, the MLE for θ_0 is most efficient but not robust. Then, robustness can be ensured only by "paying" a small loss in efficiency at the reference model $\mathbb{P}_{\theta_0}^m$. Here we propose an estimator of θ_0 that achieves the same optimality result within the class of conditionally unbiased M-estimators as in Künsch, Stefanski and Carroll (1989).

Consider the functional estimator $\overline{a}(\cdot)$ of θ_0 implicitly defined by

$$E_{\theta_0}[\psi_c(y_1^m; \overline{a}(\mathbb{P}_{\theta_0}^m))] = 0, \tag{1.19}$$

where $\psi_c(y_1^m;\theta) := A(\theta)\,\psi^{bif}(y_1^m;\theta)$,

$$\psi^{bif}(y_1^m;\theta) := \big(s(y_1^m;\theta) - \tau(y_1^{m-1};\theta)\big)\, w(y_1^m;\theta), \tag{1.20}$$

and $w(y_1^m;\theta) := \min(1,\ c\,\|A(\theta)\,\big(s(y_1^m;\theta) - \tau(y_1^{m-1};\theta)\big)\,\|^{-1})$. The non singular matrix $A(\theta) \in \mathbb{R}^p \times \mathbb{R}^p$ and the $\mathcal{F}_{m-1}$-measurable random vectors $\tau(y_1^{m-1};\theta) \in \mathbb{R}^p$ are determined, respectively, by solving the implicit equations

$$E_{\theta_0}[\psi_c(y_1^m;\theta_0)\ \psi_c(y_1^m;\theta_0)^\top] = I, \tag{1.21}$$

$$E_{\theta_0}[\psi_c(y_1^m;\theta_0)|\mathcal{F}_{m-1}] = 0. \tag{1.22}$$

We discuss below the interpretation and the computation of the A matrix and the τ-vectors. The estimating function ψ_c (or the unscaled version ψ^{bif}) is conditionally unbiased at the reference model and is a truncated version of the ML score as, by construction, $\|\psi_c(y_1^m;\theta)\| \le c$. Moreover, as $(\psi_c(y_{1+t}^{m+t};\theta_0))_{t\in\mathbb{Z}}$ is a martingale difference sequence under $\mathbb{P}_{\theta_0}$, the conditional IF of the functional estimator $\overline{a}(\cdot)$ is given by (1.18),

$$IF_{\psi_c}^{cond}(y_1^m;\overline{a}(\mathbb{P}_{\theta_0}^m)) = D(\psi_c;\theta_0)^{-1}\,\psi_c(y_1^m;\overline{a}(\mathbb{P}_{\theta_0}^m)).$$

The estimating function ψ^{bif} satisfies the following optimality criterion.

Proposition 1.1 If for a given constant of $c \ge \sqrt{p}$, equations (1.21) and (1.22) have solutions $A(\theta_0)$ and $\tau(y_1^{m-1};\theta_0)$, respectively, then ψ^{bif} minimizes $tr(V(\psi;\theta_0)\,V(\psi^{bif};\theta_0)^{-1})$ among all ψ satisfying (1.16) and

$$\sup_{z\in\mathbb{R}^m} \big(IF_\psi^{cond}(z;\theta_0)^\top\, V(\psi^{bif};\theta_0)^{-1}\, IF_\psi^{cond}(z;\theta_0)\big)^{1/2} \le c. \tag{1.23}$$

Up to multiplication by a constant matrix, ψ^{bif} is unique almost surely.

Any score function ψ^{opt} such that $V(\psi;\theta_0) - V(\psi^{opt};\theta_0)$ is positive semi-definite for all ψ satisfying (1.16) is called strongly efficient and the following corollary holds; cf. also Stefanski et. al. (1986), p. 416.

Corollary 1.1 If ψ^{bif} exists and there exists an unbiased, strongly efficient score function ψ^{opt} satisfying $\gamma(\psi^{opt}) \leq c < \infty$, then the two estimators coincide.

Proofs are given in Appendix A.1. Under general conditions (see Clarke (1983, 1986) and Heritier and Ronchetti (1994)), the optimal robust estimator $\overline{a}(\cdot)$ is consistent and asymptotically normally distributed at the reference model $\mathbb{P}_{\theta_0}^m$ with an asymptotic covariance matrix given by

$$\begin{aligned} V(\psi_c;\theta_0) &= E_{\theta_0}\left[\frac{\partial}{\partial\theta}\psi_c(y_1^m;\theta_0)^\top\right]^{-1} E_{\theta_0}\left[\frac{\partial}{\partial\theta^\top}\psi_c(y_1^m;\theta_0)\right]^{-1} \\ &= D(\psi_c;\theta_0)^{-1}\, D(\psi_c;\theta_0)^{-\top}, \end{aligned}$$

where for brevity $B^{-\top} := (B^\top)^{-1}$ for any invertible matrix B.

Remark 1.2 The conditions given in the references above guarantee the Fréchet differentiability of the functional corresponding to the estimator $\overline{a}(\cdot)$ which in turn implies the asymptotic normality in a $\epsilon\, n^{-1/2}$-neighborhood of the reference model; see also Bednarski (1993). These conditions are satisfied for M-estimators with a score function ψ which is bounded, continuous, and a.e. differentiable. Notice in particular that the PMLE is defined by an unbounded score function and is not Fréchet differentiable.

Under the reference model $\mathbb{P}_{\theta_0}^m$, the simple expression for $V(\psi_c;\theta_0)$ is implied by condition (1.22) and the Newey and West (1987) covariance matrix estimator is not necessary. Although $(s(y_{1+t}^{m+t};\theta_0))_{t\in\mathbb{Z}}$ is a martingale difference sequence under the reference model $\mathbb{P}_{\theta_0}^m$, not any truncated version of the Gaussian score function s is a martingale difference sequence. Such a property has to be explicitly imposed as in condition (1.22).

Interpretation of A and τ

The A matrix is a scaling matrix ensuring that the upper bound c on the function ψ_c is also the upper bound on the normed self-standardized influence function. Indeed, under the scaling condition (1.21),

$V(\psi_c;\theta_0)^{-1} = D(\psi_c;\theta_0)^\top D(\psi_c;\theta_0)$ and

$$\begin{aligned}\|V(\psi_c;\theta_0)^{-1/2}\, IF^{cond}_{\psi_c}(y;\theta_0)\|^2 &:= IF^{cond}_{\psi_c}(y;\theta_0)^\top\, V(\psi_c;\theta_0)^{-1}\, IF^{cond}_{\psi_c}(y;\theta_0)\\ &= \psi_c(y;\theta_0)^\top D(\psi_c;\theta_0)^{-\top}\, V(\psi_c;\theta_0)^{-1}\, D(\psi_c;\theta_0)^{-1}\psi_c(y;\theta_0)\\ &= \|\psi_c(y;\theta_0)\|^2.\end{aligned}$$

Hence, the normed self-standardized IF of $\overline{a}(\cdot)$ is equal to the Euclidian norm of the robust score function ψ_c, which is bounded by c. The A matrix is computed by a simple iterative procedure given in Section 1.4.3, where we propose an algorithm to compute the optimal estimator (1.19)–(1.22).

Further, to satisfy the conditional Fisher consistency condition (1.22), each truncated score function has to be shifted by some corresponding τ-vector. This implicitly defines the random sequence of τ-vectors $(\tau(y_{1+t}^{m-1+t};\theta_0))_{t\in\mathbb{Z}}$ associated to $(\psi_c(y_{1+t}^{m+t};\theta_0))_{t\in\mathbb{Z}}$. The existence of such a sequence is guaranteed by the continuity of the mapping $\tau(y_1^{m-1};\theta) \longmapsto (\, s(y_1^m;\theta) - \tau(y_1^{m-1};\theta)\,)\, w(y_1^m;\theta)$ and by the mean value theorem; cf. also Lemma 2.1 in Künsch et al. (1989). As $\tau(y_1^{m-1};\theta_0)$ is $\mathcal{F}_{m-1}$-measurable,

$$\tau(y_1^{m-1};\theta_0) = \frac{E_{\theta_0}[s(y_1^m;\theta_0)\, w(y_1^m;\theta_0)|\mathcal{F}_{m-1}]}{E_{\theta_0}[w(y_1^m;\theta_0)|\mathcal{F}_{m-1}]}. \tag{1.24}$$

In general, the expectations involved in (1.24) cannot be expressed analytically and numerical procedures must be used for computations. However, this makes robust estimation procedures of models of the form (1.1) very time consuming. For the conditionally Gaussian reference model $\mathbb{P}^m_{\theta_0}$ we therefore provide an accurate analytical approximation of $\tau(y_1^{m-1};\theta_0)$ for all models of the form (1.1)–(1.2). Such approximations make use crucially of the conditionally unbiasedness of a robust score function ψ_c. In the unconditionally unbiased case, the centering τ-vector defining an unconditionally unbiased robust

M-estimator is implicitly defined by the condition $E_{\theta_0}[(s(y_1^m;\theta_0) - \tau(\theta_0))\, w(y_1^m;\theta_0)] = 0$, implying[16]

$$\tau(\theta_0) = \frac{E_{\theta_0}[s(y_1^m;\theta_0) w(y_1^m;\theta_0)]}{E_{\theta_0}[w(y_1^m;\theta_0)]}. \tag{1.25}$$

In general, the expectations in (1.25) cannot be expressed analytically, except in some very particular cases like AR(p) model settings where $\tau(\theta_0) = 0$, because the computation of τ requires computing some unconditional moments under $\mathbb{P}_{\theta_0}^m$. In virtually all cases relevant for this paper, such moments are unknown. Therefore, in these cases the functional dependence of τ on θ_0 and A in (1.25) must be computed by solving some m-dimensional integrals by Monte Carlo simulation[17]. These simulations make algorithms for robust estimators computationally very demanding.

Analytical Approximations for $\tau(y_1^m;\theta_0)$

Under the reference distribution $\mathbb{P}_{\theta_0}^m$ and for models of the form (1.1)–(1.2), we provide a simple and easy implementable accurate approximation for $\tau(y_1^{m-1};\theta)$ in equation (1.24). In the following we briefly explain the procedure; detailed calculations are given in Appendix A.2. We proceed in two steps.

In the first step, given $\tau^{(0)}$ as initial value for $\tau(y_1^{m-1};\theta_0)$, we compute the real roots of the following quartic equation, with respect to the real variable $u_m(\theta_0)$,

$$\begin{aligned} 0 &= \| A(\theta_0) \left(s(y_1^m;\theta_0) - \tau^{(0)} \right) \|^2 - c^2 \\ &:= \| A(\theta_0) \left(-k_{1,m} + k_{2,m}\, u_m(\theta_0) + k_{1,m}\, u_m^2(\theta_0) - \tau^{(0)} \right) \|^2 - c^2, \end{aligned} \tag{1.26}$$

where

$$k_{1,m} := \frac{1}{2\sigma_m^2(\theta_0)} \frac{\partial \sigma_m^2(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0}, \quad k_{2,m} := \frac{1}{\sigma_m(\theta_0)} \frac{\partial \mu_m(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0}.$$

[16] In the i.i.d. setting the class of conditionally unbiased M-estimators is equivalent to the unconditional ones and $\tau(y_1^{m-1};\theta_0) \equiv \tau(\theta_0)$ as $E_{\theta_0}[\cdot|\mathcal{F}_{m-1}] = E_{\theta_0}[\cdot]$.

[17] Since the y variables are not i.i.d., sufficiently long simulations are necessary in order to adequately capture the time dependence in the $\mathcal{Y}$ process, such as in the RGMM estimator case in Ronchetti and Trojani (2001).

Notice that $k_{1,m}$ and $k_{2,m}$ are $\mathcal{F}_{m-1}$-measurable. This allows us to solve equation (1.26) with respect to $u_m(\theta_0)$. In almost all simulations and all empirical estimations equation (1.26) had only two real roots. Therefore, we only consider that case for brevity. The case of four real roots is discussed in Appendix A.2.

In the second step, we 'split' the integrals in (1.24) according to the roots determined by (1.26). Precisely, denoting the roots by $\underline{u}_m$ and $\overline{u}_m$, with $\underline{u}_m \leq \overline{u}_m$, the denominator in (1.24) is given by

$$E_{\theta_0}[w(y_1^m;\theta_0)|\mathcal{F}_{m-1}] = \int_{-\infty}^{\underline{u}_m} \frac{c}{\|A(\theta_0)(s(\upsilon;\theta_0)-\tau^{(0)})\|}\, d\Phi(u) + [\Phi(\overline{u}_m) - \Phi(\underline{u}_m)] + \int_{\overline{u}_m}^{+\infty} \frac{c}{\|A(\theta_0)(s(\upsilon;\theta_0)-\tau^{(0)})\|}\, d\Phi(u), \tag{1.27}$$

where $\upsilon := (y_1, \ldots, y_{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0)\, u)$ and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard Gaussian density and cumulative function, respectively. Typical values of $\overline{u}_m$ range from 2.7 to 3.5 (the opposite for $\underline{u}_m$), so that both $\underline{u}_m$ and $\overline{u}^m$ are quite far in the tails of a standard normal distribution. Therefore, the 'main contribution' to the expectation on the left hand side of (1.27) comes from the term in the square brackets[18]. Since the integrals on the right hand side of (1.27) are 'symmetric', we can consider only the integral on the right tail. As mentioned, full analytical solutions are not available. However, since $\overline{u}_m$ is 'quite far' in the right tail of a standard Gaussian distribution, the integral can be well approximated using the Laplace's method; cf. for instance Jensen (1995), Th. 3.1.1. This gives

$$\begin{aligned}
&\int_{\overline{u}_m}^{+\infty} \frac{c}{\|A(\theta_0)(s(\upsilon;\theta_0)-\tau^{(0)})\|}\phi(u)\, du \\
&:= \int_{\overline{u}_m}^{+\infty} q_d(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\, du \\
&= \frac{1}{\sqrt{2\pi}}\exp(-0.5\overline{u}_m^2)\frac{1}{\overline{u}_m}\left(\overline{q}_d(0) + \frac{\overline{q}_d'(0)}{\overline{u}_m} + \frac{\overline{q}_d''(0)}{\overline{u}_m^2} + O\left(\frac{1}{\overline{u}_m^3}\right)\right) \\
&=: L_d(\overline{u}_m) + O\left(\frac{1}{\overline{u}_m^3}\right),
\end{aligned} \tag{1.28}$$

[18]The Gaussian density $\phi(u)$ is roughly zero for $u \notin (-4, 4)$ and $c\,\|A(\theta_0)(s(\upsilon;\theta_0)-\tau^{(0)})\|^{-1} \downarrow 0$ for $u \to \pm\infty$.

with the obvious notation for $q_d(u)$ and where $\overline{q}_d(z) := q_d(\overline{u}_m + z)\exp(-0.5z^2)$.[19]

The integral in the numerator of (1.24) is split in the same way as in (1.27). Then the integrals on the tails are again approximated using the Laplace method. Detailed calculations are presented in Appendix A.2. The resulting formula for the computation of τ is given in the next proposition.

Proposition 1.2 Given the second order regression model (1.1)–(1.2) and the conditionally Gaussian reference model $\mathbb{P}^m_{\theta_0}$, if the quartic equation (1.26) has only two real roots $\underline{u}_m \leq \overline{u}_m$, then

$$\tau(y_1^{m-1};\theta_0) = \frac{-L_n(\underline{u}_m) - k_{1,m}\left[\Phi(\overline{u}_m) - \Phi(\underline{u}_m)\right] + k_{2,m}M_{1,m} + k_{1,m}M_{2,m} + L_n(\overline{u}_m)}{-L_d(\underline{u}_m) + \left[\Phi(\overline{u}_m) - \Phi(\underline{u}_m)\right] + L_d(\overline{u}_m)} + O\big(\underline{u}_m^{-3}\big) + O\big(\overline{u}_m^{-3}\big),$$

where $M_{1,m} := \phi(\underline{u}_m) - \phi(\overline{u}_m)$, $M_{2,m} := \underline{u}_m\phi(\underline{u}_m) - \overline{u}_m\phi(\overline{u}_m) + \Phi(\overline{u}_m) - \Phi(\underline{u}_m)$. $L_n(\cdot)$ and $L_d(\cdot)$ are defined in Appendix A.2 and correspond to some Laplace's approximations for the integrals in the numerator and in the denominator of (1.24).

We recall that Proposition 1.2 is available because $\tau(y_1^{m-1};\theta)$, $k_{1,m}$ and $k_{2,m}$ are $\mathcal{F}_{m-1}$-measurable. This allows us to 'split' the integrals involved in (1.24) and to approximate the Gaussian integrals.

Remark 1.3 Intuitively, the real roots $\underline{u}_m$ and $\overline{u}_m$ in equation (1.26) determine the range where the standardized innovation $u_m(\theta_0)$ is 'not influential' (in terms of self-standardized sensitivity of $\overline{a}(\cdot)$) for the arising asymptotic bias because

$$\begin{aligned} \|A(\theta_0)\,(s(y_1^m;\theta_0) - \tau^{(0)})\| \quad &\leq \quad c, \iff u_m(\theta_0) \in [\underline{u}_m, \overline{u}_m], \\ &> \quad c, \iff u_m(\theta_0) \in (-\infty, \underline{u}_m) \cup (\overline{u}_m, +\infty), \end{aligned}$$

and the normed self-standardized IF of the functional estimator $\overline{a}(\cdot)$ is equal to the Euclidian norm of the ψ_c-function.

Example 1.7 For the AR(1)-ARCH(1) process of Example 1.2 and the parameters $\rho_0 = 0.01$, $\rho_1 = 0.8$, $\alpha_0 = 0.02$ and $\alpha_1 = 0.8$ Figure 1.2 shows $\|A(\theta_0)\,s(y_1^3;\theta_0)\|$, i.e. the normed self-standardized IF

[19] For the Laplace's approximation of the integral on the left tail replace $L_d(\overline{u}_m)$ by $-L_d(\underline{u}_m)$ in (1.28).

of the PMLE, as a function of $u_3(\theta_0)$ and for some values of y_1 and y_2. Clearly, $\|A(\theta_0)\, s(y_1^3;\theta_0)\|$ is unbounded and increases very rapidly for some values of y_1 and y_2. For the PMLE ($\tau^{(0)} = 0$ and $w(y_1^3;\theta_0) = 1$), the real roots in equation (1.26) are given by the intersections of $\|A(\theta_0)\, s(y_1^3;\theta_0)\|$ with a constant line at some given value of c. For instance, the most robust estimator has $c = 2$ and even in such a case equation (1.26) has two real roots for the two cases depicted in Figure 1.2. In our Monte Carlo simulations in Section 1.6 we consider a robust estimator of the AR(1)-ARCH(1) model with $c = 9$.

1.4.3 Algorithm

To compute the robust estimator defined in (1.19)–(1.22) an iterative algorithm has to be adopted because the weights $w(y_1^m;\overline{a}(\mathbb{P}_{\theta_0}^m))$, the matrix $A(\overline{a}(\mathbb{P}_{\theta_0}^m))$ and the random vectors $\tau(y_1^{m-1};\overline{a}(\mathbb{P}_{\theta_0}^m))$ depend on the value of the estimator itself in a nonlinear way. Given a constant $c \geq \sqrt{p}$ (cf. Hampel et al. (1986), p. 228), the robust estimator is computed by the following four steps algorithm.

1. Fix a starting value $\theta^{(0)}$ for θ_0, and initial values $\tau_t^{(0)} := \tau(\tilde{y}_{t-m+1}^{t-1};\theta^{(0)}) = 0$, for all $t = 1,\ldots,n$ and $A^{(0)}$ such that

$$A^{(0)^\top} A^{(0)} = \left[n^{-1} \sum_{t=1}^{n} s(\tilde{y}_{t-m+1}^{t};\theta^{(0)})\; s(\tilde{y}_{t-m+1}^{t};\theta^{(0)})^\top \right]^{-1}.$$

2. Compute, for all $t = 1,\ldots,n$, the real roots of equations (1.26), and the associated new values $\tau_t^{(1)} := \tau(\tilde{y}_{t-m+1}^{t-1};\theta^{(0)})$ for τ_t and the new matrix $A^{(1)}$ for A defined by

$$\tau_t^{(1)} := \frac{-L_n(\underline{u}_t) - k_{1,t}\,[\Phi(\overline{u}_t) - \Phi(\underline{u}_t)] + k_{2,t} M_{1,t} + k_{1,t} M_{2,t} + L_n(\overline{u}_t)}{-L_d(\underline{u}_t) + [\Phi(\overline{u}_t) - \Phi(\underline{u}_t)] + L_d(\overline{u}_t)},$$

$$(A^{(1)^\top} A^{(1)})^{-1} := n^{-1} \textstyle\sum_{t=1}^{n} \left(s(\tilde{y}_{t-m+1}^{t};\theta^{(0)}) - \tau_t^{(0)} \right) \left(s(\tilde{y}_{t-m+1}^{t};\theta^{(0)}) - \tau_t^{(0)} \right)^\top \times \min^2(1,\, c\,\|A^{(0)}\,(s(\tilde{y}_{t-m+1}^{t};\theta^{(0)}) - \tau^{(0)})\|^{-1}).$$

3. Compute the robust estimator $\theta^{(1)}$ implied by (1.19) for given $A^{(1)}$ and $\tau_t^{(1)}$ as the solution of the implicit equation

$$\sum_{t=1}^{n} \left(s(\tilde{y}_{t-m+1}^t; \theta^{(1)}) - \tau_t^{(1)} \right) \min(1,\ c \, \| A^{(1)} \, (s(\tilde{y}_{t-m+1}^t; \theta^{(0)}) - \tau_t^{(1)}) \|^{-1}) = 0.$$

4. Replace $A^{(0)}$ by $A^{(1)}$ and $\tau_t^{(0)}$ by $\tau_t^{(1)}$ for all $t = 1, \ldots, n$ and iterate Steps 2 and 3 above until convergence of the sequence $(\theta^{(i)})_{i \in \mathbb{N}}$ of estimators associated to (1.19) and to the sequence $(A^{(i)},\ \overline{\tau}^{(i)})_{i \in \mathbb{N}}$, where $\overline{\tau}^{(i)} := (\tau_1^{(i)}, \ldots, \tau_n^{(i)})$.

Starting values for $\theta^{(0)}$ could be the PML estimate of θ_0 or the result of a grid search algorithm. We wrote a Matlab code to implement the algorithm and we used the Matlab function '`roots`' to compute the real roots of equation (1.26). Analytical expressions for the τ-vectors avoid "internal" simulations to compute the robust estimator. This largely reduces the computation time. For comparison, we implemented a second algorithm in which the τ integrals were computed numerically using the Matlab function '`quadl`'. This algorithm is unfeasible as the computation time of $\overline{a}$ is almost two hours already for a simple AR(1)-ARCH(1) model. For further comparison, we also implemented a robust GMM estimator as in Ronchetti and Trojani (2001) with moment conditions $A(\theta)(s(y_1^m; \theta) - \tau(\theta))\, w(y_1^m; \theta)$, where $\tau(\theta)$ is given by (1.25). In our simulations of Section 1.6, the computation time of the estimator (1.19)–(1.22) was about 20% the one of such a robust GMM estimator.

1.5 Robust Testing Procedures

The robust estimator defined in (1.19)–(1.22) allows us to derive the optimal robust version of several ML-type tests to control the maximal bias on the level and the power of the tests induced by local distributional misspecifications of a null or an alternative hypothesis. Precisely, the robust versions of the classical Wald, score and likelihood ratio tests based on the robust estimator in (1.19)–(1.22) can

be derived following the general approach proposed by Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001). Such robust tests satisfy the optimality criterion of maximizing the asymptotic power subject to a bound on the asymptotic bias of the level and the power test. In the following, we present the robust testing procedure and discuss in detail only the robust version of the Wald test. The score and likelihood ratio test can be similarly derived.

Consider a general null hypothesis of the form

$$g(\overline{a}(\mathbb{P}^m_{\theta_0})) = 0, \tag{1.29}$$

for a smooth function $g : \Theta \longrightarrow \mathbb{R}^r$ such that $(\partial/\partial\,\overline{a})\, g(\overline{a}(\mathbb{P}^m_\theta))^\top$ is of full column rank r for all $\theta \in \Theta$. The aim is to analyze the local stability properties of a ML-type test based on the robust estimator defined in Section 1.4.2. We consider test statistics $n\,Q$ that are quadratic forms of a functional U,

$$n\,Q(\mathbb{P}^m_n) := n\,U(\mathbb{P}^m_n)^\top U(\mathbb{P}^m_n), \quad n \in \mathbb{N}, \tag{1.30}$$

where $\mathbb{P}^m_n$ is the empirical m-dimensional distribution of the observations $\tilde{y}_1, \ldots, \tilde{y}_n$. The functional U associated with the Wald test is[20]

$$U^W(\mathbb{P}^m_n) := \left[\frac{\partial g(\theta)}{\partial \theta^\top}\, V(\psi_c;\theta)\, \frac{\partial g(\theta)^\top}{\partial \theta}\right]^{-1/2}_{\theta=\overline{a}(\mathbb{P}^m_n)} g(\overline{a}(\mathbb{P}^m_n)).$$

Under the reference model $\mathbb{P}^m_{\theta_0}$ and the null hypothesis (1.29), $n\,Q(\mathbb{P}^m_n)$ converges in distribution to a χ^2 distribution with r degrees of freedom. To apply the methodology in Heritier and Ronchetti (1994), we make the following assumption.

Assumption 1.1 Let a bounded-influence estimator $\overline{a}$ of $a(\mathbb{P}^m_{\theta_0})$ be given. Then,

$$\sqrt{n}\,\left(\overline{a}(\mathbb{P}^m_n) - \overline{a}(\mathbb{P}^m_{\eta(\epsilon,n)})\right) \to \mathcal{N}(0, V(\psi;\theta_0)),\ n \to \infty \tag{1.31}$$

in distribution, uniformly over the sequence $(\mathcal{U}^{\eta(\epsilon,n)}(\mathbb{P}^m_{\theta_0}))_{n\in\mathbb{N}}$ of $\eta(\epsilon,n)$-neighborhoods of $\mathbb{P}^m_{\theta_0}$ defined by (1.11) for $\eta := \eta(\epsilon,n) = \epsilon\, n^{-1/2}$ and $\mathbb{G} \in dom(\overline{a})$.

[20]For the functional U associated with the score and the likelihood ratio test see Heritier and Ronchetti (1994), p. 898.

Assumption 1.1 is implied by the Fréchet differentiability of the functional $\overline{a}(\cdot)$; cf. Remark 1.2 after Corollary 1.1. Under Assumption 1.1 the following proposition holds.

Proposition 1.3 Let $\overline{a}$ be the robust estimator defined by (1.19)–(1.22) and denote by α the level functional of the test based on the functional $Q(\cdot)$ in (1.30). Further let $(\mathbb{P}^m_{\eta(\epsilon,n)})_{n\in\mathbb{N}}$ be a sequence of $\eta(\epsilon,n)$-contaminations of the underlying null distribution $\mathbb{P}^m_{\theta_0}$, each of them belonging to the corresponding neighborhood $\mathcal{U}^{\eta(\epsilon,n)}(\mathbb{P}^m_{\theta_0})$ as defined in (1.11). Then,

$$\lim_{n\to\infty} \alpha(\mathbb{P}^m_{\eta(\epsilon,n)}) = \alpha_0 + \mu\,\epsilon^2 \left\| \int_{\mathbb{R}^m} IF(z; U(\mathbb{P}^m_{\theta_0}))\, d\mathbb{G}^m(z) \right\|^2 + o(\epsilon^2), \tag{1.32}$$

for all $\mathbb{G}^m \in dom(\alpha)$, where $U(\cdot)$ is the U-functional associated with the corresponding test,

$$\mu := -\left.\frac{\partial}{\partial\beta} H_r(q_{1-\alpha_0};\beta)\right|_{\beta=0},$$

$H_r(\cdot;\beta)$ is the cumulative distribution function of a noncentral $\chi^2(r;\beta)$ distribution with r degrees of freedom and noncentrality parameter $\beta \geq 0$, $q_{1-\alpha_0}$ is the $1-\alpha_0$ quantile of a $\chi^2(r;0)$ distribution and $\alpha_0 = \alpha(\mathbb{P}^m_{\theta_0})$ is the nominal level of the test. Moreover, the bias of $\alpha(\mathbb{P}^m_{\eta(\epsilon,n)})$ is uniformly bounded by the inequality

$$\lim_{n\to\infty} |\alpha(\mathbb{P}^m_{\eta(\epsilon,n)}) - \alpha_0| \leq \mu\,\epsilon^2 \sup_{z\in\mathbb{R}^m} \|V(\psi_c;\theta_0)^{-1/2}\, IF^{cond}_{\psi_c}(z;\overline{a})\|^2 + o(\epsilon^2).$$

As a consequence of Proposition 1.3, the maximal asymptotic bias for the level of the test based on $\overline{a}$ is bounded by

$$\lim_{n\to\infty} |\alpha(\mathbb{P}^m_{\eta(\epsilon,n)}) - \alpha_0| \leq \mu\,(\epsilon\, c)^2 + o(\epsilon^2). \tag{1.33}$$

The "power" counterpart of Proposition 1.3 can also be obtained. Hence, also the maximal asymptotic bias of the power induced by local contaminations of the alternative distribution can be controlled by imposing a bound on the self-standardized sensitivity of $\overline{a}$; cf. Ronchetti and Trojani (2001), Th. 2.

An important issue in the application of our robust inference procedures is the selection of an appropriate tuning constant c in (1.19). It appears from the optimality result in Section 1.4.2 and from (1.33) that such a constant controls the degree of robustness of our robust procedure. This constant can be chosen according to different criteria. For instance, if one takes a formal inference point of view, one can use (1.33) to set c as 'fraction' of the expected degree of contamination ϵ and the maximal allowed bias on the level of the test. A table providing guidance on how to select c according to such a formal inference criterion is given in Ronchetti and Trojani (2001), p. 54. Alternatively, a more informal way is to compute the robust estimates for different values of c by imposing a given percentage of downweighted observations.

1.6 Monte Carlo Simulations

In this section we compare by Monte Carlo simulations the performance of the classical PMLE and the robust estimator at the reference model and in the presence of model contaminations. We estimate the AR(1)-ARCH(1) model presented in Example 1.2. We simulate the following contaminated models "near" the reference model $\mathbb{P}_{\theta_0}$.

1. *Standard Gaussian innovations.* In this experiment, the innovation $u_t(\theta_0)$ has a standard Gaussian distribution, the PMLE is the MLE and we compare the efficiency of the robust and the MLE under the reference model $\mathbb{P}_{\theta_0}$.

2. *Replacement model* (cf. for instance Martin and Yohai (1986)). Under such a model the observed process $\mathcal{X} := (x_t)_{t\in\mathbb{Z}}$ is generated according to the following data generating process,

$$x_t = (1 - \vartheta_t^\eta)y_t + \vartheta_t^\eta \xi_t, \tag{1.34}$$

where the clean process $\mathcal{Y} := (y_t)_{t\in\mathbb{Z}}$ is generated by the reference model $\mathbb{P}_{\theta_0}$ and $(\vartheta_t^\eta)_{t\in\mathbb{Z}}$ is an

i.i.d. 0-1 random sequence independent of $\mathcal{Y}$ with the property $\mathbb{P}(\vartheta_t^\eta = 1) = \eta$. Hence, at a time $t \in \mathbb{Z}$, the clean observation y_t is replaced by ξ_t with probability η. In our simulations we set $\eta = 0.5\%$ and $\xi_t = 1.5$ for all t. Such a low probability of contaminations is motivated by some difficulties of the standard PMLE to converge when higher probabilities of contaminations occur (for e.g. $\eta = 1\%$). In this experiment the model (1.1)–(1.2) is dynamically "slightly" misspecified as the dynamic equations (1.1) are not satisfied. This experiment allows to compare the performances of the PML and the robust estimator when very few observations deviate from the assumed model.

3. *Innovative outlier model* (cf. for instance Bustos and Yohai (1986)). Under such a contamination the innovations are given by $u_t(\theta_0) = \check{u}_t(\theta_0)\,[(1-\epsilon) + \epsilon\,\varrho^2]^{-1/2}$, where $\check{u}_t(\theta_0)$ is distributed as the following mixture distribution

$$\check{u}_t(\theta_0) \sim (1-\epsilon)\,\mathcal{N}(0,1) + \epsilon\,\mathcal{N}(0,\varrho^2). \tag{1.35}$$

We set $\epsilon = 1\%$ and $\varrho = 3$. Contamination (1.35) describes situations where a given shock (or outlier) affects also future realizations of the process $\mathcal{Y}$. Furthermore, as $u_t(\theta_0) \sim i.i.d.(0,1)$, the dynamic equations in model (1.1)–(1.2) are satisfied and the model is dynamically correctly specified. Hence, this is a typical situation in which the PMLE is applied (and it is not the MLE) and there are no theoretical efficiency reasons to prefer one estimator to the other.

The simulation design covers a good range of local deviations from the reference model $\mathbb{P}_{\theta_0}$. The tail indices (cf. Gasko and Rosenberger (1983), p. 322) of the innovation u_t under the given distributions are 1 for the standard Gaussian distribution, approximately 1.08 for the replacement model (1.34) and 1.03 for the innovative outlier model (1.35). For comparison, a standard Student t_5 distribution has a tail index of 1.34. Therefore, all simulated distributions are very close and samples from different processes are virtually undistinguishable.

We simulate the AR(1)-ARCH(1) model (1.4) for the following parameter choice: $\rho_0 = 0.01$, $\rho_1 = 0.8$, $\alpha_0 = 0.02$, $\alpha_1 = 0.8$, under the different distributions for y_t presented above and for the sample size $n = 1{,}000$. The tuning constant for the robust estimator was set at $c = 9$. Such a rather large value implies that very few observations were downweighted[21].

Each model is simulated 5,000 times. For each simulation we compute the PML and the robust estimates for θ_0 and the corresponding covariance matrices. Then, for each parameter we compute the corresponding confidence interval at the 95% confidence level.

1.6.1 Point Estimation

Estimation results are presented in Tables 1.1–1.3. For each estimated parameter, the first row contains summary statistics for the PML estimates and the second row for the robust estimates. In Figures 1.3–1.5 we plot the estimated densities of the classical and robust estimator. Table 1.1 shows that the efficiency loss of the robust estimator at the reference model $\mathbb{P}_{\theta_0}$ is almost negligible. Specifically, the mean squared errors of all parameter estimates are very close. This is confirmed by Figure 1.3. We recall that in this experiment the PMLE is the MLE. Table 1.2 and Figure 1.4 show instead large mean squared errors of PML estimates. By contrast, robust estimates maintain low mean squared errors. It is somehow surprising that such inefficiencies in PML estimates are induced by contaminating (on average) only 0.5% of the sample observations. Finally, Table 1.3 and Figure 1.5 show that, in terms of mean squared error, both estimators correctly estimate the conditional mean parameters ρ_0 and ρ_1. However, the robust estimator always outperforms the PMLE, especially in the estimation of the conditional variance parameters α_0 and α_1. We recall that, under the innovative outlier model (1.35), the PMLE should "perform well" as model (1.1)–(1.2) is correctly specified.

[21] For instance, under the reference model $\mathbb{P}_{\theta_0}$ only 3 or 4 (out of 1,000) observations were slightly downweighted with weights of 0.8–0.9.

In a simulation study not reported here we compared the performances of the RGMM estimates introduced in Section 1.4.3 and our robust estimator under the reference model $\mathbb{P}_{\theta_0}$ and the replacement model (1.34) for $\eta = 5\%$, $\xi_t \sim \mathcal{N}(0,1)$ and $c = 4$ for both estimators. The two performances were quite close, up to the large differences in the computation time.

1.6.2 Interval estimation

Figures 1.6–1.8 show the boxplots of the estimated confidence interval lengths for the PML and the robust estimates. Actual confidence interval coverages are close to the nominal 95% in both cases. An exception is the confidence interval of the parameter α_0 which is 78% for PML and 92% for the robust version under the replacement model (1.34). Figure 1.6 shows that, *under the reference model* $\mathbb{P}_{\theta_0}$, the confidence intervals lengths for both techniques are almost identical. However, Figure 1.7 shows that, under the replacement model (1.34), the PML confidence intervals are much larger than the robust ones, denoting large inaccuracy on the inference results. Finally, Figure 1.8 shows that confidence intervals are tighter for robust estimates than for PML estimates, especially for the conditional variance parameters α_0 and α_1.

1.6.3 Hypothesis Testing

To analyze the performance, outside the contamination model, of the classical PML estimator and our robust estimator from the perspective of hypothesis testing we also simulated 1,000 sample paths of an AR(1)-ARCH(1) model for the parameter choice $\rho_0 = 0$ and $\rho_1 = 0$, 0.05, 0.10 and $\alpha_0 = 0.02$, $\alpha_1 = 0.8$ under scaled Student t_3 and scaled Student t_5 innovations, respectively. We do not necessarily believe that the innovations follow these distributions but we take t_3 and t_5 as examples of distributions which are very close to the normal model. Under scaled Student t innovations the model (1.1)–(1.2) is dynamically correctly specified and hence the PMLE should perform well. In our experiment we tested

the joint null hypothesis $\rho_0 = 0$ and $\rho_1 = 0$ by means of a corresponding classical and robust Wald statistic, respectively. The empirical rejection frequencies of a Wald test based on the classical PML estimator and a Wald test based on our robust estimator are calculated for a fixed nominal level of the test of 5%. The results are presented in Table 1.4. The estimated standard error of the empirical rejection frequency $\hat{p}$ (using the binomial distribution) is 0.7%, 1.4% and 1.5% for $\hat{p} = 5\%$, 30%, 60%, respectively. Table 1.4 shows that the robust Wald test performs very well across all models, while the classical test is oversized in finite sample and shows a lower power than the robust one.

The low power of classical tests under even slight departures from conditional normality suggests that robust tests could be useful in application to unmask some possible 'dynamics' in the data hidden by the presence of influential observations.

1.7 Empirical Application

We apply classical and robust Wald tests for ARCH to weekly exchange rate returns of the Swedish krona against the US dollar over the period November 29^{th}, 1993 until November 17^{th}, 2003. The data were downloaded from Datastream and consist of 522 observations. The first ten sample autocorrelations of squared and absolute returns are not significantly different from zero. Moreover, the Jarque-Bera test has a p-value of 0.47 not rejecting normality. Classical PML estimates for the parameters ρ_0, ρ_1, α_0 and α_1 of an AR(1)-ARCH(1) model (and Wald test p-values for the hypothesis that the corresponding parameter is zero) are 0.02 (0.73), -0.030 (0.53), 1.86 (0), 0.06 (0.22). The robust estimates under a tuning constant $c = 4$ are 0.01 (0.88), 0.014 (0.75), 1.64 (0), 0.47 (0). Therefore, as in typical financial return series, the conditional mean parameters are not significantly different from zero. Moreover, the PML estimate of the ARCH parameter α_1 is also not significant. Hence, the classical Wald test does not reject the homoscedasticity hypothesis. By contrast, the robust estimate

of this ARCH parameter is highly significant, showing that ARCH effects in the data are possibly obscured by some outlying observations detected by the robust weights presented in the bottom panel of Figure 1.9. These results are consistent with the low power of PML tests under non normal conditional returns in Section 1.6.3. Finally, it is interesting to notice that one would expect outliers to *enhance* the ARCH structure. Instead, because the estimation of the volatility by classical techniques is inflated, the potential ARCH structure is hidden by the presence of a few outlying observations.

1.8 Conclusions

We derived optimal bounded-influence estimators for the parameters of conditional location and scale models under a conditionally Gaussian reference model. Based on these results, we obtained optimal bounded-influence versions of the classical likelihood-based tests for parametric hypotheses. We proposed an efficient algorithm for the computation of our robust estimators, which strongly reduces the necessary computation time by avoiding the simulation of multidimensional integrals. Monte Carlo simulations show that our robust estimators maintain a very high efficiency under ideal model conditions and have good robustness properties under local departures from conditional normality, both in estimation and inference. On the contrary, classical PML estimators are highly inefficient even under small departures from conditional Gaussianity. An application to exchange rate data confirms these patterns.

true	mean	median	q_{25}	q_{75}	Stdv	$q_{75} - q_{25}$	MSE%
ρ_0	0.0100	0.0099	0.0066	0.0133	0.0051	0.0067	0.0026
0.01	0.0100	0.0100	0.0066	0.0134	0.0051	0.0067	0.0026
ρ_1	0.7983	0.7989	0.7896	0.8074	0.0140	0.0178	0.0199
0.8	0.7977	0.7985	0.7890	0.8070	0.0142	0.0181	0.0208
α_0	0.0200	0.0199	0.0189	0.0210	0.0015	0.0021	0.0002
0.02	0.0200	0.0200	0.0189	0.0211	0.0015	0.0021	0.0002
α_1	0.7976	0.7986	0.7461	0.8496	0.0758	0.1034	0.5756
0.8	0.8007	0.8016	0.7490	0.8525	0.0765	0.1035	0.5850

Table 1.1: Summary statistics for each estimated parameter $\hat{\theta}_0 := (\hat{\rho}_0 \quad \hat{\rho}_1 \quad \hat{\alpha}_0 \quad \hat{\alpha}_1)^\top$ for the MLE (first row) and the robust estimator (second row) based on 5,000 simulations of 1,000 observations of the AR(1)-ARCH(1) model in Example 1.2 under the reference model $\mathbb{P}_{\theta_0}$, i.e. a standard Gaussian distribution for innovations; cf. Figure 1.3.

true	mean	median	q_{25}	q_{75}	Stdv	$q_{75} - q_{25}$	MSE%
ρ_0	0.0166	0.0161	0.0102	0.0227	0.0094	0.0125	0.0132
0.01	0.0112	0.0111	0.0074	0.0150	0.0057	0.0076	0.0034
ρ_1	0.7930	0.7952	0.7787	0.8089	0.0280	0.0302	0.0832
0.8	0.7959	0.7965	0.7864	0.8063	0.0154	0.0200	0.0254
α_0	0.0298	0.0290	0.0246	0.0341	0.0070	0.0095	0.0145
0.02	0.0222	0.0218	0.0204	0.0236	0.0027	0.0032	0.0012
α_1	0.8037	0.8052	0.7069	0.9125	0.1337	0.2055	1.7880
0.8	0.8081	0.8064	0.7486	0.8704	0.0896	0.1218	0.8097

Table 1.2: Summary statistics for each estimated parameter $\hat{\theta}_0 := (\hat{\rho}_0 \quad \hat{\rho}_1 \quad \hat{\alpha}_0 \quad \hat{\alpha}_1)^\top$ for the PMLE (first row) and the robust estimator (second row) based on 5,000 simulations of 1,000 observations of the AR(1)-ARCH(1) model in Example 1.2 under the replacement model (1.34); cf. Figure 1.4.

true	mean	median	q_{25}	q_{75}	Stdv	$q_{75} - q_{25}$	MSE%
ρ_0	0.0100	0.0100	0.0065	0.0135	0.0055	0.0070	0.0030
0.01	0.0100	0.0100	0.0067	0.0132	0.0051	0.0064	0.0026
ρ_1	0.7982	0.7988	0.7882	0.8086	0.0159	0.0204	0.0256
0.8	0.7978	0.7983	0.7889	0.8074	0.0143	0.0185	0.0209
α_0	0.0199	0.0199	0.0189	0.0210	0.0016	0.0022	0.0003
0.02	0.0194	0.0194	0.0184	0.0203	0.0014	0.0019	0.0002
α_1	0.7989	0.7992	0.7405	0.8594	0.0895	0.1189	0.8006
0.8	0.7727	0.7748	0.7225	0.8221	0.0747	0.0995	0.6328

Table 1.3: Summary statistics for each estimated parameter $\hat{\theta}_0 := (\hat{\rho}_0 \;\; \hat{\rho}_1 \;\; \hat{\alpha}_0 \;\; \hat{\alpha}_1)^\top$ for the PMLE (first row) and the robust estimator (second row) based on 5,000 simulations of 1,000 observations of the AR(1)-ARCH(1) model in Example 1.2 under the innovative outlier model (1.35); cf. Figure 1.5.

	t_3		t_5	
ρ_1	PML	ROB	PML	ROB
0.00	0.08	0.05	0.07	0.05
0.05	0.17	0.24	0.22	0.26
0.10	0.46	0.65	0.62	0.74

Table 1.4: Each entry in the Table corresponds to the empirical rejection frequency of the joint hypothesis $\rho_0 = 0$ and $\rho_1 = 0$ obtained using 5% critical values for the χ^2 test and based on 1,000 simulations of 1,000 observations of the AR(1)-ARCH(1) model in Example 1.2 under scaled t_3 and scaled t_5 innovations, respectively.

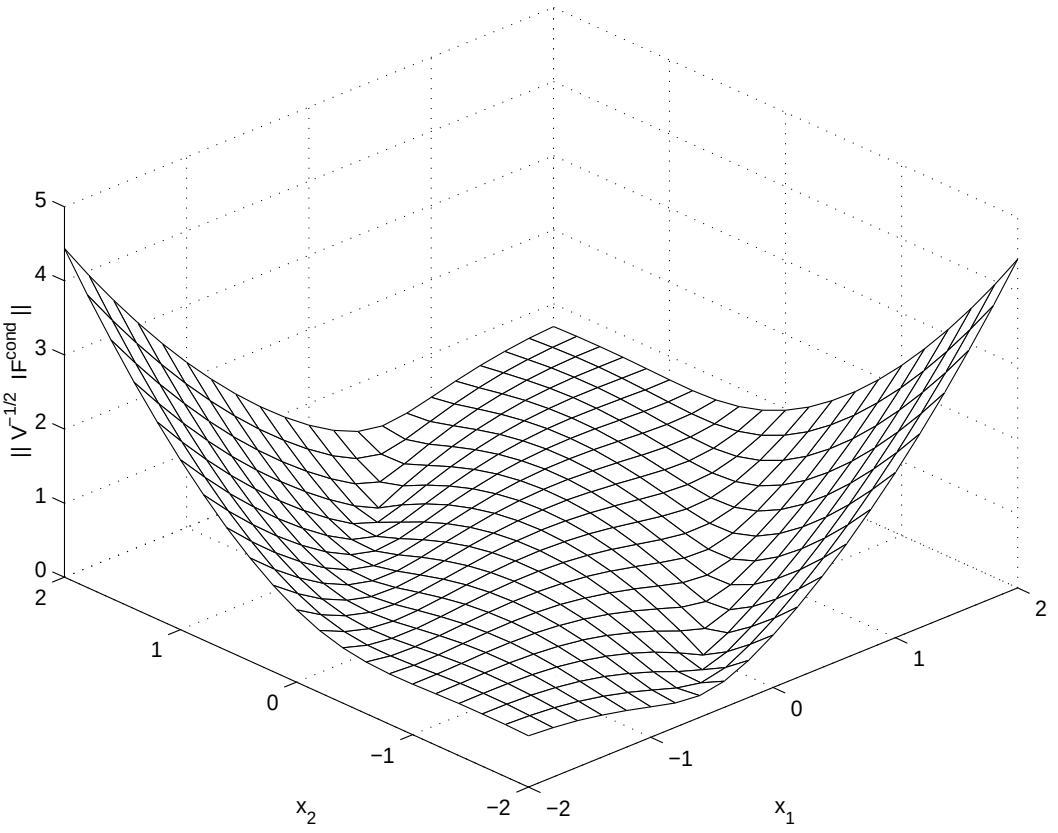

Figure 1.1: Normed self-standardized IF $\|V(s;\theta_0)^{-1/2} IF^{cond}(x_1, x_2; \theta_0)\|$ of the MLE, under the reference model $\mathbb{P}^2_{\theta_0}$, for the AR(1) model: $y_2 = 0.8\, y_1 + \varepsilon_2$, where $\varepsilon_2 \sim \mathcal{N}(0, 2)$, as a function of x_1 and x_2.

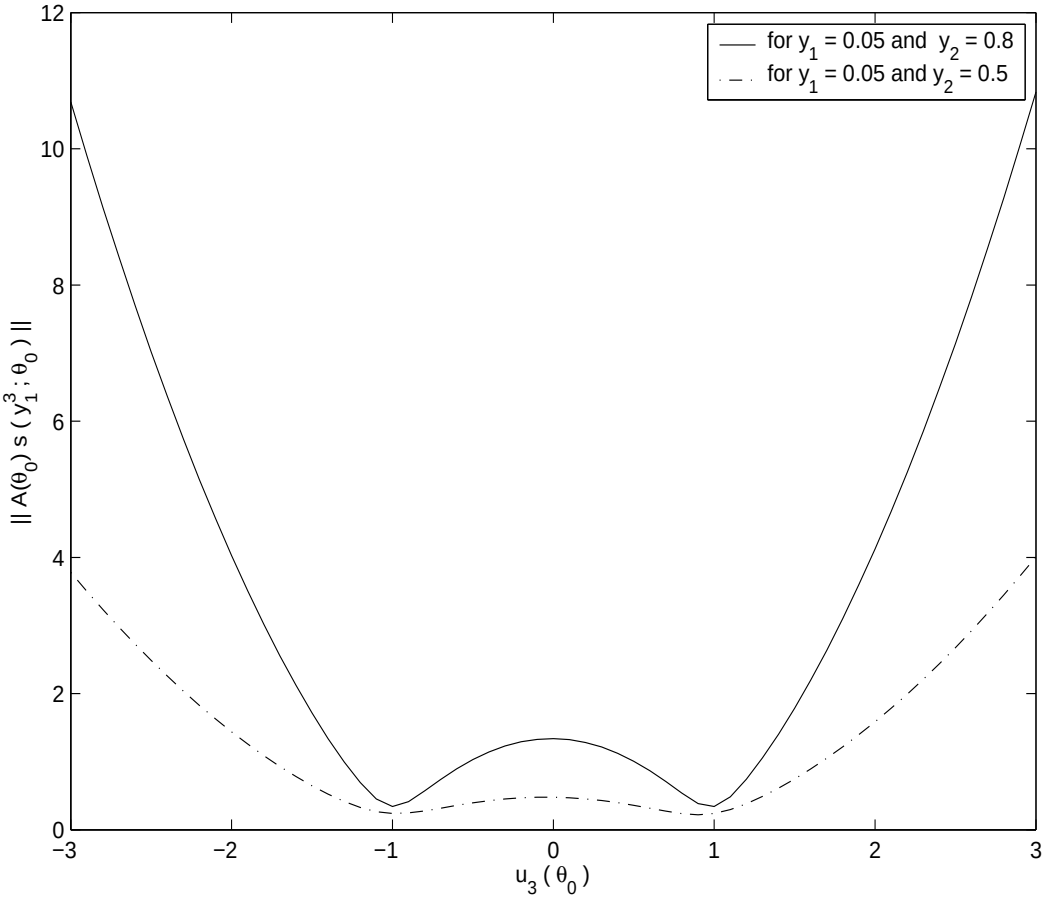

Figure 1.2: Normed self-standardized IF $\|A(\theta_0)\, s(y_1^3;\theta_0)\|$ of the MLE, under the reference model $\mathbb{P}^3_{\theta_0}$, for the AR(1)-ARCH(1) model: $y_3 = 0.01 + 0.8\, y_2 + \varepsilon_3(\theta_0)$, $\varepsilon_3(\theta_0) = \sigma_3(\theta_0)\, u_3(\theta_0)$, $\sigma_3^2(\theta_0) = 0.02 + 0.8\, \varepsilon_2^2(\theta_0)$, as a function of $u_3(\theta_0)$.

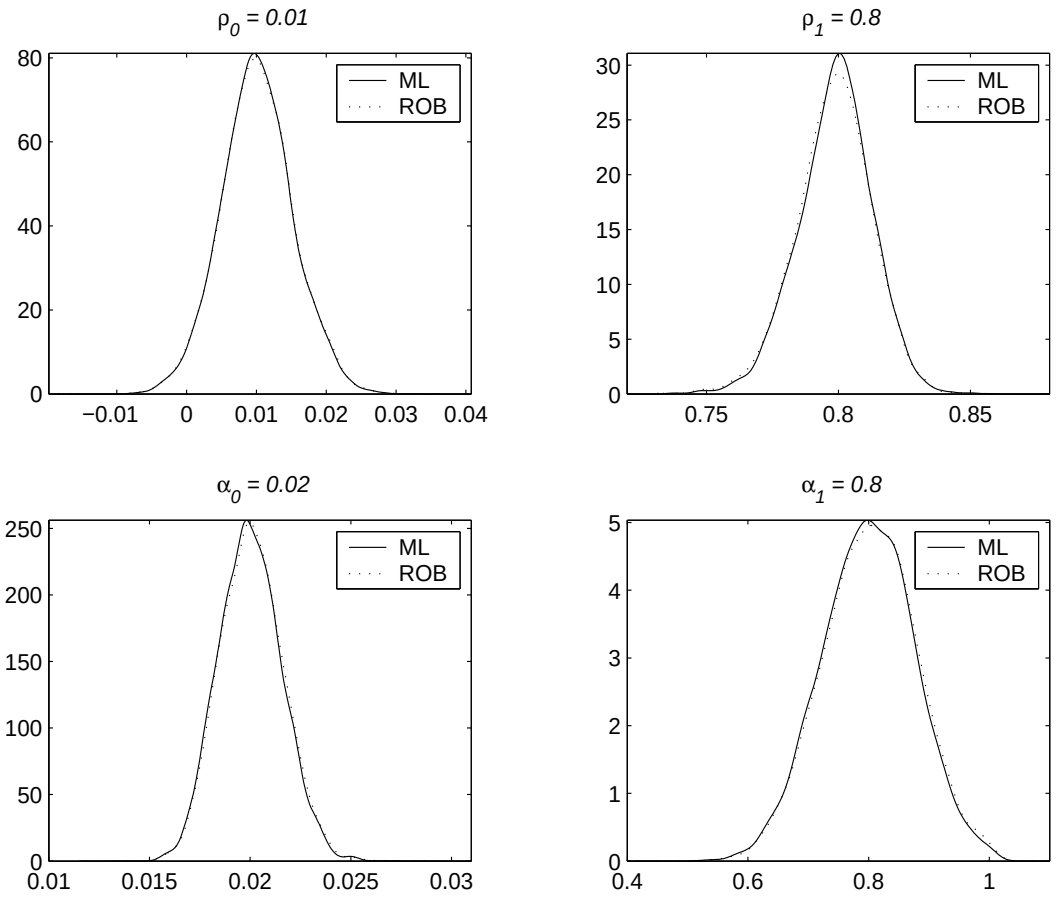

Figure 1.3: Estimated densities of $\hat{\theta}_0 := (\hat{\rho}_0 \;\; \hat{\rho}_1 \;\; \hat{\alpha}_0 \;\; \hat{\alpha}_1)^\top$ based on 5,000 simulations of 1,000 observations of the AR(1)-ARCH(1) process in Example 1.2 under the reference model $\mathbb{P}_{\theta_0}$, i.e. Gaussian distribution for innovations; cf. Table 1.1.

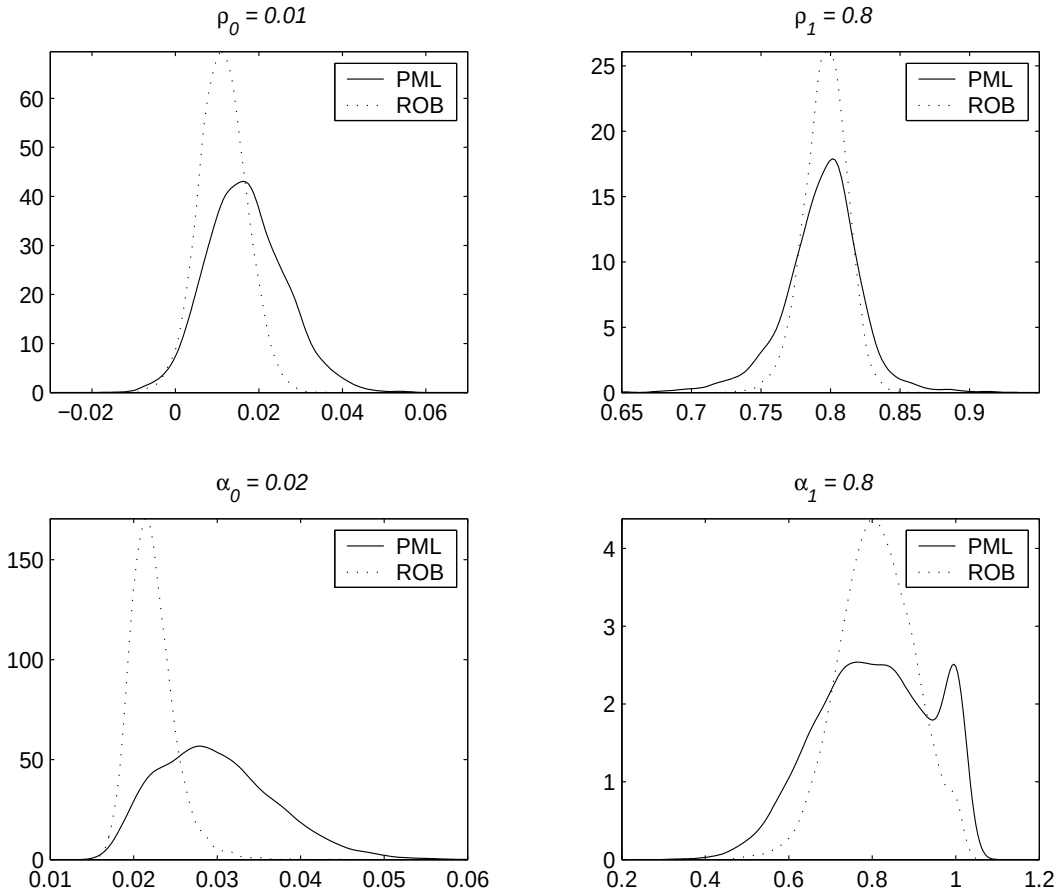

Figure 1.4: Estimated densities of $\hat{\theta}_0 := (\hat{\rho}_0 \quad \hat{\rho}_1 \quad \hat{\alpha}_0 \quad \hat{\alpha}_1)^\top$ based on 5,000 simulations of 1,000 observations of the AR(1)-ARCH(1) process in Example 1.2 under the replacement model (1.34); cf. Table 1.2.

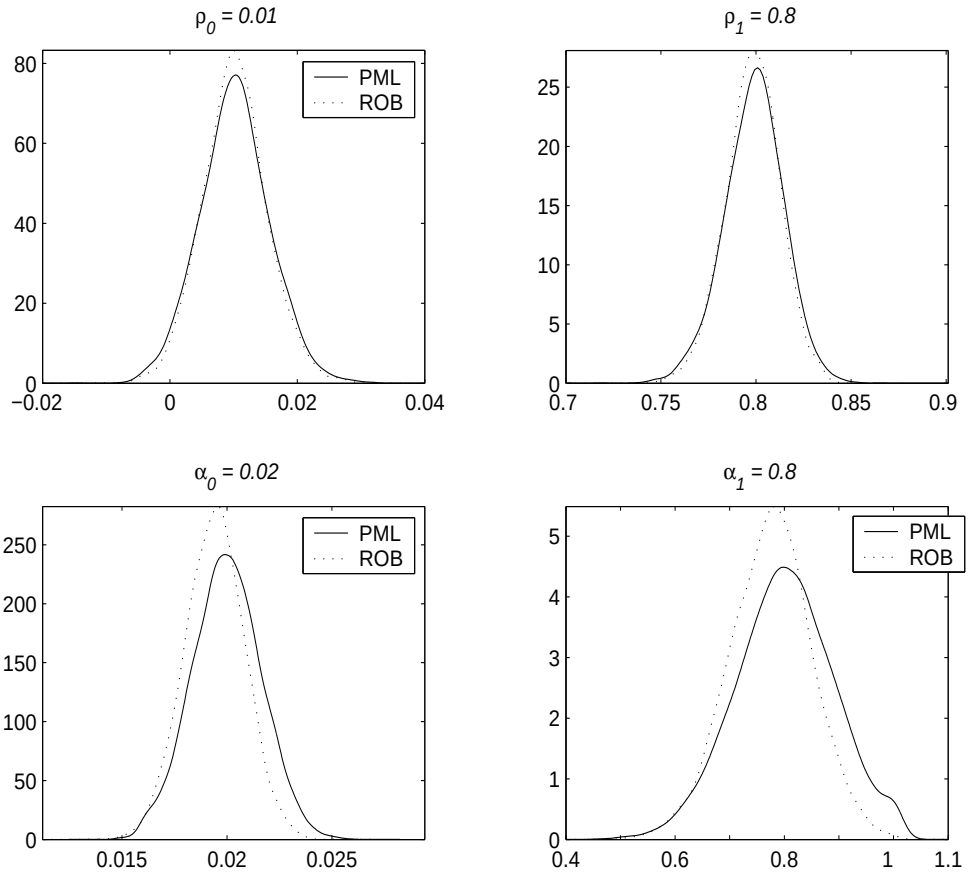

Figure 1.5: Estimated densities of $\hat{\theta}_0 := (\hat{\rho}_0 \quad \hat{\rho}_1 \quad \hat{\alpha}_0 \quad \hat{\alpha}_1)^\top$ based on 5,000 simulations of 1,000 observations of the AR(1)-ARCH(1) process in Example 1.2 under the innovative outlier model (1.35); cf. Table 1.3.

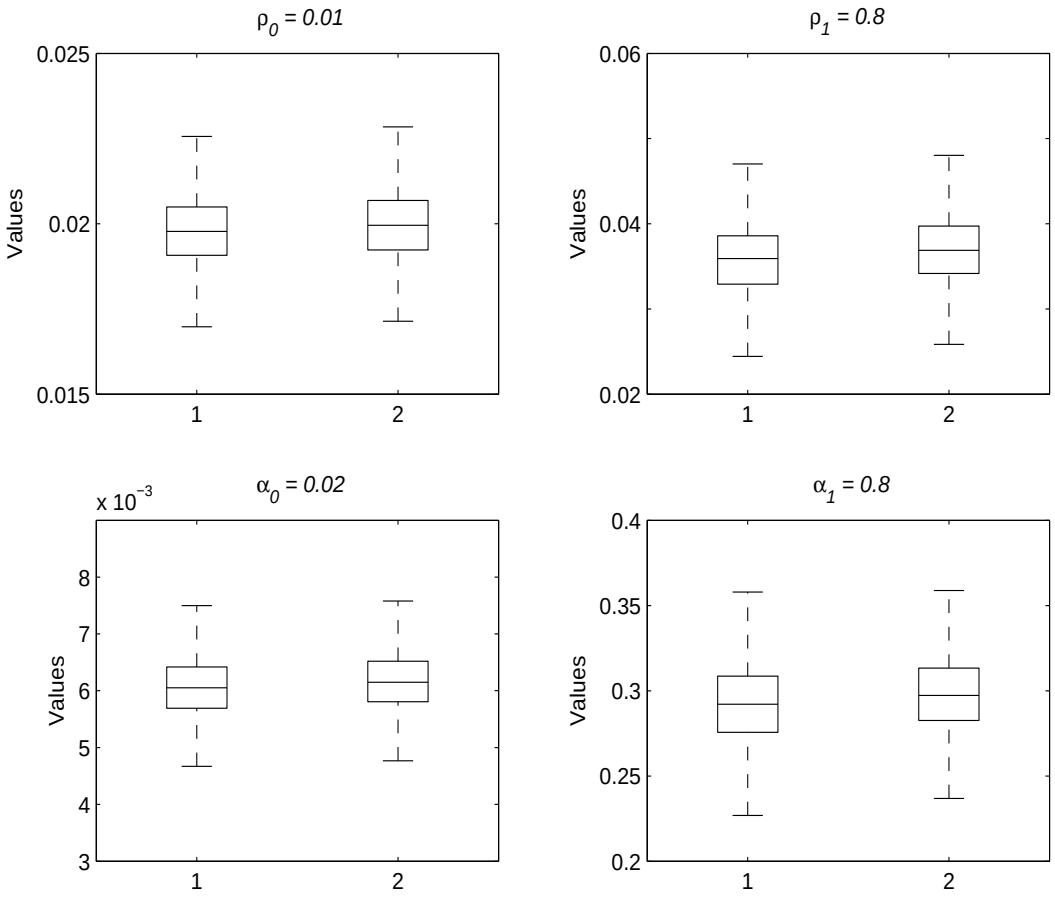

Figure 1.6: Boxplot of the lengths of ML (column 1) and robust (column 2) confidence intervals for $\hat{\theta}_0 := (\hat{\rho}_0 \;\; \hat{\rho}_1 \;\; \hat{\alpha}_0 \;\; \hat{\alpha}_1)^\top$ (cf. Figure 1.3) under the reference model $\mathbb{P}_{\theta_0}$.

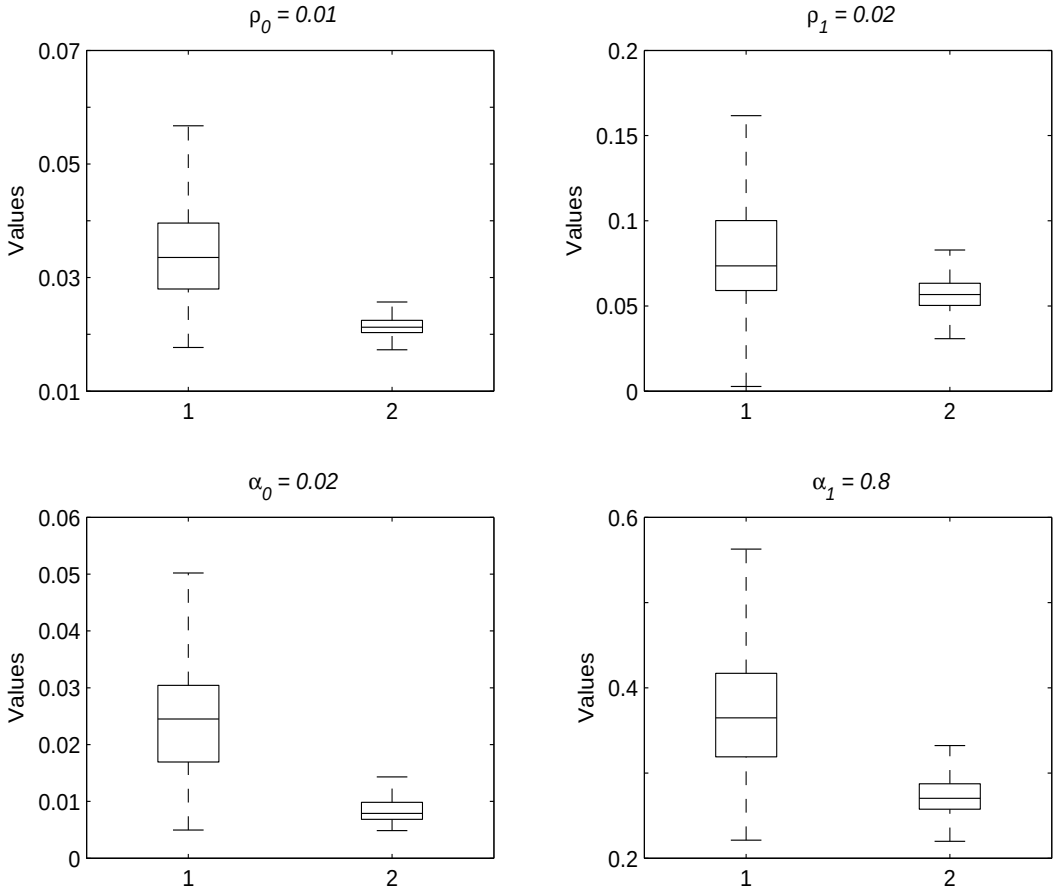

Figure 1.7: Boxplot of the lengths of PML (column 1) and robust (column 2) confidence intervals for $\hat{\theta}_0 := (\hat{\rho}_0 \;\; \hat{\rho}_1 \;\; \hat{\alpha}_0 \;\; \hat{\alpha}_1)^\top$ (cf. Figure 1.4) under the replacement model (1.34).

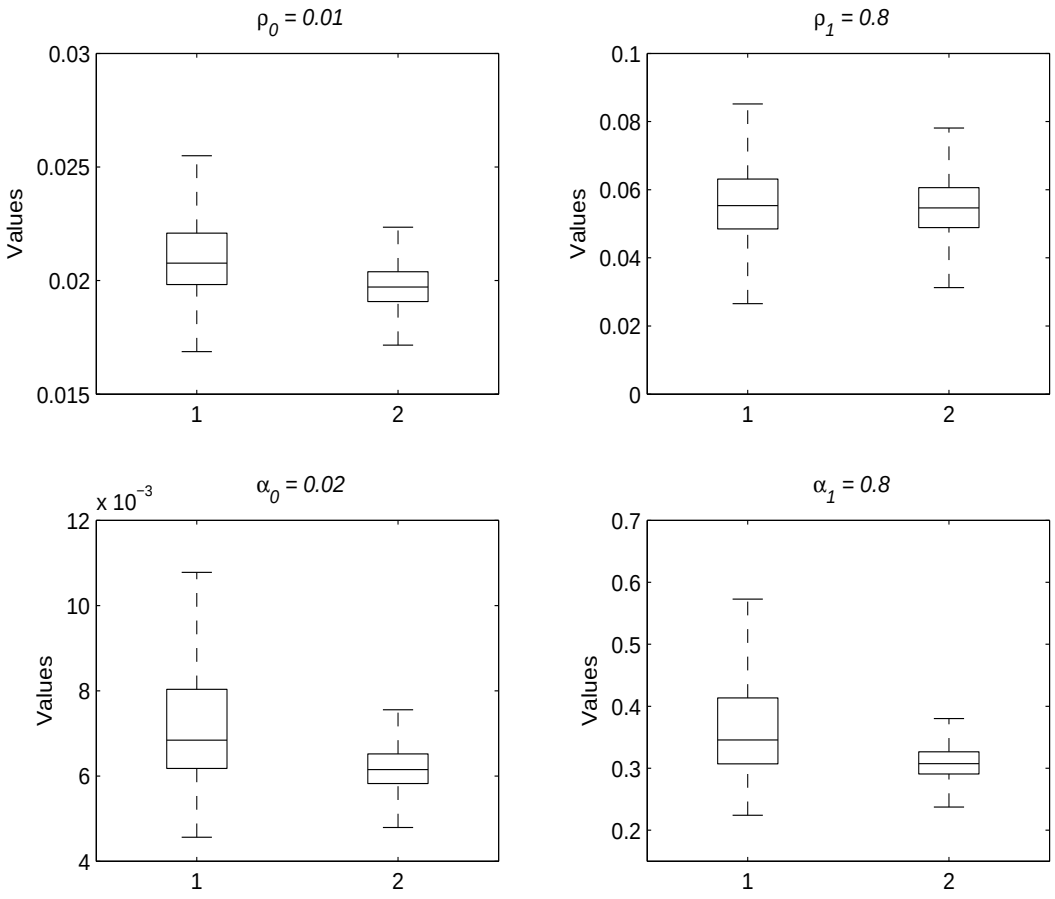

Figure 1.8: Boxplot of the lengths of PML (column 1) and robust (column 2) confidence intervals for $\hat{\theta}_0 := (\hat{\rho}_0 \;\; \hat{\rho}_1 \;\; \hat{\alpha}_0 \;\; \hat{\alpha}_1)^\top$ (cf. Figure 1.5) under the innovative outlier model (1.35).

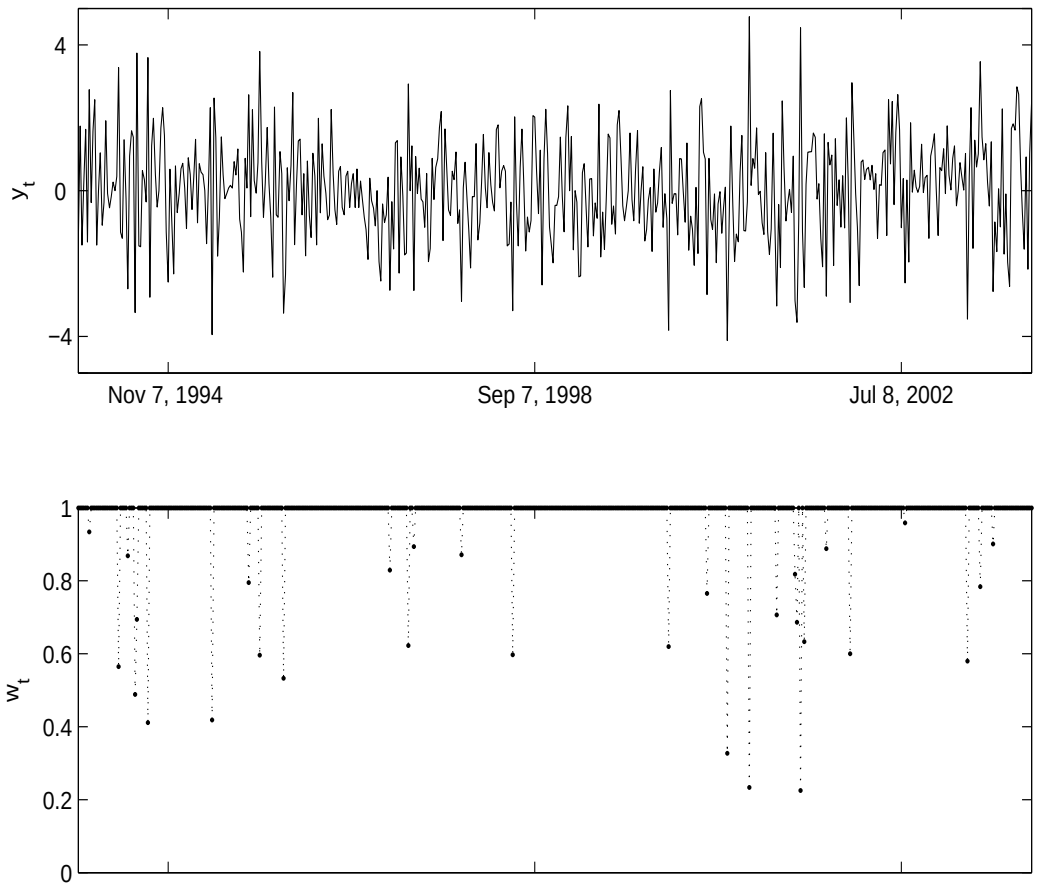

Figure 1.9: Weekly exchange rate returns of the Swedish krona versus the US dollar, for the period 11/29/1993 until 11/17/2003 (top panel) and the weights implied by the robust estimate of the AR(1)-ARCH(1) model with $c = 4$ (bottom panel).

Chapter 2

Robust Volatility Estimations for VaR Predictions

2.1 Abstract

This paper proposes a robust procedure to estimate asset volatilities in order to obtain accurate risk measure predictions. A model of the (G)ARCH family for historical portfolio returns is estimated with an efficient robust estimator. Resampling procedures are applied on standardized residuals to estimate VaR risk measures. In some Monte Carlo simulations we show that our robust approach performs well compared to competing approaches based on non robust volatility estimation procedures. Backtesting on four stock price series show that our robust approach gives more stable VaR profiles and reserve amounts than classical approaches and a similar backtesting performance.

2.2 Introduction

In this paper we propose a robust procedure for risk measure estimates, based on the robust filtering of historically simulated portfolio returns. Precisely, a model of the (G)ARCH family for portfolio returns volatilities is estimated by an efficient robust estimator. Historical residual returns are standardized by past conditional volatilities to make them suitable for historical simulation. Then, they are scaled by the current conditional volatility forecast to reflect current market conditions. Simulated financial returns based on these residuals allow to estimate the future distribution of portfolio returns. From such a distribution several measures of market risk can be easily computed; cf. Barone-Adesi, Giannopoulos and Vosper (1999, 2000).

Large portfolios of traded assets held by most financial institutions have made the measurement of market risk, i.e. the risk of losses on the trading book due to adverse market movements, a primary concern for regulators and for internal risk managers. In particular, banks are required to hold a certain amount of capital against adverse market movements. According to the Basle Committee (1996), the risk capital of any bank must be sufficient to cover losses on the trading book over a ten days holding period 99% of the times. Such a risk capital is usually called *Value at Risk* (VaR). Of course, holding periods and confidence levels may vary according to the form of investigated problems. For internal purposes, most banks use VaR at a 95% confidence level and a horizon of one day. From a statistical viewpoint the VaR is simply the quantile of the profit and loss (P&L) distribution of a given portfolio over a certain holding period.

In the financial literature other measures of market risk have been proposed. For instance, Artzner et al. (1999) showed the theoretical deficiencies of VaR as a measure of market risk and proposed the expected shortfall (or tail conditional expectation), i.e. the expected loss given that the loss exceeds the VaR, as a sounder alternative. The main problem in implementing all such risk measures is to

obtain accurate estimates for the tails of P&L distributions. Two main approaches have been proposed for estimating P&L distributions[1]: nonparametric historical simulation (HS) methods and parametric methods based on dynamic models for the conditional volatility of asset returns.

In the HS method the estimated P&L distribution of a portfolio is simply the empirical distribution of past gains and losses that the bank would have earned by holding that portfolio. Hence, the method is easy to implement and avoids "ad hoc" assumptions on the shape of the P&L distribution (such as normality). However, this method suffers from serious drawbacks. Firstly, the implicit assumption of independent and identically distributed (i.i.d.) historically simulated returns is unrealistic. It is well-known that the volatility of asset returns tends to change in time and that periods of high volatilities tend to cluster; cf., for instance, Cont (2001). The method is unable to distinguish between periods of high and low volatility and can not reflect the present market conditions.

Parametric models of volatility dynamics, such as GARCH models, yield risk measure estimates which reflect the current level of volatilities. However, such models often assume conditional normality of innovations (see, e.g., JP Morgan's Riskmetrics (1995)), while real data seems to exhibit excess kurtosis. Hence, these methods tend to underestimate the portfolio riskiness.

The filtered historical simulation (FHS) method offers the advantages of the previous two approaches. The conditional heteroscedasticity of asset returns is taken into account by modelling asset returns with a GARCH-type model. The scaled residuals (or "filtered" returns) are, at least approximately, an i.i.d. series suitable for resampling procedures and historical simulation. The non-normality of conditional asset returns typically induces a heavy tailed distribution for the scaled residuals.

So far the FHS method have been applied based on the PML estimates of GARCH-type models; cf., for instance, Barone-Adesi et al. (1999) and McNeil and Frey (2000). Unfortunately, it is known that such estimates can be largely determined by a few outlying observations, inducing misleading

[1]See for instance Duffie and Pan (1997) and the references therein.

inference results; cf., for instance, Künsch (1984), Sakata and White (1998) and Mancini, Ronchetti and Trojani (2004). Of course, outlying observations (due for instance to some liquidity patterns or market crashes) are a fundamental component of the data generating process and in particular of the riskiness we aim to measure. However, such outlying observations are typically unexplained by parametric volatility models, which can reasonably well-describe only some "structural" part of the data generating process. Moreover, these "shocks" can induce biased PML estimates, i.e. parameter estimates not representative for the parametric reference model. In these cases, filtering and forecasting procedures can be negatively affected using biased inadequate parameter estimates.

In this paper we propose to "filter" the historical portfolio returns by some robustly estimated GARCH model (cf. Mancini et al. (2004)) and then to estimate the future distribution of financial returns by applying the FHS method. For brevity, we call this procedure robust FHS method. Using our robust estimator, efficient parameter estimates for the parametric reference model are obtained under general local deviations of the true data generating process from such a reference model. In a second step, historical portfolio returns are filtered using such a reference model and VaR estimates are computed applying resampling procedures.

In order to verify the accuracy of VaR figures reported by financial institutions, the Basle Committee (1996) recommends the so-called "backtesting" procedure. If an institution correctly reports, say, the daily VaR's at 1% confidence level the actual daily losses should exceed (on average) the reported VaR's only 1% of the time. The trading day in which the actual loss exceeds the reported VaR is considered a violation. If the number of violations is "high" regulators will force the bank to increase the capital charge. As interest rates paid on such capital charges are virtually negligible, the bank will suffer a substantial opportunity cost. In risk management, the stability over time of the reserve amounts is also a desirable feature, as for a financial institution it is difficult to rapidly adjust the capital base. Furthermore, stable VaR profiles over time also allow to adapt outstanding risk

exposures to VaR limits more smoothly and, thus, more efficiently. In the backtest of Section 2.5 we show that the proposed robust FHS method induces more stable reserve estimates then the classical ones and a similar backtesting performance.

Recently, Diebold, Schuermann and Stroughair (1998) and McNeil and Frey (2000) proposed an approach to estimate risk measures based on the FHS method, where each tail of the filtered return distribution is separately estimated using an Extreme Value Theory approach. This procedure offers a parametric form for the tails of the conditional return distributions and allows for some extrapolation beyond the range of the data, even if care is required at this point[2]. A potentially alternative approach is to compute risk measures "directly", i.e. without using the FHS method, applying statistical tools designed to estimate regression quantiles and conditional distributions; see for instance Koenker and Bassett (1978), Foresi and Peracchi (1995) and Peracchi (2002). However, possible drawbacks of regression quantiles are their behaviour under heteroschedasticity and the non robustness to "bad" leverage points; see Koenker and Bassett (1982).

The structure of the paper is as follows. Section 2.3 briefly recalls the FHS method and introduces the robust estimator applied in the robust FHS method. Section 2.4 shows some Monte Carlo simulations comparing classical and robust FHS methods. Section 2.5 presents the out of sample backtesting performances on four stock price series of classical and robust FHS methods. Section 2.6 concludes and outlines some directions for future research.

---

[2]A potential disadvantage of this approach is due to the separate estimation of the tails distribution, because an increase in the market risk warned by large positive returns remains undetected by this procedure.

2.3 The Model

Let $\mathcal{Y} := (Y_t)_{t\in\mathbb{Z}}$ be a stationary time series process on the probability space $(\mathbb{R}^\infty, \mathcal{F}, \mathbb{P}_*)$ modelling daily rate of returns on a financial asset with price P_t at time t, i.e. $Y_t := P_t/P_{t-1} - 1$. Assume that $\mathbb{P}_*$ can be "approximated" by some parametric model $\mathcal{P} := \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$. Precisely, $\mathbb{P}_* \in \mathcal{U}(\mathbb{P}_{\theta_0})$ for some $\theta_0 \in \Theta$ and $\mathcal{U}(\mathbb{G})$ denotes some nonparametric neighborhood of $\mathbb{G}$. Under $\mathbb{P}_{\theta_0}$, Y_t satisfies

$$Y_t = \mu_t(\theta_0) + \sigma_t(\theta_0)\, Z_t, \tag{2.1}$$

where $\mu_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ parameterize the conditional mean and the conditional variance of Y_t, given the information $\mathcal{F}_{t-1}$ up to time $t-1$. Under $\mathbb{P}_{\theta_0}$ the innovations Z_t's are a strict white noise, i.e. $Z_t \sim i.i.d.\mathcal{N}(0,1)$ for all $t \in \mathbb{Z}$. Denote by $Y_1^m := (Y_1, \ldots, Y_m)$ the finite random sequence of $\mathcal{Y}$ and by $\mathbb{P}_*^m$ ($\mathbb{P}_{\theta_0}^m$) the m-dimensional marginal distribution of Y_1^m induced by $\mathbb{P}_*$ ($\mathbb{P}_{\theta_0}$). Denote by $F_{t,h}(\cdot)$ the conditional returns distribution induced by $\mathbb{P}_*$ over h days $Y_{t+h,h} := P_{t+h}/P_t - 1$, given the information $\mathcal{F}_t$. For $0 < \alpha < 1$ and horizon h days, the corresponding quantile of $F_{t,h}(\cdot)$ is

$$y_{t+h}^{\alpha,h} := \inf\{y \in \mathbb{R} : F_{t,h}(y) \geq \alpha\}.$$

Formally, the required capital or Value at Risk (VaR) at time t for an institution investing in a financial asset with market price P_t is $VaR_t^{\alpha,h} = P_t + R_t^{\alpha,h}$, where $R_t^{\alpha,h}$ is the reserve amount such that the probability of a loss over the next h days is "small" and equal[3] to some level α

$$\alpha = \mathbb{P}_*(P_{t+h} + R_t^{\alpha,h} < 0\,|\mathcal{F}_t) = \mathbb{P}_*(P_{t+h} - P_t < -VaR_t^{\alpha,h}\,|\mathcal{F}_t). \tag{2.2}$$

Hence, $-VaR_t^{\alpha,h} = P_t\, y_{t+h}^{\alpha,h}$ is the α quantile of the P&L distribution under $\mathbb{P}_*$ over the next h days, given the available information $\mathcal{F}_t$. From an economic viewpoint, $VaR_t^{\alpha,h}$ is the minimum loss over the next h days with probability α.

[3]For simplicity, in equation (2.2) we consider a continuous P&L distribution under $\mathbb{P}_*$.

An alternative measure of market risk is the expected shortfall,

$$S_{t+h}^{\alpha,h} := E_*[Y_{t+h,h} \,|\, Y_{t+h,h} < y_{t+h}^{\alpha,h}\,,\, \mathcal{F}_t],$$

where $E_*[\cdot]$ denotes expectation with respect to $\mathbb{P}_*$; cf. Artzner et al. (1999). When $h = 1$ the previous risk measures simplify to

$$y_t^{\alpha,1} = \mu_t(\theta_0) + \sigma_t(\theta_0)\, z_\alpha \quad \text{and} \quad S_t^{\alpha,1} = \mu_t(\theta_0) + \sigma_t(\theta_0)\, E_*[Z \,|\, Z < z_\alpha],$$

where z_α is the α quantile of the distribution of Z_t, which by assumption does not depend on t.

Our aim is to derive robust inference on such risk measures (for different horizons h and confidence levels α) when the true underlying distribution $\mathbb{P}_*$ is unknown and belongs to some nonparametric neighborhood of the parametric reference model $\mathbb{P}_{\theta_0}$. Then, model (2.1) is regarded as an "approximate" description of the true data generating process $\mathbb{P}_*$.

2.3.1 Estimation Procedures

Typically, model (2.1) is estimated by a Pseudo Maximum Likelihood (PML) approach (cf. Gourieroux, Monfort and Trognon (1984)), under the *nominal* assumption of Gaussian innovations. The functional PML estimator $a : \mathcal{P} \longrightarrow \Theta$ is implicitly defined by the estimating equation

$$E_*[s(Y_1^m; a(\mathbb{P}_{\theta_0}^m))] = 0,$$

where

$$s(Y_1^m; \theta) = \frac{1}{\sigma_m^2(\theta)} \frac{\partial \mu_m(\theta)}{\partial \theta} \varepsilon_m(\theta) + \frac{1}{2\sigma_m^2(\theta)} \frac{\partial \sigma_m^2(\theta)}{\partial \theta} \left(\frac{\varepsilon_m^2(\theta)}{\sigma_m^2(\theta)} - 1 \right) \tag{2.3}$$

and $\varepsilon_m(\theta) := \sigma_m(\theta) Z_m$. Under the model $\mathbb{P}_{\theta_0}^m$ the PMLE is the MLE and has an asymptotic normal distribution with covariance matrix $V(s; \theta_0) = I(\theta_0)^{-1}$, where $I(\theta) = E_*[s(Y_1^m; \theta)\, s(Y_1^m; \theta)^\top]$ is the information matrix.

However, if indeed model (2.1) is slightly different from the true data generating process of $\mathcal{Y}$, PML estimators can induce biased and inefficient inference results on θ_0; cf., for instance, Künsch (1984), Sakata and White (1998) and Mancini, Ronchetti and Trojani (2004). As one could expect, this will induce also inaccurate risk measure estimates; cf. Section 2.4.

In the present time series setting, to describe the linearized asymptotic bias of a statistical functional $\tilde{a}$ in a neighborhood of $\mathbb{P}_{\theta_0}$ we introduce the following concept from the theory of robust statistics; cf. Künsch (1984) and Hampel et al. (1986).

Definition 2.1 The conditional influence function of a statistical functional[4] $\tilde{a}$ is the function $IF : \mathbb{R}^m \times \Theta \longrightarrow \mathbb{R}^p$ such that

$$i) \quad \lim_{\eta \downarrow 0} \frac{\tilde{a}((1-\eta)\mathbb{P}^m_{\theta_0} + \eta\, \mathbb{G}^m) - \tilde{a}(\mathbb{P}^m_{\theta_0})}{\eta} = \int_{\mathbb{R}^m} IF(y; \tilde{a}(\mathbb{P}^m_{\theta_0}))\, d\mathbb{G}^m(y),$$

$$ii) \quad E_{\theta_0}[IF(y; \tilde{a}(\mathbb{P}_{\theta_0}))|\mathcal{F}_{m-1}] = 0,$$

where $E_{\theta_0}[\cdot]$ denotes the expectation with respect to the reference distribution $\mathbb{P}^m_{\theta_0}$, $\mathbb{G}^m$ is any stationary m-dimensional distribution on $\mathbb{R}^m$, and provided the limit exists.

A bounded conditional IF ensures a bounded linearized asymptotic bias of the estimator induced by any distribution in a neighborhood of $\mathbb{P}^m_{\theta_0}$. Unfortunately, the bias of the PMLE is given by

$$a(\mathbb{P}^m_*) - a(\mathbb{P}^m_{\theta_0}) \propto \int_{\mathbb{R}^m} s(y; \theta_0)\, d\mathbb{G}^m(y),$$

where the score function s is defined in equation (2.3) and is unbounded at least in $\varepsilon_m(\theta)$.

In order to obtain robust estimates for θ_0 in the neighborhood of $\mathbb{P}^m_{\theta_0}$, we use the robust estimator proposed in Mancini et al. (2004).

---

[4]In the following we will always assume that the domain of the given statistical functional is an open convex subset of $\mathcal{P}$ containing $\mathbb{P}_{\theta_0}$ and all empirical measures.

Let $\psi_c(Y_1^m;\theta) := A(\theta)\ \big(s(Y_1^m;\theta) - \tau(Y_1^{m-1};\theta)\big)\, w(Y_1^m;\theta)$. We define a robust functional M-estimator $\overline{a}(\cdot)$ of θ_0 implicitly by

$$E_*[\psi_c(Y_1^m;\overline{a}(\mathbb{P}_{\theta_0}))] = 0, \tag{2.4}$$

where $w(Y_1^m;\theta) := \min(1,\ c\,\|A(\theta)\,\big(s(Y_1^m;\theta) - \tau(Y_1^{m-1};\theta)\big)\,\|^{-1})$. The non singular matrix $A(\theta) \in \mathbb{R}^p \times \mathbb{R}^p$ and the $\mathcal{F}_{m-1}$-measurable random vectors $\tau(Y_1^{m-1};\theta) \in \mathbb{R}^p$ are determined by the implicit equations

$$E_{\theta_0}[\psi_c(Y_1^m;\theta_0)\ \psi_c(Y_1^m;\theta_0)^\top] = I, \tag{2.5}$$

$$E_{\theta_0}[\psi_c(Y_1^m;\theta_0)|\mathcal{F}_{m-1}] = 0. \tag{2.6}$$

The estimating function ψ_c is a truncated version of the ML score (2.3) and can be interpreted as a weighted ML score as, by construction, $\|\psi_c(Y_1^m;\theta)\| \leq c$. The constant $c \geq \sqrt{p}$ is chosen by the researcher and controls for the degree of robustness of the robust estimator[5]. When $c = \infty$, $\overline{a}$ is the MLE of θ_0 under $\mathbb{P}_{\theta_0}^m$.

The robust estimator defined in (2.4)–(2.6) is the most efficient estimator among all robust estimators $\overline{a}$ satisfying equation (2.6) and the robustness constraint $\sup_{z\in\mathbb{R}^m} \|V(\psi_c;\theta_0)^{-1/2} IF(z;\overline{a}(\mathbb{P}_{\theta_0}^m))\| \leq c$; cf. Mancini et al. (2004), Proposition 1.

2.3.2 Filtering Historical Simulations

In order to implement an estimation procedure for VaR risk measures, the dynamics for the conditional mean and volatility of asset returns in model (2.1) have to be specified. Many different models have been proposed in the econometric literature such as GARCH models (cf. Bollerslev et

[5]See Ronchetti and Trojani (2001) for a choice of c based on the maximal bias in the level and the power of tests derived by the corresponding robust estimator.

al. (1992)), HARCH models (Müller et al. (1995)) and stochastic volatility models (cf., for instance, Shepard (1996)). In this paper we assume an AR(1) model for the conditional mean and a GARCH(1,1) model for the volatility process; the approach we propose extends easily to more general models. Hence, the conditional mean and variance of the innovation $\varepsilon_t(\theta) = Y_t - \mu_t(\theta)$ are

$$\mu_t(\theta_0) = \rho_0 + \rho_1 Y_{t-1}, \tag{2.7}$$

$$\sigma_t^2(\theta_0) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2(\theta_0) + \beta \sigma_{t-1}^2(\theta_0), \tag{2.8}$$

where α_0, α_1, $\beta > 0$, $|\rho_1| < 1$ and $\alpha_1 + \beta < 1$.

To compute VaR risk measures, we follow the approach of filtering the historical simulation as proposed by Barone-Adesi, Bourgoin and Giannopoulos (1998). We briefly review the method which is in two steps. Given the observed daily rate of returns $\{y_1, y_2, \ldots, y_T\}$ at the end of day T,

Step 1. Fit the model (2.7)–(2.8) to the sample data to get the parameter estimate $\hat{\theta} = (\hat{\rho}_0\ \hat{\rho}_1\ \hat{\alpha}_0\ \hat{\alpha}_1\ \hat{\beta})^\top$. Then, estimate the scaled residuals (or "filter" the observed returns)

$$\hat{z}_t = \frac{y_t - \mu_t(\hat{\theta})}{\sigma_t(\hat{\theta})},$$

for all $t = 1, \ldots, T$, where the conditional mean and standard deviation series $(\mu_t(\hat{\theta}), \sigma_t(\hat{\theta}))_{t=1,\ldots,T}$ are computed recursively from (2.7)–(2.8) after having substituted sensible starting values.

Step 2. Randomly select one estimated scaled innovation, say $z_1^\star$, from $\{\hat{z}_1, \ldots, \hat{z}_T\}$, and compute $\varepsilon_{T+1}^\star := \hat{\sigma}_{T+1} z_1^\star$ to simulate

$$y_{T+1}^\star = \hat{\rho}_0 + \hat{\rho}_1 y_T + \varepsilon_{T+1}^\star \quad \text{and} \quad p_{T+1}^\star = p_T\,(1 + y_{T+1}^\star).$$

Randomly select (with replacement) a second scaled innovation, say $z_2^\star$, and compute $\varepsilon_{T+2}^\star := \sigma_{T+2}^\star z_2^\star$, where $\sigma_{T+2}^\star = (\hat{\alpha}_0 + \hat{\alpha}_1 \varepsilon_{T+1}^{\star 2} + \hat{\beta}\sigma_{T+1}^2)^{1/2}$. Simulate

$$y_{T+2}^\star = \hat{\rho}_0 + \hat{\rho}_1 y_{T+1}^\star + \varepsilon_{T+2}^\star \quad \text{and} \quad p_{T+2}^\star = p_{T+1}^\star\,(1 + y_{T+2}^\star)$$

and so on up to $p^{\star}_{T+h}$. Repeat 10,000 times the procedure in Step 2 to estimate the distribution of $Y_{T+h,h}|\mathcal{F}_T$. The h-days-ahead VaR forecast at a 1% level is the 1% percentile of such a distribution. Similarly, the expected shortfall measure is readily computed. The 1-day-ahead VaR forecast can also be computed using the empirical version of the quantile z_α.

The main advantage of this procedure is that it accounts for the heteroscedasticity observed in most financial time series returns and that it avoids ad hoc assumptions on the error distribution (such as normality). This procedure is semi parametric as volatilities are modeled by a parametric GARCH model and the innovation distribution is estimated by a nonparametric resampling method.

The robust methods developed in Mancini et al. (2004) do not apply to situations in which the conditional variance depends on the entire data history, as for instance in GARCH models. However, as in Sakata and White (1998), we can heuristically expect our robust estimator to accurately perform under well behaved GARCH models with sufficient memory decay such that the effects of outliers distant in time may die out. The robust methods developed in Mancini et al. (2004) do not explicitly study such a setting. Nevertheless, it seems plausible that the estimator $\overline{a}$ defined in (2.4)–(2.6) may be shown to have relatively high robustness even for such GARCH models. Therefore, the robust estimator $\overline{a}(\cdot)$ allows to obtain efficient parameter estimates of the model (2.7)–(2.8) under general local deviations of the true data generating process from such a reference model. Clearly, accurate parameter estimates are of critical importance as they enter both in the filtering of historical returns and in the estimation of the distribution $Y_{T+h,h}|\mathcal{F}_T$. When the model (2.7)–(2.8) is estimated by the robust estimator (2.4)–(2.6), the procedure is called robust FHS method. When the model (2.7)–(2.8) is estimated by the PML estimator defined by (2.3), the procedure is called FHS method. In the next sections we compare the risk measure estimates of the classical and robust FHS methods on simulated and real data.

2.4 Monte Carlo Simulations

To compare the performance of the classical and the robust FHS method we compute the *out of sample* predictions of the VaR for confidence levels $\alpha = 5\%$, 1% and horizons $h = 1$ day, 10 days when the daily returns Y_t follow an AR(1)-GARCH(1,1) model. We simulate the following dynamics for $\mathcal{Y}$.

1. *Student t_5 innovation model.* In this experiment, the scaled innovation in (2.1) is given by

   $$Z_t = ((\nu - 2)/\nu)^{1/2} T_{\nu,t},$$

   where $T_{\nu,t}$ has a Student t distribution with $\nu > 2$ degrees of freedom for all $t \in \mathbb{Z}$. Hence, $Z_t \sim i.i.d.(0,1)$, the model (2.1) is dynamically correctly specified and the estimator defined by (2.3) is indeed a PMLE. Therefore, there are no theoretical reasons to prefer our robust estimator to the PML estimator. Notice that the t_5 distribution is quite close to the normal distribution under which the estimator defined by (2.3) is the MLE.

2. *Laplace innovation model.* The scaled innovation in (2.1) is given by

   $$Z_t = 2^{-1/2} L_t,$$

   where L_t has a Laplace (or Double exponential) distribution for all $t \in \mathbb{Z}$. Such a distribution has a symmetric convex density and displays more heavy tails than the t_5. Also in this experiment $Z_t \sim i.i.d.(0,1)$ and model (2.1) is dynamically correctly specified.

3. *Replace-innovative model.* Under such a model the observed process $\mathcal{Y} := (Y_t)_{t\in\mathbb{Z}}$ is generated according to the data generating process

   $$Y_t = \begin{cases} \rho_0 + \rho_1 Y_{t-1} + \varepsilon_t, & 1-q, \\ \check{Y}_t, & q, \end{cases}$$

   where $\check{Y}_t \sim \mathcal{N}(0, \varrho^2)$, $\varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ and σ_t^2 is given by (2.8). Hence, at a time t the observed Y_t is not generated by the GARCH dynamic with probability q. The possible "shock" $\check{Y}_t$ will

affect future realizations of Y, mainly "inflating" the conditional variance on subsequent days. In this experiment the model (2.1) is "slightly" misspecified as the dynamic equations (2.7)–(2.8) are not satisfied. We set $\varrho = 10$ and $q = 0.2\%$. The choice for ϱ is quite extreme and allows to compare the performance of the two approaches under very infrequent, dramatic, symmetric shocks that could occur over short time periods in real data, as for instance in daily equity or exchange rate returns.

We simulate an AR(1)-GARCH(1,1) model for the following parameter choice[6] $\rho_0 = \rho_1 = 0.01$, $\alpha_0 = 0.03$, $\alpha_1 = 0.1$ and $\beta = 0.8$ under the above distribution for Y_t and for a sample size $T = 1{,}000$. Hence, under the reference model $\mathbb{P}_{\theta_0}$ the volatility of Y_t is about 10% on an annual base. The tuning constant of the robust estimator $\overline{a}$ was set at $c = 9$. Such a rather large value of the tuning constant c implies that only a few observations are downweighted and this is motivated by the high quality of financial data. Each model is simulated 1,000 times. For each simulation we compute $y_{T+h}^{\alpha,h}$, i.e. the out of sample VaR predictions (as a percentage of the simulated price p_T) for confidence levels $\alpha = 5\%, 1\%$ and horizons $h = 1$ day, 10 days. We apply the FHS method based on the classical and the robust estimates of the model (2.7)–(2.8).

Table 2.1 shows bias and mean squared error of the PML and robust parameter estimates for the AR(1)-GARCH(1,1) model (2.7)–(2.8). The robust estimator largely outperforms the classical PML estimator in terms of bias and mean squared error in all experiments and for almost all estimated parameters.

---

[6]Similar parameter estimates have been reported by Bollerslev, Engle and Nelson (1994) for the daily rate of returns of the Deutschemark versus the U.S. Dollar exchange rate from January 2$^{\text{nd}}$, 1981 to July 9$^{\text{th}}$, 1992.

2.4.1 VaR Violations

Table 2.2 shows for the classical and robust FHS method the number of violations of the VaR estimates for horizons $h = 1$ day, 10 days and confidence levels $\alpha = 5\%$, 1%. In the i-th simulation, a violation occurs when the actual loss is larger then the predicted VaR, i.e. $I(i) := \mathbf{1}_{\{y_{T+h,h}(i) < y^{\alpha,h}_{T+h}(i)\}} = 1$. Under the null hypothesis that the proposed method correctly estimates the VaR, the test statistic $\sum_{i=1}^{1000} I(i)$ is binomially distributed $Bin(1000, \alpha)$; cf. for instance Christoffersen et al. (1998). Hence, for $\alpha = 0.05$ and 0.01 the expected number of violations for both methods are 50 and 10 and two-sided confidence intervals at a 95% level are $[37, 64]$ and $[4, 17]$, respectively. Both methods exhibit numbers of violations within such intervals.

2.4.2 VaR Predictions

Tables 2.3–2.4 show the performances of the classical and robust FHS method in predicting the VaR for horizons $h = 1$ day, 10 days and confidence levels $\alpha = 5\%$, 1% in the different experiments. To evaluate the overall performances, we compute the bias and the mean squared error (MSE) of the VaR estimates. Then, we split our simulation results in two subsets. The first subset collects all simulations where the predicted VaR's exceed (in absolute value) the true VaR's and the second subset the opposite cases. The first subset denoted by VaR^+ reflects an "opportunity cost" for a bank associated with an unduly high reserve amount. The second subset denoted by VaR^- reflects an "hidden risk" undetected by the VaR estimates. Under a given experiment, for each simulation, we compute the true VaR by simulating 100,000 times the true dynamic for $\mathcal{Y}$ over the relevant time horizon $[T, T + h]$ and then by computing the relevant quantiles of the corresponding empirical distribution of $Y_{T+h,h}|\mathcal{F}_T$.

In the first two experiments (Student t_5 and Laplace innovation model), the classical and the robust FHS method have similar biases, but the robust FHS method always displays a lower MSE; cf. also the first two panels in Figures 2.1–2.2. For the horizon $h = 1$ day the reduction in the MSE's is (about)

10% for the Student t_5 and 25% for the Laplace innovation model. For the horizon $h = 10$ days these reductions tend to decrease (especially at 1% confidence level and for the Laplace innovation model). The classical FHS method displays a rather low bias in the VaR estimates under the Laplace innovation model. For the horizon $h = 1$ day the statistics for VaR^+ and VaR^- are almost always in favor of the robust FHS method. For the horizon $h = 10$ days the robust FHS method still outperforms the classical ones.

In the third experiment (replace-innovative model), for the horizon $h = 1$ day, the robust FHS method largely outperforms the classical one in terms of MSE's, means and standard deviations of VaR^+ and VaR^-. For the horizon $h = 10$ days the classical FHS method breaks down, while the robust FHS method still delivers quite reasonable estimates for the VaR at a 5% confidence level. For the VaR at a 1% confidence level also the robust FHS method breaks down; cf. also the third panel in Figures 2.1–2.2.

2.4.3 VaR Levels in High Volatility Periods

Since accurate VaR estimates are particularly important in "high volatility" periods, for each experiment we collect VaR predictions where σ_T^2, i.e. the conditional variance at the end of the "estimating period", was in the upper decile of its simulated distribution. Then, given such VaR predictions for each confidence level and horizon we compute the actual probability of a violation, i.e. $\mathbb{P}_*(Y_{T+h,h} < \hat{y}_{T+h}^{\alpha,h} | \mathcal{F}_T)$, which is equal to α under the null hypothesis that the proposed method correctly estimates the VaR. Figures 2.3–2.4 show the boxplots of such probabilities for horizons $h = 1$ day, 10 days. In the first two experiments the classical FHS method tends to underestimate the VaR (being the corresponding, actual probabilities larger than α), especially for the 1% confidence level. The robust FHS method performs generally well and only under the Laplace innovation model for the horizon $h = 10$ days seems to be too conservative. In the last experiment the robust FHS method

largely outperforms the classical one, even though such differences decrease from horizon $h = 1$ day to horizon $h = 10$ days. Similar boxplots for the "medium" and the "low" volatility cases (not reported here) show that in such cases the overall performances of the two methods are quite close. As one could expect, accurate model parameter estimates are more valuable in "high" volatility periods than in the "low" volatility ones.

2.5 Backtesting

We backtest the method on four historical series of daily rate of returns: the Dow Jones index, the BMW, Microsoft and Nestle share prices from December 10^{th}, 1993 to December 10^{th}, 2003. The data were downloaded from Datastream. To backtest the FHS methods on a historical series $y_1, \ldots, y_l$, where $l \gg n$, we compute $y_{T+h}^{\alpha,h}$ for $T \in \mathcal{T} = \{n, n+1, \ldots, l-h\}$ using a time window of n days for each estimate. In our implementation we set $n = 1{,}000$, hence using about the last four years of data for each prediction. For each day $T \in \mathcal{T}$ we estimate an AR(1)-GARCH(1,1) model using the PML estimator defined by (2.3) and our robust estimator (2.4)–(2.6). Then, using resampling methods we compute the out of sample VaR and reserve estimates based on classical and robust filtered returns for the usual horizons and confidence levels.

2.5.1 VaR Violations

Table 2.5 shows the number of violations and the two-sided p-values for the null hypothesis that the proposed method correctly predicts the VaR's; cf. also Figures 2.5–2.8. Both methods perform well with the exception of the classical FHS method in the backtest on Microsoft returns. In this case, the high number of violations is due to a few large spikes in the data which "inflate" the classical estimates of the conditional variances, inducing a "thin" tailed distribution for the classical residuals and an

underestimation of actual VaR's. By contrast, our robust estimator yields parameter estimates of α_0 (which controls the level of volatility) 60% lower on average than the corresponding PML estimates. Hence, robust residuals have a more heavy tailed distribution.

2.5.2 Reserve Amounts

Tables 2.6–2.7 show summary statistics of the reserve amounts defined by (2.2) and estimated using the classical and robust FHS methods for horizons $h = 1$ day and $h = 10$ days, respectively; cf. again Figures 2.5–2.8. For comparisons, we compute reserve amounts for portfolios invested in each single asset and valued 100 units of domestic currency at December 10^{th}, 1993. In the backtestings on the Dow Jones and BMW returns, for both horizons $h = 1$ day, 10 days, in terms of standard deviations and interquartile ranges, the robust FHS method yields more stable reserve estimates (about 20%) then the classical ones, especially for the reserves at 1% confidence level. For the backtest on the Nestle returns, the overall performances of the two methods are quite close.

2.6 Conclusions and Future Developments

We proposed an efficient robust method to estimate asset return volatilities. Then, resampling procedures on standardized residuals are applied in order to obtain accurate out of sample VaR predictions. By means of Monte Carlo simulations we showed that the proposed method performs well compared to competing approaches based on non robust volatility estimation procedures. Backtesting on four stock price series showed that our robust approach gives more stable VaR profiles and reserve amounts than classical (non robust) approaches and similar backtesting performances in terms of number of violations.

The present work suggests some possible directions for future research. First, it would be interesting

to develop a robust VaR estimation procedure based on the present robust volatility estimation and on a robust resampling method or a robust estimation of the tails of innovation distributions. Second, the backtesting considered in Section 2.5 could be repeated using asymmetric/threshold GARCH models for stock returns able to account for possible leverage effects in the data. Finally, backtesting performances of the proposed method could be investigated also for other price series, such as interest or exchange rates.

	$\rho_0 = 0.01$		$\rho_1 = 0.01$		$\alpha_0 = 0.03$		$\alpha_1 = 0.1$		$\alpha_1 = 0.8$	
	bias	MSE%	bias	MSE%	bias	MSE%	bias	MSE%	bias	MSE%
cl	0.0000	0.0271	−0.0019	0.1354	0.0016	0.0246	0.0038	0.1933	−0.0076	0.7730
rob	0.0005	0.0133	0.0004	0.0134	−0.0035	0.0162	−0.0024	0.1541	0.0049	0.6162
cl	0.0006	0.0268	−0.0010	0.1367	0.0010	0.0191	0.0018	0.1510	−0.0036	0.6041
rob	−0.0001	0.0001	−0.0001	0.0001	−0.0012	0.0012	0.0021	0.0088	−0.0042	0.0351
cl	−0.0023	0.1833	−0.0027	0.6162	0.0674	2.0589	0.0356	1.2348	−0.0712	4.9392
rob	−0.0003	0.0282	−0.0002	0.0301	0.0116	0.1899	−0.0023	0.1616	0.0046	0.8464

Table 2.1: Bias and MSE of estimated parameters of the AR(1)-GARCH(1,1) model under Student t_5 innovation (first panel), Laplace innovation (second panel), replace-innovative model (third panel).

	Stud t_5	Laplace	rep-innov	Stud t_5	Laplace	rep-innov
cl - 5%	59	54	53	45	47	37
rob - 5%	61	51	53	44	49	41
cl - 1%	15	15	12	8	11	8
rob - 1%	15	13	9	7	10	8

Table 2.2: Number of violations for horizons $h = 1$ day (left panel) and $h = 10$ days (right panel).

	bias	MSE	mean VaR^+	std VaR^+	mean VaR^-	std VaR^-
cl - 5%	-0.0079	0.0052	0.0485	0.0467	-0.0528	0.0546
rob - 5%	-0.0067	0.0045	0.0472	0.0460	-0.0488	0.0482
cl - 1%	-0.0390	0.0254	0.1063	0.0955	-0.1270	0.1113
rob - 1%	-0.0381	0.0234	0.1100	0.0956	-0.1211	0.0996
cl - 5%	0.0005	0.0051	0.0572	0.0510	-0.0520	0.0413
rob - 5%	0.0031	0.0034	0.0498	0.0394	-0.0423	0.0319
cl - 1%	-0.0026	0.0215	0.1182	0.1029	-0.1072	0.0854
rob - 1%	0.0048	0.0168	0.1089	0.0906	-0.0950	0.0684
cl - 5%	-0.0079	0.0801	0.0823	0.1030	-0.1141	0.3761
rob - 5%	-0.0059	0.0117	0.0415	0.0374	-0.0526	0.1324
cl - 1%	0.0163	0.1379	0.1555	0.1832	-0.1565	0.4620
rob - 1%	-0.0064	0.0304	0.0778	0.0680	-0.0887	0.2045

Table 2.3: Statistics of VaR estimates for horizon $h = 1$ day, confidence levels 5% and 1% under Student t_5 innovation (first panel), Laplace innovation (second panel) and replace-innovative model (third panel). VaR^+ denotes cases where the estimated VaR's exceed the true VaR and VaR^- opposite cases.

	bias	MSE	mean VaR^+	std VaR^+	mean VaR^-	std VaR^-
cl - 5%	-0.0237	0.0645	0.1946	0.1613	-0.2045	0.1527
rob - 5%	-0.0156	0.0541	0.1831	0.1463	-0.1868	0.1363
cl - 1%	0.1341	0.2034	0.3968	0.3086	-0.2813	0.2161
rob - 1%	0.1614	0.2011	0.4027	0.3126	-0.2489	0.1987
cl - 5%	0.0039	0.0682	0.2088	0.1619	-0.2053	0.1568
rob - 5%	0.0154	0.0671	0.2081	0.1731	-0.1951	0.1497
cl - 1%	-0.0116	0.1891	0.3492	0.2899	-0.3341	0.2503
rob - 1%	0.0379	0.1861	0.3646	0.3073	-0.2980	0.2350
cl - 5%	0.0296	0.9212	0.4301	0.5081	-0.4249	1.1342
rob - 5%	-0.0053	0.1552	0.2570	0.2415	-0.2469	0.3506
cl - 1%	0.6214	6.6174	2.2312	2.6798	-0.7710	1.1158
rob - 1%	0.6032	6.1515	2.3041	2.9306	-0.6261	0.4203

Table 2.4: Statistics of VaR estimates for horizon $h = 10$ days, confidence levels 5% and 1% under Student t_5 innovation (first panel), Laplace innovation (second panel) and replace-innovative model (third panel). VaR^+ denotes cases where the estimated VaR's exceed the true VaR and VaR^- opposite cases.

	Expect. 5%	cl - 5%	rob - 5%	Expect. 1%	cl - 1%	rob - 1%
Dow Jones	80	79 (0.88)	83 (0.76)	16	17 (0.82)	18 (0.63)
BMW	80	77 (0.70)	82 (0.85)	16	20 (0.32)	20 (0.32)
Microsoft	80	104 (0.01)	93 (0.15)	16	13 (0.44)	19 (0.46)
Nestle	80	81 (0.94)	76 (0.62)	16	18 (0.63)	16 (0.99)
Dow Jones	80	83 (0.88)	83 (0.88)	16	26 (0.39)	26 (0.39)
BMW	80	76 (0.84)	76 (0.84)	16	17 (0.90)	17 (0.90)
Microsoft	80	130 (0.06)	109 (0.22)	16	28 (0.27)	18 (0.79)
Nestle	80	84 (0.84)	85 (0.80)	16	21 (0.51)	19 (0.67)

Table 2.5: Backtesting results. Number of violations using classical and robust FHS method for horizon $h = 1$ day (first panel) and $h = 10$ days (second panel); p-values in parenthesis.

	mean	median	stdv	q_{25}	q_{75}	$q_{75} - q_{25}$
cl - 5%	243.2	214.0	138.9	150.8	308.3	157.4
rob - 5%	225.7	209.4	112.2	151.5	281.0	129.5
cl - 1%	546.2	499.1	223.3	394.9	642.5	247.6
rob - 1%	519.3	485.8	171.8	406.3	587.7	181.4
cl - 5%	915.7	855.9	284.7	713.5	1079.3	365.8
rob - 5%	893.4	854.5	239.8	730.6	1029.6	299.0
cl - 1%	1643.2	1554.5	456.6	1339.2	1879.0	539.8
rob - 1%	1623.9	1574.0	395.4	1347.2	1855.0	507.8
cl - 5%	3122.3	2959.3	1313.7	2397.9	3644.8	1246.9
rob - 5%	3375.9	3141.1	1479.8	2418.6	4101.0	1682.5
cl - 1%	6888.9	6572.5	2627.9	5352.3	8043.8	2691.6
rob - 1%	6242.9	5972.7	2369.3	4819.9	7371.9	2551.9
cl - 5%	314.0	273.4	173.6	189.1	397.3	208.2
rob - 5%	323.1	271.8	194.5	181.9	416.1	234.2
cl - 1%	702.4	644.5	299.5	492.0	840.2	348.2
rob - 1%	715.0	635.5	331.9	479.3	872.0	392.7

Table 2.6: Summary statistics of the reserve amounts determined by classical and robust FHS method for horizon $h = 1$ day, confidence levels 5%, 1% for the Dow Jones index (first panel), the BMW share price (second panel), the Microsoft share price (third panel) and the Nestle share price (fourth panel).

	mean	median	stdv	q_{25}	q_{75}	$q_{75} - q_{25}$
cl - 5%	1330.3	1268.5	416.8	1079.9	1528.8	448.9
rob - 5%	1271.1	1261.2	337.7	1087.4	1453.0	365.6
cl - 1%	2288.8	2203.8	632.9	1897.6	2600.3	702.7
rob - 1%	2142.5	2107.8	492.1	1873.1	2400.7	527.6
cl - 5%	3692.4	3617.3	875.6	3127.8	4145.6	1017.7
rob - 5%	3602.6	3592.4	753.6	3138.9	4012.1	873.2
cl - 1%	5692.9	5619.5	1246.1	4840.2	6355.3	1515.0
rob - 1%	5495.4	5486.5	1059.9	4817.2	6058.0	1240.8
cl - 5%	11536.0	11815.5	3447.6	9953.4	13293.2	3339.8
rob - 5%	12686.4	11924.6	5342.8	9747.4	15083.8	5336.4
cl - 1%	19033.4	19264.8	5748.5	16138.2	22216.2	6078.0
rob - 1%	20155.7	19167.3	8027.3	15285.4	23912.9	8627.5
cl - 5%	1530.8	1484.0	519.2	1140.7	1785.9	645.2
rob - 5%	1507.6	1488.7	497.7	1127.9	1762.0	634.1
cl - 1%	2524.4	2454.1	802.1	1897.5	2961.5	1064.0
rob - 1%	2498.2	2463.7	792.8	1885.3	2941.9	1056.6

Table 2.7: Summary statistics of the reserve amounts determined by classical and robust FHS method for horizon $h = 10$ days, confidence levels 5%, 1% for the Dow Jones index (first panel), the BMW share price (second panel), the Microsoft share price (third panel) and the Nestle share price (fourth panel).

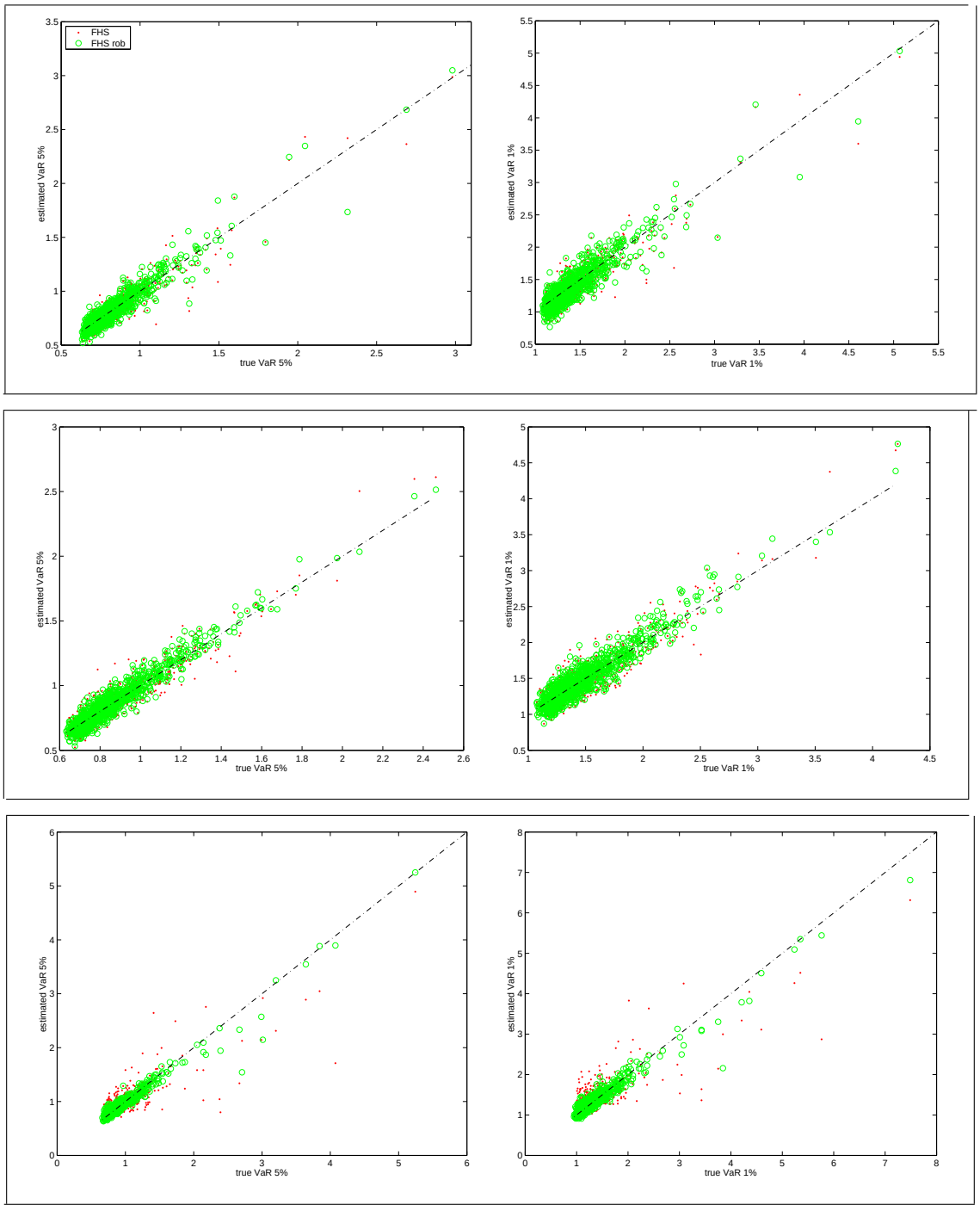

Figure 2.1: True VaR's versus estimated VaR's for horizon $h = 1$ day and confidence levels 5%, 1% under Student t_5 innovation (first panel), Laplace innovation (second panel) and replace-innovative model (third panel).

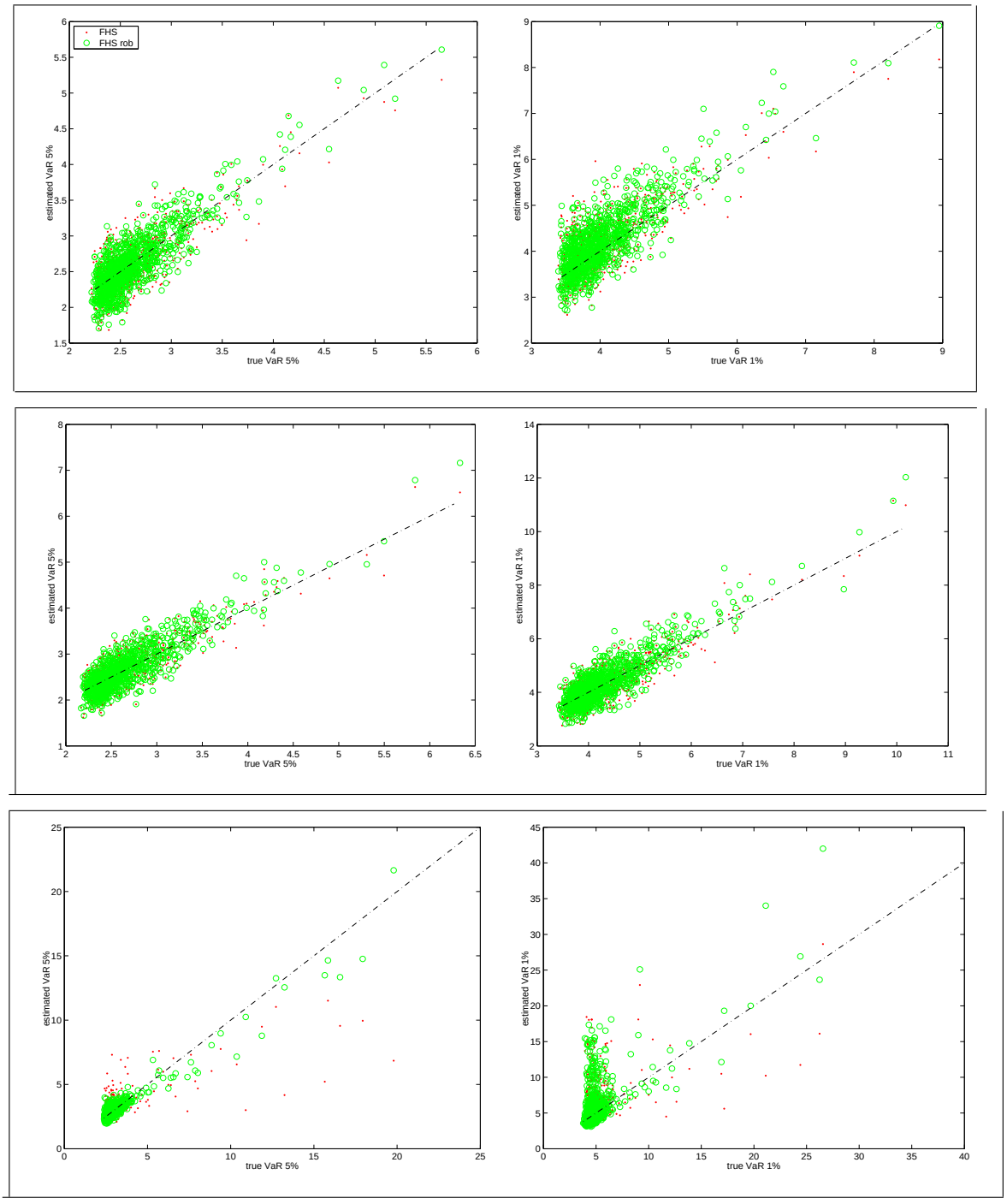

Figure 2.2: True VaR's versus estimated VaR's for horizon $h = 10$ days and confidence levels 5%, 1% under Student t_5 innovation (first panel), Laplace innovation (second panel) and replace-innovative model (third panel).

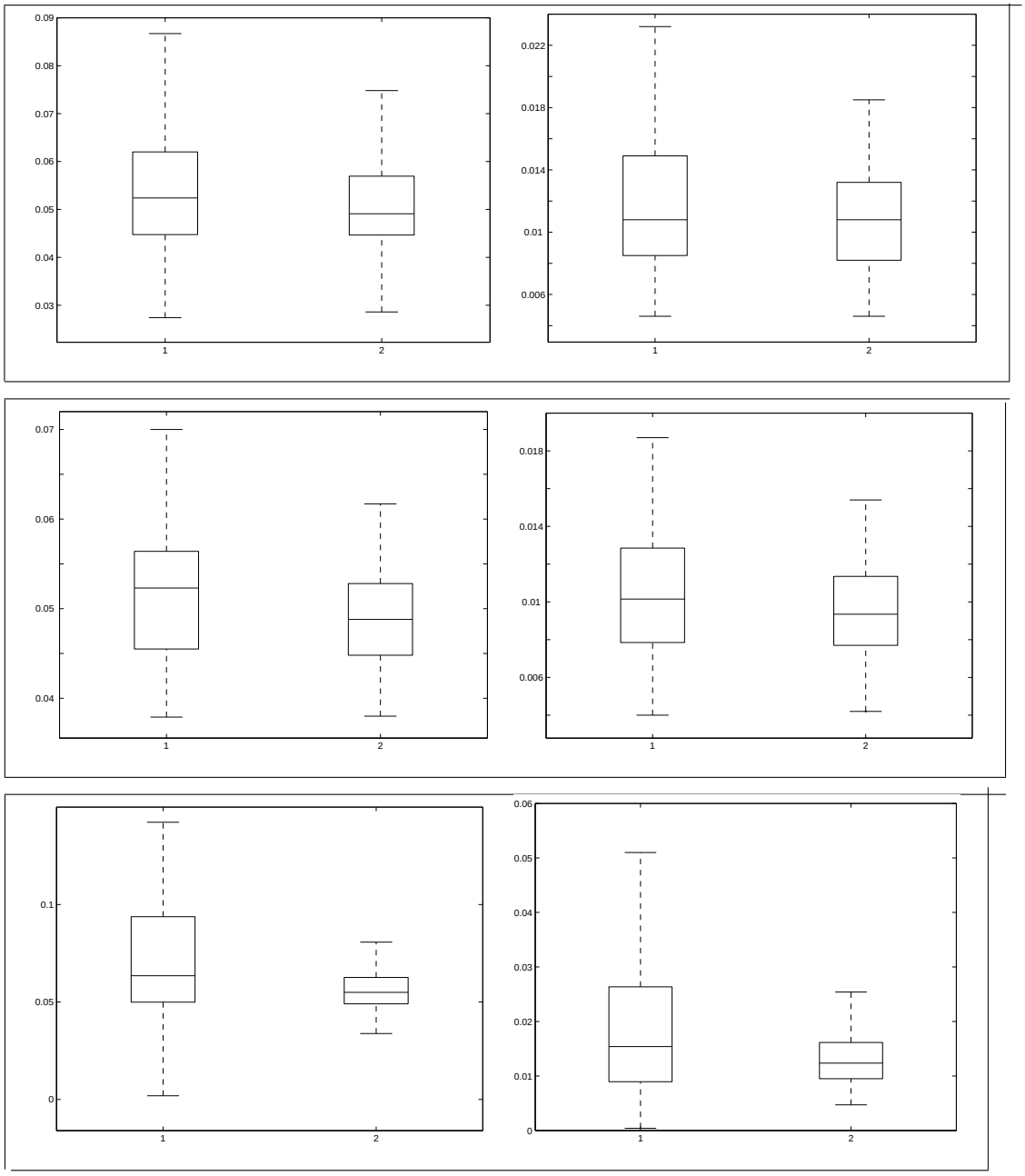

Figure 2.3: Actual VaR levels for horizon $h = 1$ day and nominal confidence levels 5%, 1% for classical (column 1) and robust (column 2) FHS method under Student t_5 innovation (first panel) Laplace innovation (second panel) and replace-innovative model (third panel) when the volatility is "high".

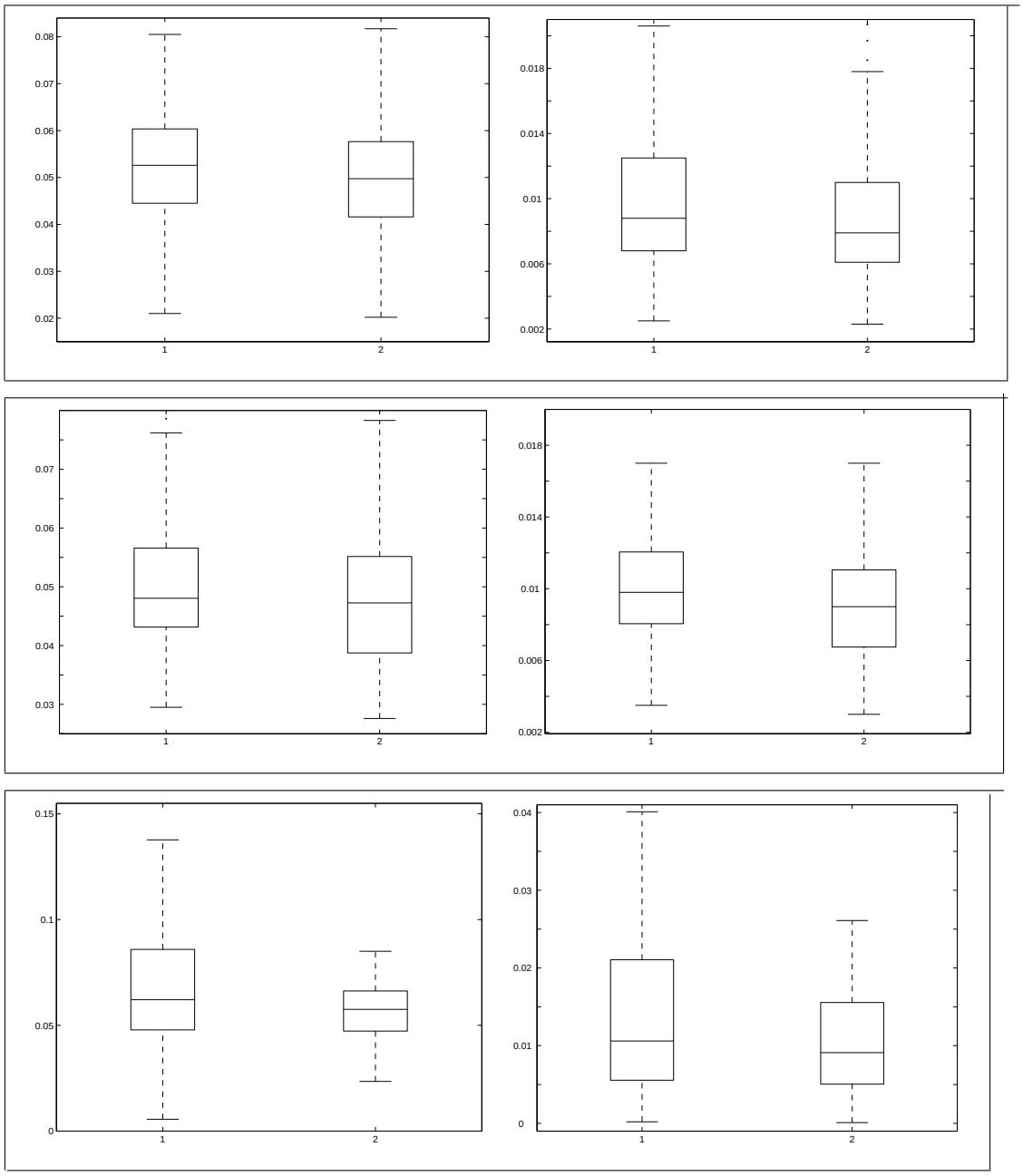

Figure 2.4: Actual VaR levels for horizon $h = 10$ days and nominal confidence levels 5%, 1% for classical (column 1) and robust (column 2) FHS method under Student t_5 innovation (first panel) Laplace innovation (second panel) and replace-innovative model (third panel) when the volatility is "high".

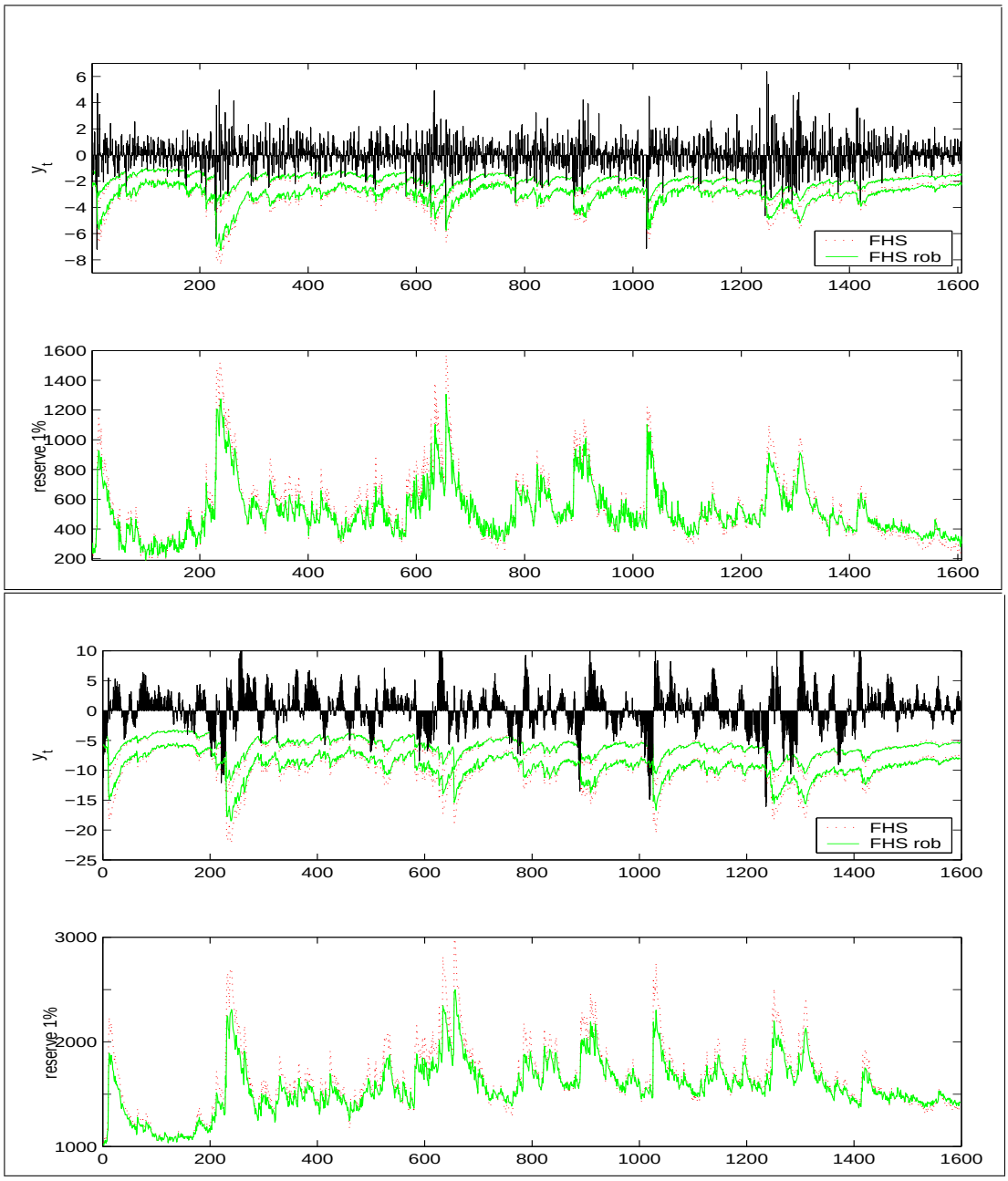

Figure 2.5: Backtesting on the Dow Jones index. Out of sample estimates of VaR's at 5%, 1% confidence levels (superimposed on the rate of returns in percentage) and reserve amounts for horizons $h = 1$ day (first panel) and $h = 10$ days (second panel) using the classical and robust FHS method.

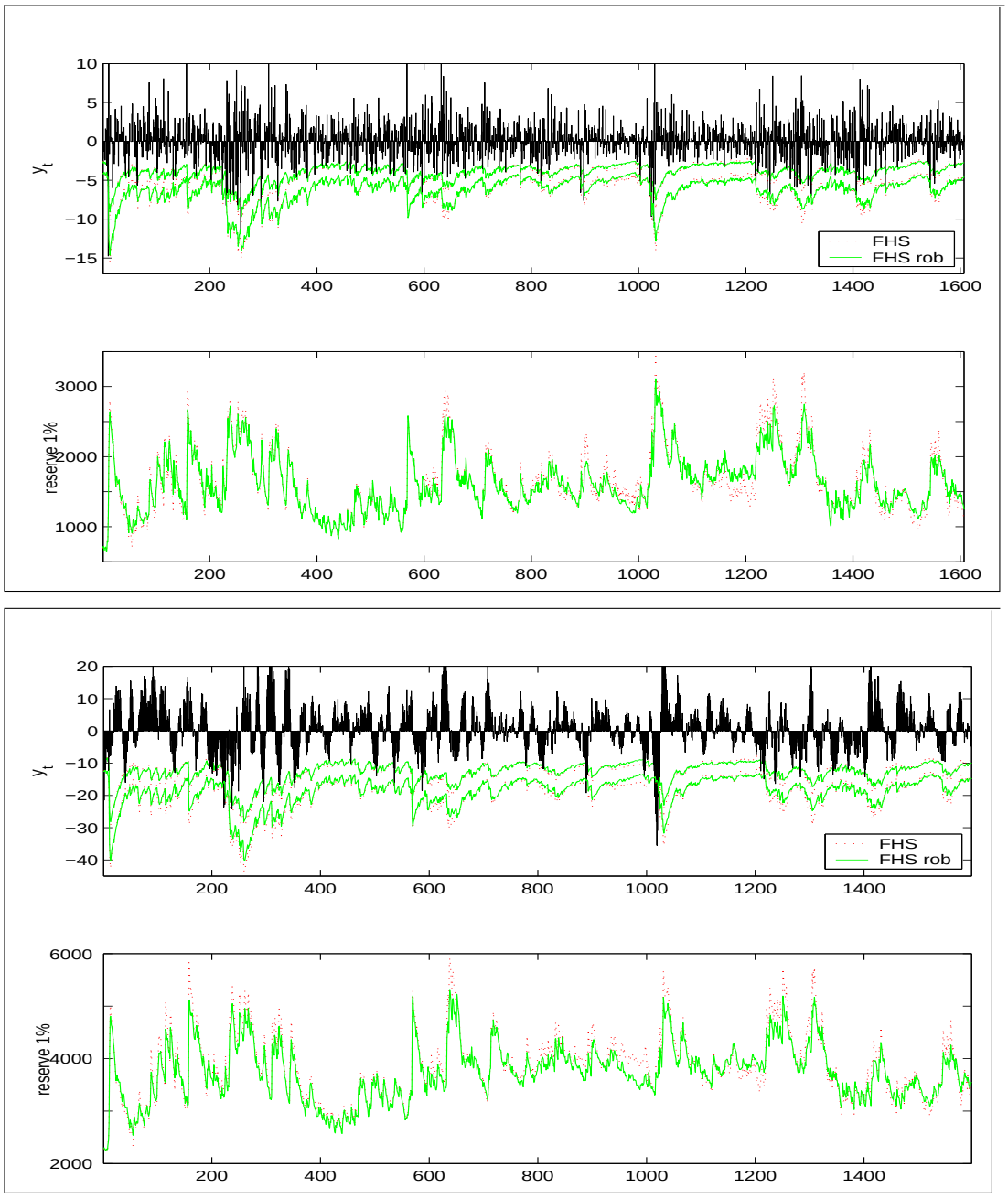

Figure 2.6: Backtesting on the BMW share price. Out of sample estimates of VaR's at 5%, 1% confidence levels (superimposed on the rate of returns in percentage) and reserve amounts for horizons $h = 1$ day (first panel) and $h = 10$ days (second panel) using the classical and robust FHS method.

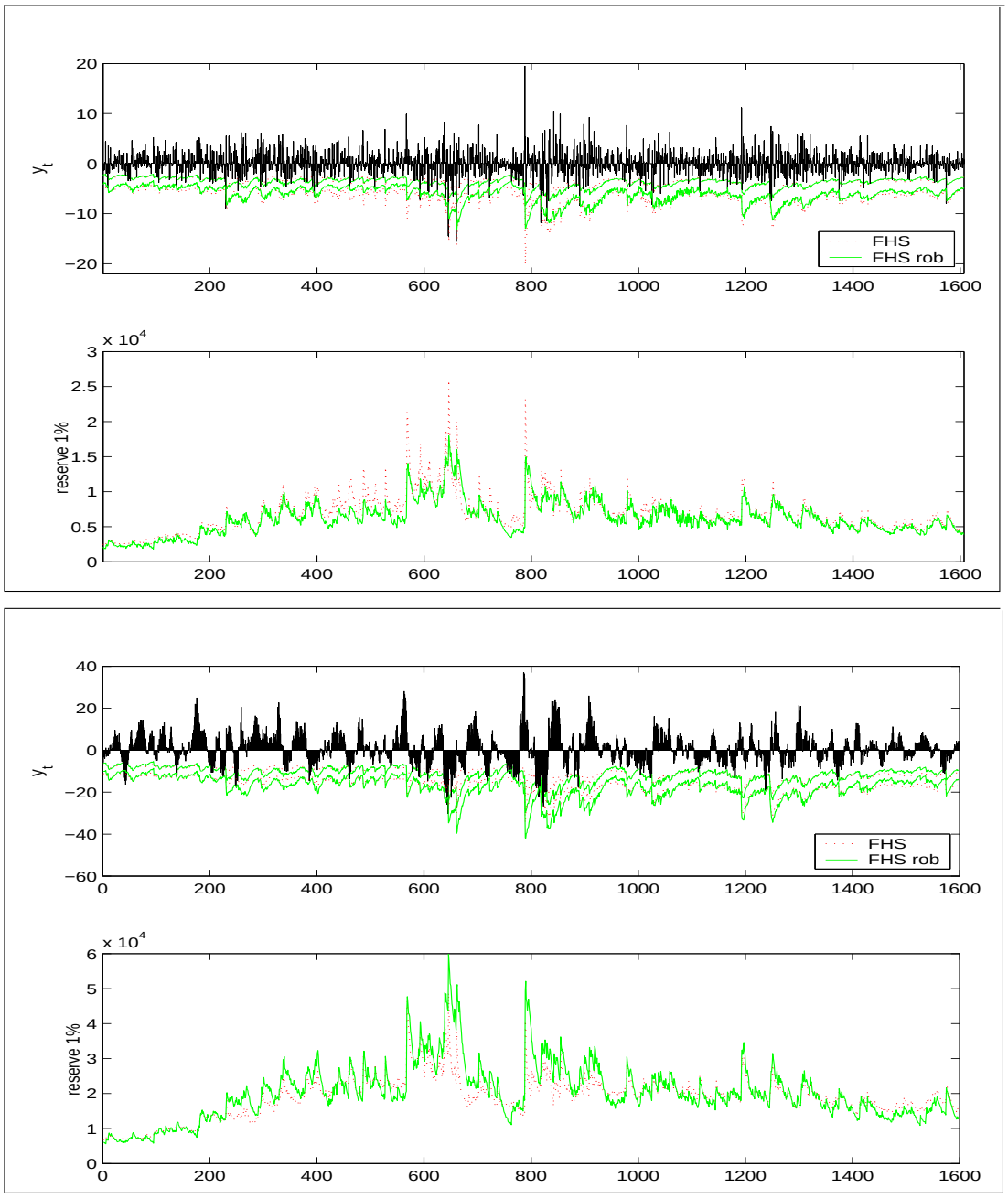

Figure 2.7: Backtesting on the Microsoft share price. Out of sample estimates of VaR's at 5%, 1% confidence levels (superimposed on the rate of returns in percentage) and reserve amounts for horizons $h = 1$ day (first panel) and $h = 10$ days (second panel) using the classical and robust FHS method.

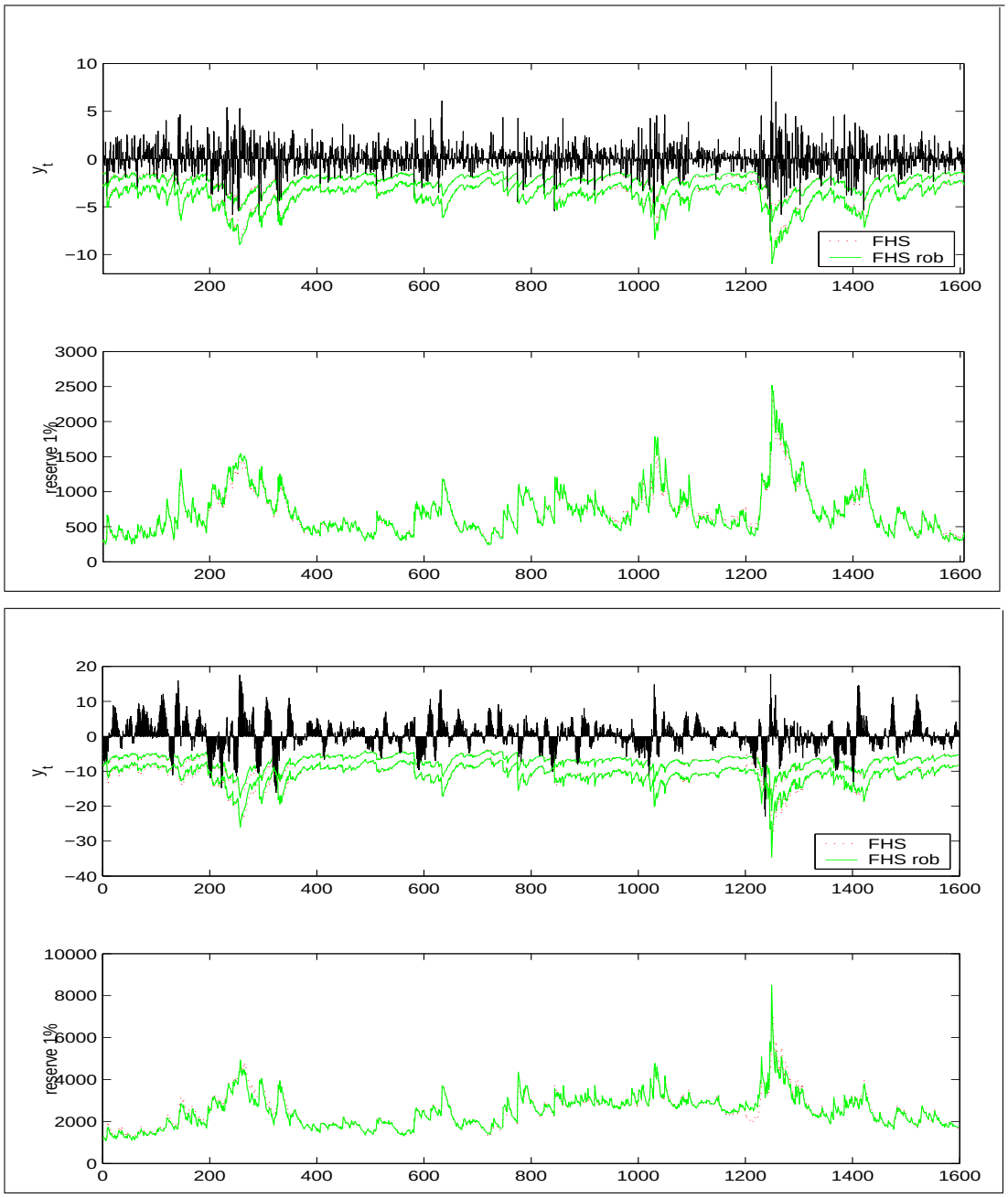

Figure 2.8: Backtesting on the Nestle share price. Out of sample estimates of VaR's at 5%, 1% confidence levels (superimposed on the rate of returns in percentage) and reserve amounts for horizons $h = 1$ day (first panel) and $h = 10$ days (second panel) using the classical and robust FHS method.

Appendix A

Proofs of Propositions

A.1 Proofs of Proposition 1.1 and Corollary 1.1

The proofs of Proposition 1.1 and Corollary 1.1 follow from Theorem 1 and Corollary 1 in Stefanski, Carroll and Ruppert (1986).

To prove Proposition 1.1, let ψ be any competitor with ψ^{bif}. Without loss of generality assume that $\psi = IF_{\psi}^{cond}$, i.e. ψ is in canonical form in the sense of Hampel (1974). Hence, $D(\psi;\theta_0) = I$ and

$$E_{\theta_0}[\psi(y_1^m;\theta_0)\, s(y_1^m;\theta_0)^\top] = I \tag{A.1}$$

as differentiating (1.17) with respect to θ gives $-D(\psi;\theta_0)+E_{\theta_0}[\psi(y_1^m;\theta_0)\, s(y_1^m;\theta_0)^\top] = 0$ assuming that integration and differentiation can be interchanged. Moreover, $V(\psi;\theta_0) = E_{\theta_0}[\psi(y_1^m;\theta_0)\,\psi(y_1^m;\theta_0)^\top]$. Write s for $s(y_1^m;\theta_0)$, τ for $\tau(y_1^{m-1};\theta_0)$, ψ for $\psi(y_1^m;\theta_0)$, V_ψ for $V(\psi;\theta_0)$, D_{bif} for $D(\psi^{bif};\theta_0)$ and V_{bif} for $V(\psi^{bif};\theta_0)$.

Then

$$\begin{aligned}
&\min_{\psi} tr(V_{\psi}\, V_{bif}^{-1}) \\
&\quad = \min_{\psi} tr(E_{\theta_0}[\psi\,\psi^{\top}]\, V_{bif}^{-1}) \\
&\quad = \min_{\psi} tr(E_{\theta_0}[(D_{bif}^{-1}(s-\tau)-\psi)\ (D_{bif}^{-1}(s-\tau)-\psi)^{\top}]\, V_{bif}^{-1}) \\
&\quad = \min_{\psi} E_{\theta_0}[(D_{bif}^{-1}(s-\tau)-\psi)^{\top}\ V_{bif}^{-1}\ (D_{bif}^{-1}(s-\tau)-\psi)] \\
&\quad = \min_{\psi} E_{\theta_0}[(V_{bif}^{-1/2} D_{bif}^{-1}(s-\tau) - V_{bif}^{-1/2}\,\psi)^{\top}\ (V_{bif}^{-1/2} D_{bif}^{-1}(s-\tau) - V_{bif}^{-1/2}\,\psi)], \qquad \text{(A.2)}
\end{aligned}$$

where, using (1.16) and (A.1), the second equality follows from

$$E_{\theta_0}[(D_{bif}^{-1}(s-\tau)-\psi)\ (D_{bif}^{-1}(s-\tau)-\psi)^{\top}] = D_{bif}^{-1} E_{\theta_0}[(s-\tau)\ (s-\tau)^{\top}] D_{bif}^{-\top} - D_{bif}^{-1} - D_{bif}^{-\top} + V_{\psi},$$

and the first three terms in the right hand side are independent of ψ. Under condition (1.23), the objective function (A.2) is minimized by

$$\psi^{*} = D_{bif}^{-1}(s-\tau)\ \min(1, c\,(\ (s-\tau)^{\top}\, D_{bif}^{-\top}\, V_{bif}^{-1}\, D_{bif}^{-1}\,(s-\tau)\ )^{-1/2})$$

and, as $D_{bif}^{-\top}\, V_{bif}^{-1}\, D_{bif}^{-1} = E_{\theta_0}[\psi^{bif}\,\psi^{bif\top}]^{-1} = A(\theta_0)^{\top} A(\theta_0)$, $\psi^{*} = D_{bif}^{-1}\,\psi^{bif}$. Condition (A.1) ensures that ψ^{*} is unique almost surely. □

To prove Corollary 1.1, again assume that all score functions are in canonical form and satisfy (1.16). Define

$$\mathcal{S} := \{\psi : \sup_{z\in\mathbb{R}^m} \psi^{\top} V_{\psi}^{-1} \psi \le c^2\}, \quad \mathcal{S}_{bif} := \{\psi : \sup_{z\in\mathbb{R}^m} \psi^{\top}\, V_{bif}^{-1}\, \psi \le c^2\}.$$

We must show that if there exists ψ^{opt} in $\mathcal{S}$ such that $V_{\psi^{opt}} \le V_{\psi}$ (in the strong sense of positive-definiteness) for all ψ in $\mathcal{S}$, then ψ^{opt} is equivalent to $D_{bif}^{-1}\,\psi^{bif}$.

Clearly, $D_{bif}^{-1}\,\psi^{bif}$ is in $\mathcal{S}$; thus by assumption $V_{\psi^{opt}} \le V_{bif}$ (and $V_{\psi^{opt}}^{-1} \ge V_{bif}^{-1}$). From this follows that

$$\psi^{opt\top}\, V_{bif}^{-1}\, \psi^{opt} \le \psi^{opt\top}\, V_{\psi^{opt}}^{-1}\, \psi^{opt} \le c^2,$$

and hence ψ^{opt} is in $\mathcal{S}_{bif}$. Let $\mathcal{I} = \mathcal{S} \cap \mathcal{S}_{bif}$. The set $\mathcal{I}$ is nonempty; it contains $D_{bif}^{-1}\,\psi^{bif}$ and ψ^{opt}. For any ψ in $\mathcal{I}$ we know $V_{\psi^{opt}} \leq V_\psi$ and hence

$$tr(V_{\psi^{opt}}\,V_{bif}^{-1}) \leq tr(V_\psi\,V_{bif}^{-1})$$

for all ψ in $\mathcal{I}$. But Proposition 1.1 proves that $D_{bif}^{-1}\,\psi^{bif}$, when defined, is the almost everywhere unique minimizer of $tr(V_\psi\,V_{bif}^{-1})$ among all ψ in $\mathcal{I}$. Hence, the equivalence of ψ^{opt} and ψ^{bif} follows. □

A.2 Proof of Proposition 1.2

This appendix describes the computation of the correction factor $\tau(y_1^{m-1};\theta_0)$ implicitly defined by equation (1.22).

Recall that, according to (1.26), the Gaussian score function can be written as

$$s(y_1^m;\theta_0) = -k_{1,m} + k_{2,m}u_m(\theta_0) + k_{1,m}u_m^2(\theta_0),$$

where $u_m(\theta_0) \sim \mathcal{N}(0,1)$ under the reference model $\mathbb{P}_{\theta_0}^m$. We write A for $A(\theta_0)$.

Formally, the problem is to compute $\tau(y_1^{m-1};\theta_0)$ such that

$$0 = A\int_{-\infty}^{+\infty}\left(-k_{1,m} + k_{2,m}\,u + k_{1,m}\,u^2 - \tau(y_1^{m-1};\theta_0)\right)w(y_1^{m-1},\mu_m(\theta_0)+\sigma_m(\theta_0)\,u;\theta_0)\,d\Phi(u).$$

As $\tau(y_1^{m-1};\theta_0)$, $k_{1,m}$ and $k_{2,m}$ are $\mathcal{F}_{m-1}$-measurable,

$$\tau(y_1^{m-1};\theta_0) := \frac{\tau_{num}(y_1^{m-1};\theta_0)}{\tau_{den}(y_1^{m-1};\theta_0)},$$

where

$$\tau_{num}(y_1^{m-1};\theta_0) := \int_{-\infty}^{+\infty}\left(-k_{1,m} + k_{2,m}u + k_{1,m}u^2\right)w(y_1^{m-1},\mu_m(\theta_0)+\sigma_m(\theta_0)\,u;\theta_0)\,d\Phi(u) \quad \text{(A.3)}$$

and

$$\tau_{den}(y_1^{m-1};\theta_0) := \int_{-\infty}^{+\infty} w(y_1^{m-1},\mu_m(\theta_0)+\sigma_m(\theta_0)\,u;\theta_0)\,d\Phi(u). \quad \text{(A.4)}$$

Clearly, the difficult part in the computation of these integrals is the weighting function $w(y_1^m;\theta_0)$, defined by (1.20). However, as the weighting function implies that $\|\psi_c(y_1^m;\theta_0)\|^2 \leq c^2$, we can equivalently express such an inequality in terms of 'admissible' values of the standardized innovation $u_m(\theta_0)$. Specifically, we compute $\tau(y_1^{m-1};\theta_0)$ by means of the following two steps procedure.

Step 1

In the first step we compute the real roots in the real variable $u_m(\theta_0)$ of the quartic equation (1.26), i.e.

$$\begin{aligned} 0 &= \|A\,(s(y_1^m;\theta_0) - \tau^{(0)})\|^2 - c^2 \\ &:= \|A\,(-k_{1,m} + k_{2,m}\,u_m(\theta_0) + k_{1,m}\,u_m^2(\theta_0) - \tau^{(0)})\|^2 - c^2 \\ &:= a_4 u_m^4(\theta_0) + a_3 u_m^3(\theta_0) + a_2 u_m^2(\theta_0) + a_1 u_m(\theta_0) + a_0 - c^2, \end{aligned}$$

where

$$a_4 := k_{1,m}^\top A^\top A k_{1,m}, \quad a_3 := 2k_{1,m}^\top A^\top A k_{2,m},$$

$$a_2 := k_{2,m}^\top A^\top A k_{2,m} - 2k_{1,m}^\top A^\top A k_{1,m} - 2k_{1,m}^\top A^\top A \tau^{(0)},$$

$$a_1 := -a_3 - 2k_{2,m}^\top A^\top A \tau^{(0)},$$

$$a_0 := a_4 + 2k_{1,m}^\top A^\top A \tau^{(0)} + \tau^{(0)\top} A^\top A \tau^{(0)}.$$

Recall that existence of a solution is guaranteed by Lemma 2.1 in Künsch et al. (1989) when choosing $c \geq \sqrt{p}$. In general, we have either two or four real roots. As $a_4 > 0$, in the first case

$$\begin{aligned} \|A\,(s(y_1^m;\theta_0) - \tau^{(0)})\| &\leq c, \quad u_m(\theta_0) \in [\underline{u}_m, \overline{u}_m], \\ &> c, \quad u_m(\theta_0) \in (-\infty, \underline{u}_m) \cup (\overline{u}_m, +\infty) \end{aligned}$$

denoting by $\underline{u}_m \leq \overline{u}_m$ the real roots. In the second case, with real roots $\underline{\underline{u}}_m \leq \underline{u}_m \leq \overline{u}_m \leq \overline{\overline{u}}_m$,

$$\begin{aligned} \|A\,(s(y_1^m;\theta_0) - \tau^{(0)})\| &\leq c, \quad u_m(\theta_0) \in [\underline{\underline{u}}_m, \underline{u}_m] \cup [\overline{u}_m, \overline{\overline{u}}_m] \\ &> c, \quad u_m(\theta_0) \in (-\infty, \underline{\underline{u}}_m) \cup (\underline{u}_m, \overline{u}_m) \cup (\overline{\overline{u}}_m, +\infty). \end{aligned}$$

In almost all simulations and all empirical estimations there were two real roots.

Step 2

In the second step we 'split' the integral in equations (A.3) and (A.4) according to the roots determined in Step 1. Assume first that there are two real roots, then

$$
\begin{aligned}
&\tau_{num}(y_1^{m-1};\theta_0) \\
&= \int_{-\infty}^{\underline{u}_m} \overbrace{\left(-k_{1,m}+k_{2,m}u+k_{1,m}u^2\right)\frac{c}{\|A\left(s(\upsilon;\theta_0)-\tau^{(0)}\right)\|}}^{q_n(u):=} d\Phi(u) \\
&\quad + \int_{\underline{u}_m}^{\overline{u}_m} \left(-k_{1,m}+k_{2,m}u+k_{1,m}u^2\right) d\Phi(u) \\
&\quad + \int_{\overline{u}_m}^{+\infty} \left(-k_{1,m}+k_{2,m}u+k_{1,m}u^2\right)\frac{c}{\|A\left(s(\upsilon;\theta_0)-\tau^{(0)}\right)\|} d\Phi(u) \\
&:= \int_{-\infty}^{\underline{u}_m} q_n(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du - k_{1,m}\left[\Phi(\overline{u}_m)-\Phi(\underline{u}_m)\right] + k_{2,m}M_{1,m} + k_{1,m}M_{2,m} \\
&\quad + \int_{\overline{u}_m}^{+\infty} q_n(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du.
\end{aligned}
$$

Notice that $q_n : \mathbb{R} \longrightarrow \mathbb{R}^p$ with the same functional form in each component. We recall that $M_{1,m}$, $M_{2,m}$ are defined in Proposition 1.2 and $\upsilon := (y_1, \ldots, y_{m-1}, \mu_m(\theta_0) + \sigma_m(\theta_0)\, u)$.

Notice that $M_{1,\cdot}$ and $M_{2,\cdot}$ are the truncated first and second moment of a standard Gaussian random variable and integration by parts yields

$$
M_{1,\cdot} := \int_a^b u\, d\Phi(u) = \phi(a) - \phi(b), \quad M_{2,\cdot} := \int_a^b u^2\, d\Phi(u) = a\,\phi(a) - b\,\phi(b) + \Phi(b) - \Phi(a).
$$

The remaining univariate integrals are approximated 'componentwise' using the Laplace's method. Under standard regularity conditions[1] on the real function $q(\cdot)$, for $\alpha \to \infty$

$$
\begin{aligned}
\int_0^\infty \alpha\exp(-\alpha u)\; q(u)\, du &= q(0) + \frac{q'(0)}{\alpha} + \frac{q''(0)}{\alpha^2} + O\left(\frac{1}{\alpha^3}\right) \\
&=: \mathcal{L}(q,\alpha) + O\left(\frac{1}{\alpha^3}\right)
\end{aligned}
$$

[1] See, for instance, Jensen (1995), p. 58.

by iterating integration by parts. $\mathcal{L}(q,\alpha)$ is the Laplace's approximation of the integral up to the third order[2]. Therefore,

$$\begin{aligned}
&\int_{\overline{u}_m}^{+\infty} q_n(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du \\
&= \frac{1}{\sqrt{2\pi}}\exp(-0.5\overline{u}_m^2)\frac{1}{\overline{u}_m}\int_0^{+\infty}\overline{u}_m\exp(-\overline{u}_m z)\;q_n(\overline{u}_m+z)\exp(-0.5z^2)\,dz \\
&= \frac{1}{\sqrt{2\pi}}\exp(-0.5\overline{u}_m^2)\frac{1}{\overline{u}_m}\left(\mathcal{L}(\overline{q}_n,\overline{u}_m)+O\left(\frac{1}{\overline{u}_m^3}\right)\right) \\
&=: L_n(\overline{u}_m)+O\left(\frac{1}{\overline{u}_m^3}\right),
\end{aligned}$$

where the first equality follows from the substitution $z=u-\overline{u}_m$ and $\overline{q}_n(z):=q_n(\overline{u}_m+z)\exp(-0.5z^2)$. Similarly,

$$\begin{aligned}
\int_{-\infty}^{\underline{u}_m} q_n(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du &= -\frac{1}{\sqrt{2\pi}}\exp(-0.5\underline{u}_m^2)\frac{1}{\underline{u}_m}\left(\mathcal{L}(\underline{q}_n,\underline{u}_m)+O\left(\frac{1}{\underline{u}_m^3}\right)\right) \\
&=: -L_n(\underline{u}_m)+O\left(\frac{1}{\underline{u}_m^3}\right),
\end{aligned}$$

where $\underline{q}_n(z):=q_n(\underline{u}_m+z)\exp(-0.5z^2)$ and using the substitution $z=u-\underline{u}_m$.

The procedure for computing the denominator of τ in (A.4) is completely analogous. Specifically,

$$\begin{aligned}
&\tau_{den}(y_1^{m-1};\theta_0) \\
&= \int_{-\infty}^{\underline{u}_m}\frac{c}{\|A\left(s(\upsilon;\theta_0)-\tau^{(0)}\right)\|}\,d\Phi(u)+\int_{\underline{u}_m}^{\overline{u}_m}d\Phi(u)+\int_{\overline{u}_m}^{+\infty}\overbrace{\frac{c}{\|A\left(s(\upsilon;\theta_0)-\tau^{(0)}\right)\|}}^{q_d(u):=}\,d\Phi(u) \\
&= \int_{-\infty}^{\underline{u}_m} q_d(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du+[\Phi(\overline{u}_m)-\Phi(\underline{u}_m)] \\
&\quad +\int_{\overline{u}_m}^{+\infty} q_d(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du.
\end{aligned}$$

The Laplace's approximations of the remaining integrals are

$$\begin{aligned}
\int_{\overline{u}_m}^{+\infty} q_d(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du &= \frac{1}{\sqrt{2\pi}}\exp(-0.5\overline{u}_m^2)\frac{1}{\overline{u}_m}\left(\mathcal{L}(\overline{q}_d,\overline{u}_m)+O\left(\frac{1}{\overline{u}_m^3}\right)\right) \\
&=: L_d(\overline{u}_m)+O\left(\frac{1}{\overline{u}_m^3}\right),
\end{aligned}$$

[2]We use third order Laplace's approximations as the contribution of higher order terms is negligible.

where $\overline{q}_d(z) := q_d(\overline{u}_m + z)\exp(-0.5z^2)$ and

$$\begin{aligned}\int_{-\infty}^{\underline{u}_m} q_d(u)\frac{1}{\sqrt{2\pi}}\exp(-0.5u^2)\,du &= -\frac{1}{\sqrt{2\pi}}\exp(-0.5\underline{u}_m^2)\frac{1}{\underline{u}_m}\left(\mathcal{L}(\underline{q}_d, \underline{u}_m) + O\left(\frac{1}{\underline{u}_m^3}\right)\right) \\ &=: -L_d(\underline{u}_m) + O\left(\frac{1}{\underline{u}_m^3}\right),\end{aligned}$$

where $\underline{q}_d(z) := q_d(\underline{u}_m + z)\exp(-0.5z^2)$. Collecting the terms one yields the expression for τ stated in Proposition 1.2.

In the general case where the quartic equation (1.26) has four real roots $\underline{\underline{u}}_m \leq \underline{u}_m \leq \overline{u}_m \leq \overline{\overline{u}}_m$, for instance the integral in (A.4) can be written as follows

$$\begin{aligned}&\tau_{den}(y_1^{m-1};\theta_0) \\ &= \int_{-\infty}^{\underline{\underline{u}}_m} \frac{c}{\|A\left(s(\upsilon;\theta_0) - \tau^{(0)}\right)\|}\,d\Phi(u) + \int_{\underline{\underline{u}}_m}^{\underline{u}_m} d\Phi(u) + \int_{\underline{u}_m}^{\overline{u}_m} \frac{c}{\|A\left(s(\upsilon;\theta_0) - \tau^{(0)}\right)\|}\,d\Phi(u) \\ &\quad + \int_{\overline{u}_m}^{\overline{\overline{u}}_m} d\Phi(u) + \int_{\overline{\overline{u}}_m}^{+\infty} \frac{c}{\|A\left(s(\upsilon;\theta_0) - \tau^{(0)}\right)\|}\,d\Phi(u)\end{aligned}$$

and Laplace's approximations hold only for the first and the last integral. However, numerical results (not reported here) show that the error when neglecting the weighting function in the central integral is very small. When $\underline{u}_m$ and $\overline{u}_m$ are close enough and/or $\|A\left(s(\upsilon;\theta_0) - \tau^{(0)}\right)\|$ is not very large, the error is essentially zero. Finally, the implementation of such a case (which happens very rarely) in the algorithm would largely increase the computational time, without improving the inference results.

A.3 Proof of Proposition 1.3

We prove the statement of the proposition only for the Wald statistics. The proof is extracted from Heritier and Ronchetti (1994) and Ronchetti and Trojani (2001). Those for the other tests can be proved similarly.

Under regularity conditions (cf. Clarke (1986) and Heritier and Ronchetti (1994)), $\overline{a}$ is Fréchet differentiable and this implies the Fréchet differentiability of U. Then, using (1.31) and a first order

von Mises (1947) expansion of U gives, up to terms of order $o(\epsilon)$,

$$\sqrt{n}\left(U(\mathbb{P}_n^m) - U(\mathbb{P}_{\eta(\epsilon,n)}^m)\right) \to \mathcal{N}(0, I_r), \quad n \to \infty,$$

in distribution uniformly for all $\mathbb{G} \in dom(U)$. Hence, the test statistic (1.30) is asymptotically $\chi^2(r; \beta(\epsilon))$ with $\beta(\epsilon) = n\, U(\mathbb{P}_{\eta(\epsilon,n)}^m)^\top U(\mathbb{P}_{\eta(\epsilon,n)}^m)$ and up to terms of order $O(1/n)$

$$\alpha(\mathbb{P}_{\eta(\epsilon,n)}^m) = 1 - H_r(q_{1-\alpha_0}; \beta(\epsilon)).$$

Let $b(\epsilon) := -H_r(q_{1-\alpha_0}; \beta(\epsilon))$ and $\mu := -[(\partial/\partial\beta)H_r(q_{1-\alpha_0}; \beta)]_{\beta=0}$. Then, up to terms of order $O(1/n)$,

$$\alpha(\mathbb{P}_{\eta(\epsilon,n)}^m) - \alpha_0 = b(\epsilon) - b(0) = \epsilon\, b'(0) + \frac{1}{2}\,\epsilon^2 b''(0) + o(\epsilon^2). \tag{A.5}$$

Now,

$$b'(0) = \mu\, \frac{\partial \beta(\epsilon)}{\partial \epsilon}\bigg|_{\epsilon=0} = 2\,\mu\, n \left[\frac{\partial U(\mathbb{P}_{\eta(\epsilon,n)}^m)^\top}{\partial \epsilon}\right]_{\epsilon=0} U(\mathbb{P}_{\theta_0}^m) = 0,$$

as $U(\mathbb{P}_{\theta_0}^m) = 0$ under the null hypothesis (1.29) and

$$b''(0) = \mu\, \frac{\partial^2 \beta(\epsilon)}{\partial \epsilon^2}\bigg|_{\epsilon=0} = 2\,\mu\, n \left[\frac{\partial U(\mathbb{P}_{\eta(\epsilon,n)}^m)^\top}{\partial \epsilon}\, \frac{\partial U(\mathbb{P}_{\eta(\epsilon,n)}^m)}{\partial \epsilon}\right]_{\epsilon=0} = 2\,\mu \left\| \int_{\mathbb{R}^m} IF(z; U(\mathbb{P}_{\theta_0}^m))\, d\mathbb{G}(z) \right\|^2$$

and substituting in (A.5), equation (1.32) follows. Write IF^c for IF^{cond}. To end the proof, it suffices to compute the IF of the functional U associated to the Wald test.

$$\begin{aligned}
IF(z; U^W) &= \left[\frac{\partial g(\theta)}{\partial \theta^\top}\, V(\psi_c; \theta)\, \frac{\partial g(\theta)^\top}{\partial \theta}\right]_{\theta=\theta_0}^{-1/2} \frac{\partial g(\theta_0)}{\partial \theta^\top} IF^c(z; \overline{a}) \\
&\equiv \left[\frac{\partial g(\theta)}{\partial \theta^\top} V(\psi_c; \theta)^{1/2} V(\psi_c; \theta)^{1/2} \frac{\partial g(\theta)^\top}{\partial \theta}\right]_{\theta=\theta_0}^{-1/2} \frac{\partial g(\theta_0)}{\partial \theta^\top} V(\psi_c; \theta_0)^{1/2} V(\psi_c; \theta_0)^{-1/2} IF^c(z; \overline{a}) \\
&=: \left[B^\top B\right]^{-1/2} B^\top V(\psi_c; \theta_0)^{-1/2} IF^c(z; \overline{a}),
\end{aligned}$$

with obvious notations and to note that

$$\begin{aligned}
\|IF(z; U^W)\|^2 &= IF^c(z; \overline{a})^\top V(\psi_c; \theta_0)^{-1/2}\; B[B^\top B]^{-1} B^\top\; V(\psi_c; \theta_0)^{-1/2} IF^c(z; \overline{a}) \\
&\leq IF^c(z; \overline{a})^\top V(\psi_c; \theta_0) IF^c(z; \overline{a})
\end{aligned}$$

by the orthogonal projection property of the matrix $B[B^\top B]^{-1} B^\top$. □

Bibliography

[1] Audrino F. and G. Barone-Adesi (2002), "A Multivariate FGD Technique to Improve VaR Computation in Equity Markets", *Computational Management Science*, forthcoming.

[2] Artzner P., F. Delbaen, J. Eber and D. Heath (1999), "Coherent Measures of Risk", *Mathematical Finance*, **9**, 203–228.

[3] Barone-Adesi G., F. Bourgoin and K. Giannopoulos (1998), "Don't Look Back", *Risk*, **11**, 100–103.

[4] Barone-Adesi G., R. Engle and L. Mancini (2004), "GARCH Options in Incomplete Markets", FINRISK working paper.

[5] Barone-Adesi G., K. Giannopoulos and L. Vosper (1999), "VaR Without Correlations for Nonlinear Portfolios", *Journal of Futures Markets*, **19**, 583–602.

[6] ———— (2000), "Filtering Historical Simulation. Backtest Analysis", Working Paper, University of Lugano, Switzerland.

[7] Basle Committee (1996), *Overview of the Amendment of the Capital Accord to Incorporate Market Risk*, Basle Committee on Banking Supervision.

[8] Bednarski T. (1993), "Fréchet Differentiability of Statistical Functionals and Implications to Robust Statistics", in: Morgenthaler S., E. Ronchetti and W.A. Stahel (Ed.), *New Directions in Statistical Data Analysis and Robustness*, Basel: Birkhäuser, 26–34.

[9] Black F. and M. Scholes (1973), "The Valuation of Options and Corporate Liabilities", *Journal of Political Economy*, **81**, 637–654.

[10] Bollerslev T.P. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, **31**, 307–327.

[11] Bollerslev T.P., R.Y. Chou and K.F. Kroner (1992), "ARCH Modelling in Finance: a Selective Review of the Theory and Empirical Evidence", *Journal of Econometrics*, **52**, 5–59.

[12] Bollerslev T.P., R.F. Engle and D.B. Nelson (1994), "ARCH models", in: Engle R.F. and D. McFadden (Ed.), *Handbook of Econometrics*, Vol. IV. North-Holland, Amsterdam, 2959–3038.

[13] Bollerslev T.P. and J.M. Wooldridge (1992), "Quasi-Maximum Likelihood Estimates and Inference in Dynamic Models with Time Varying Covariances", *Econometric Reviews*, **11**, 143–172.

[14] Bustos O. and V.J. Yohai (1986), "Robust Estimates for ARMA Models", *Journal of the American Statistical Association*, **81**, 155–168.

[15] Busse J.A. (1999), "Volatility Timing in Mutual Funds: Evidence from Daily Returns", *Review of Financial Studies*, **12**, 1009–1041.

[16] Cantoni E. and E. Ronchetti (2001), "Robust Inference For Generalized Linear Models", *Journal of the American Statistical Association*, **96**, 1022–1030.

[17] Chernick M.R., D.J. Downing and D.H. Pike (1982), "Detecting Outliers in Time Series Data", *Journal of the American Statistical Association*, **77**, 749–747.

[18] Christoffersen P., F. Diebold and T. Schuermann (1998), "Horizon Problems and Extreme Events in Financial Risk Management", *Economic Policy Review, Federal Reserve Bank of New York*, October, 109–118.

[19] Clarke B.R. (1983), "Uniqueness and Fréchet Differentiability of Functional Solutions to Maximum Likelihood-Type Equations", *Annals of Statistics*, **11**, 1196–1205.

[20] ———— (1986), "Nonsmooth Analysis and Fréchet Differentiability of M-Functionals", *Probability Theory and Related Fields*, **73**, 197–209.

[21] Cont R. (2001), "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues", *Quantitative Finance*, **1**, 223–236.

[22] Danielsson J. and C. de Vries (1997), "Tail Index and Quantile Estimation with Very High Frequency Data", *Journal of Empirical Finance*, **4**, 241–257.

[23] Dell'Aquila R., E. Ronchetti and F. Trojani (2003), "Robust GMM Analysis of Models for the Short Rate Process", *Journal of Empirical Finance*, **10**, 373–397.

[24] Diebold F.X., Schuermann T. and Stroughair J. (1998), "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management", in: A.-P. N. Refenes, J.D. Moody and A.N. Burgess (Ed.), *Advances in Computational Finance*, Kluwer Academic Publishers, Amsterdam, 3–12.

[25] Duffie D. and J. Pan (1997), "An Overview of Value at Risk", *Journal of Derivatives*, **4**, 7–49.

[26] Engle R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimate of the Variance of United Kingdom Inflation", *Econometrica*, **50**, 987–1007.

[27] ———— (1995), *ARCH: Selected Readings*, Oxford University Press, Oxford, UK.

[28] Engle R.F., D.M. Lilien and R.P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica*, **55**, 391–407.

[29] Fama E. (1965), "The Behaviour of Stock Prices", *Journal of Business*, **38**, 34–105.

[30] Fernholz L.T. (1983), *Von Mises Calculus for Statistical Functionals*, Springer, New York.

[31] Ferson W.E. and A.F. Siegel (2001), "The Efficient Use of Conditioning Information in Portfolios", *Journal of Finance*, **56**, 967–982.

[32] Fleming J., C. Kirby and B. Ostdiek (2001), "The Economic Value of Volatility Timing", *Journal of Finance*, **56**, 329–352.

[33] ———— (2003), "The Economic Value of Volatility Timing using 'Realized' Volatility", *Journal of Financial Economics*, **67**, 473–509.

[34] Foresi S. and F. Peracchi (1995), "The Conditional Distribution of Excess Returns: An Empirical Analysis", *Journal of the American Statistical Association*, **90**, 451–466.

[35] Gasko M. and J.L. Rosenberger (1983), "Comparing Location Estimators: Means, Medians, and Trimean", in: Hoaglin D.C., F. Mosteller and J.W. Tukey (Ed.), *Understanding Robust and Exploratory Data Analysis*, Wiley, New York, 297–338.

[36] Genton M.G. and E. Ronchetti (2003), "Robust Indirect Inference", *Journal of the American Statistical Association*, **98**, 67–76.

[37] Glosten L.R., R. Jagannathan and D.E. Runkle (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*, **48**, 1779–1801.

[38] Gourieroux C. and A. Monfort (1989), "A General Framework for Testing a Null Hypothesis in a Mixed Form", *Econometric Theory*, **5**, 63–82.

[39] Gourieroux C., A. Monfort and A. Trognon (1984), "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, **52**, 681–700.

[40] Hampel F.R. (1968), "Contributions to the Theory of Robust Estimation", Ph.D. thesis, University of California, Berkeley.

[41] ———— (1974), "The Influence Curve and its Role in Robust Estimation", *Journal of the American Statistical Association*, **69**, 383–393.

[42] Hampel F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

[43] Heritier S. and E. Ronchetti (1994), "Robust Bounded-Influence Tests in General Parametric Models", *Journal of the American Statistical Association*, **89**, 897–904.

[44] Heston S.L. and S. Nandi (2000), "A Closed-Form GARCH Option Valuation Model", *Review of Financial Studies*, **13**, 585–625.

[45] Huber P.J. (1964), "Robust Estimation of a Location Parameter", *Annals of Mathematical Statistics*, **35**, 73–101.

[46] ———— (1981), *Robust Statistics*, Wiley, New York.

[47] Hull J.C. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance*, **42**, 281–300.

[48] Jensen J.L. (1995), *Saddlepoint Approximations*, Clarendon Press, Oxford.

[49] Koenker R.W. (1982), "Robust Methods in Econometrics", *Econometric Review*, **1**, 213–255.

[50] Koenker R.W. and G. Bassett (1978), "Regression Quantiles", *Econometrica*, **46**, 33–50.

[51] Koenker R.W. and G. Bassett (1982), "Robust Tests for Heteroskedasticity based on Regression Quantiles", *Econometrica*, **50**, 43–61.

[52] Krasker W.S. and R.E. Welsh (1982), "Efficient Bounded-Influence Regression Estimation", *Journal of the American Statistical Association*, **77**, 595–604.

[53] Krishnakumar J. and E. Ronchetti (1997), "Robust Estimators for Simultaneous Equations Models", *Journal of Econometrics*, **78**, 295–314.

[54] Künsch H. (1984), "Infinitesimal Robustness for Autoregressive Processes", *Annals of Statistics*, **12**, 843–863.

[55] Künsch H., L.A. Stefanski and R.J. Carroll (1989), "Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models", *Journal of the American Statistical Association*, **84**, 460–466.

[56] Li C.W. and W.K. Li (1996), "On a Double Threshold Autoregressive Heteroscedastic Time Series Model", *Journal of Applied Econometrics*, **11**, 253–274.

[57] Mancini L., E. Ronchetti and F. Trojani (2004), "Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models", *Journal of the American Statistical Association*, forthcoming.

[58] Mancini L. and F. Trojani (2004), "Robust Volatility Estimations for VaR Predictions", manuscript.

[59] Mandelbrot B. (1963), "The Variation of Certain Speculative Prices", *Journal of Business*, **36**, 394–419.

[60] Markatou M. and E. Ronchetti (1997), "Robust Inference: The Approach Based on Influence Functions", in: Maddala G.S. and C.R. Rao (Ed.), *Handbook of Statistics*, vol. 15, North Holland, 49–75.

[61] Martin R.D. and V.J. Yohai (1986), "Influence Functionals for Time Series", *Annals of Statistics*, **14**, 781–818.

[62] McNeil A. and R. Frey (2000), "Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach", *Journal of Empirical Finance*, **7**, 271–300.

[63] Merton R.C. (1973), "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science*, **4**, 141–183.

[64] Muler N. and V.J. Yohai (1999), "Robust Estimates for ARCH Processes", *Journal of Time Series Analysis*, **23**, 341–375.

[65] Müller O., M. Dacarogna, R. Davé, R. Olsen, O. Pictet and J. von Weizsäcker (1997), "Volatilities of Different Time Resolutions—Analyzing the Dynamics of Market Components", *Journal of Empirical Finance*, **4**, 213–240.

[66] Newey W.K. and K.D. West (1987), "A Simple Positive-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica*, **55**, 703–708.

[67] Ortelli C. and F. Trojani (2004), "Robust Efficient Method of Moments", *Journal of Econometrics*, forthcoming.

[68] Pagan A. (1996), "The Econometrics of Financial Markets", *Journal of Empirical Finance*, **3**, 15–102.

[69] Peracchi F. (1990), "Bounded-influence Estimators for the Tobit Model", *Journal of Econometrics*, **44**, 107–126.

[70] ———— (1990), "Robust M-Estimators", *Econometric Reviews*, **9**, 1–30.

[71] ———— (1991), "Bounded-influence Estimators for the SURE Model", *Journal of Econometrics*, **48**, 119–134.

[72] ———— (1991), "Robust M-tests", *Econometric Theory*, **7**, 69–84.

[73] ———— (2002), "On Estimating Conditional Quantiles and Distribution Functions", *Computational Statistics and Data Analysis*, **38**, 433–447.

[74] Poon S.-H. and C.W.J. Granger (2003), "Forecasting Volatility in Financial Markets: A Review", *Journal of Economic Literature*, **41**, 478–539.

[75] Pritsker M. (2001), "The Hidden Dangers of Historical Simulation", Finance and Economics Discussion Series 2001-27. Washington: Board of Governors of the Federal Reserve System.

[76] Riskmetrics (1995), *RiskMetrics Technical Document, 3rd ed.*, J.P. Morgan.

[77] Ronchetti E. (1982), "Robust Testing in Linear Models: the Infinitesimal Approach", Ph.D. thesis, ETH, Zurich.

[78] ———— (1982), "Robust Alternatives to the F-Test for the Linear Model", in: Grossmann W., G.C. Pflug and W. Wertz (Ed.), *Probability and Statistical Inference*, Reidel Publishing Company, 329–342.

[79] ———— (1997), "Robust Inference by Influence Functions", *Journal of Statistical Planning and Inference*, **57**, 59–72.

[80] Ronchetti E. and M.-P. Victoria-Feser (1997), "Robust Estimation for Grouped Data", *Journal of the American Statistical Association*, **92**, 333–340.

[81] Ronchetti E., C. Field and W. Blanchard (1997), "Robust Linear Model Selection by Cross-Validation", *Journal of the American Statistical Association*, **92**, 1017–1023.

[82] Ronchetti E. and A.H. Welsh (1998), "Bias-Calibrated Estimation from Sample Surveys Containing Outliers", *Journal of the Royal Statistical Society, Series B*, **60**, 413–428.

[83] Ronchetti E. and F. Trojani (2001), "Robust Inference with GMM Estimators", *Journal of Econometrics*, **101**, 37–69.

[84] Ruppert D. (1985), "On the Bounded-Influence Regression Estimator of Krasker and Welsch", *Journal of the American Statistical Association*, **80**, 205–208.

[85] Sakata S. and H. White (1998), "High Breakdown Point Conditional Dispersion Estimation with Application to S&P 500 Daily Returns Volatility", *Econometrica*, **66**, 529–567.

[86] Scott L.O. (1987), "Option Pricing When the Variance Changes Randomly: Theory, Estimation, and an Application", *Journal of Financial and Quantitative Analysis*, **22**, 419–438.

[87] Shepard N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility Models", in: Cox D.R., D.V. Hinkley and O.E. Barndorff-Nielsen (Ed.), *Time series Models in Econometrics, Finance and Other Fields*, Chapman and Hall, London, UK, 1–67.

[88] Stefanski L.A., R.J. Carroll and D. Ruppert (1986), "Optimally Bounded Score Functions for Generalized Linear Models with Applications to Logistic Regression", *Biometrika*, **73**, 413–425.

[89] Stein E.M. and J.C. Stein (1991), "Stock Price Distributions with Stochastic Volatility: An Analytic Approach", *Review of Financial Studies*, **4**, 727–752.

[90] Trojani F. and P. Vanini (2002), "A Note on Robustness in Merton's Model of Intertemporal Consumption and Portfolio Choice", *Journal of Economic Dynamics and Control*, **26**, 423–435.

[91] VON MISES R. (1947), "On the Asymptotic Distribution of Differentiable Statistical Functions", *Annals of Mathematical Statistics*, **18**, 309–348.

[92] WIGGINS J.B. (1987), "Option Values under Stochastic Volatilities", *Journal of Financial Economics*, **19**, 351–372.