Robust M-Estimation of Multivariate GARCH Models

Kris Boudt* Christophe Croux
Faculty of Business and Economics, K.U.Leuven, Belgium

Abstract

In empirical work on multivariate financial time series, it is common to postulate a Multivariate GARCH model. We show that the popular Gaussian quasi-maximum likelihood estimator of MGARCH models is very sensitive to outliers in the data. We propose to use robust M-estimators and provide asymptotic theory for M-estimators of MGARCH models. The Monte Carlo study and empirical application document the good robustness properties of the M-estimator with a fat-tailed Student $t$ loss function and volatility models with the property of bounded innovation propagation.

JEL classification: C13; C32; C51

Keywords: GARCH models, M-estimators, multivariate time series, outliers, robust methods

*Correspondence to: Kris Boudt, Faculty of Business and Economics, 69 Naamsestraat, B-3000 Leuven, Belgium. E-mail: Kris.Boudt@econ.kuleuven.be Tel: +32 16 326728 Fax: +32 16 326624
We thank Tim Bollerslev, Geert Dhaene, Sébastien Laurent, Oliver Linton and two anonymous referees for very helpful comments and suggestions. Financial support from the FWO and Research Fund K.U.Leuven is gratefully acknowledged. The usual disclaimer applies.
1 Introduction

It is widely recognized that the volatility and the correlations of daily and weekly financial returns may be time-varying. There is no doubt that incorporating these features into the estimation of the conditional covariance matrix of financial returns can lead to better decisions on portfolio optimization, asset pricing and risk management. One way to do this is to specify the conditional covariance matrix as a deterministic function of the innovations in the mean equation. This approach defines the class of Multivariate GARCH (MGARCH) models reviewed by Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2008). We propose M-estimators and MGARCH volatility forecasting models that are less influenced by outliers in the data than classical MGARCH methods.

It is common to estimate MGARCH models by maximizing a Gaussian likelihood function. Jeantheau (1998) shows that this estimation procedure, called the Gaussian Quasi-Maximum Likelihood (QML) estimation, can produce consistent parameter estimates, even when the true distribution is not Gaussian. This is an important result, since it is common to find that after correcting the returns for the dynamics in the conditional covariance matrix, the distribution of the standardized return series is still heavy tailed. Previous research has interpreted this stylized fact as either a manifestation of a fat
tailed density function such as the multivariate Student $t$ distribution (Fiorentini et al., 2003, and Pesaran and Pesaran, 2007) or the presence of outliers in the data (Franses and Ghijsels, 1999).

Our view is that one must consider both the presence of fat tails in the density function of the innovations and the occurrence of outliers as drivers of the leptokurtosis observed in the standardized residuals of MGARCH models. We do this by proposing a robust estimation method for a wide class of MGARCH models including those with a heavy tailed innovation density function. The method possesses the three features required by Huber (1981) to be robust: (i) it has a reasonably good efficiency at the assumed model, (ii) small deviations from the model assumptions impair the performance only slightly, and (iii) somewhat larger deviations from the model do not cause a catastrophe.

The effect of outliers on the estimation of univariate GARCH models has been studied in the literature and robust estimators have been proposed. One approach consists in pruning iteratively the outliers and fitting the model to the remaining data until no more outliers are detected (Franses and Ghijsels, 1999). Another approach is to use an estimator that is robust to outliers. Sakata and White (1998), Park (2002) and Peng and Yao (2003) estimate the GARCH model by minimizing a robust measure of scale of the residuals. Mancini et al. (2005) and Muler and Yohai (2002,2007) propose a robust M-estimator that assigns a much lower weight to outliers than the Gaussian ML
estimator does. All authors find that in the presence of outliers, the Gaussian QML estimator is very inaccurate and the robust estimators achieve Huber’s requirements.

To our best knowledge, we are the first to study the effect of outliers on the Gaussian QML estimator of *multivariate* GARCH models and to propose robust alternatives. Outliers are the observations that are unlikely to follow the imposed model. We identify outliers as the observations with a large squared Mahalanobis Distance (MD), defined as the inner product of the return standardized by its MGARCH volatility prediction. The standardization is necessary because in the presence of time-varying volatility, an outlying return in periods of low volatility may no longer be outlying in periods of high volatility.

Outliers have a very high influence on the Gaussian QML estimate. We therefore propose to use M-estimators with loss functions for which outliers have little influence on the estimator. Moreover outliers tend to affect future volatility differently than the regular observations. Indeed, Hamilton and Susmel (1994) and Bauwens and Storti (2000) report that often volatility goes up proportionally less after outlying shocks than it does after small and moderate shocks. Nevertheless, in most MGARCH models outlying shocks have the same effect on volatility as small and moderate shocks. To reduce the effect of outliers on the predicted volatility, we advocate in this paper the use of a
bounded innovation propagation MGARCH model, where the effect of outlying returns on volatility predictions is bounded. We show that both a robust loss function and a MGARCH model with bounded innovation propagation are needed for the estimator to be robust to outliers.

The remainder of the paper is organized as follows. In Section 2 we define the class of M-estimators for MGARCH models with elliptical innovations and in Section 3 we provide sufficient conditions under which these M-estimators are consistent and asymptotically normal. In Section 4 we show that in the presence of outliers the M-estimator with a loss function that downweights observations with a large MD will be more accurate than the Gaussian QML estimator. Furthermore, we propose a modification of classical MGARCH models allowing innovations to have a bounded effect on future volatility. This yields the Bounded Innovation Propagation (BIP) model, proposed in the univariate case by Muler and Yohai (2002, 2007). Sections 5 and 6 illustrate the new methodology through a Monte Carlo study and a financial application. We show that our proposal for a robust M-estimator has a fairly high efficiency at the assumed model and that, in contrast with the Gaussian QML estimator, it remains reliable when the data is outlier contaminated. In the empirical application of daily returns on the Nasdaq and NYSE indices, we find that the MGARCH model with bounded influence of news on future volatility captures better the difference in reaction of volatility to the 1987 stock market crash.
and to the burst of the internet bubble in 2000. Section 7 summarizes our conclusions and outlines the implications for further research.

2 M-estimation

2.1 Definitions

We dispose of $T$ realizations of a $N$-dimensional vector stochastic process $\{r_t\}$: $r_1, \ldots, r_T$. We have a model stating that, conditional on the sigma field $I_{t-1}$ generated by the information in the realizations of $\{r_t\}$ until time $t - 1$, the random variable $r_t$ has mean zero and covariance matrix equal to $H_{t,\theta}$

$$r_t|I_{t-1} = H_{t,\theta}^{1/2} u_t \quad \text{with} \quad E[u_t|I_{t-1}] = 0 \quad \text{and} \quad \text{Cov}(u_t|I_{t-1}) = I_N. \quad (1)$$

We assume that $H_{t,\theta}$ is parameterized by a MGARCH model as a measurable function $H_{\theta}(\cdot)$ of the past realizations or $r_t$, with $\theta$ the unknown parameter vector. Under this model, we have that for $t > 2$

$$H_{t,\theta} = H_{\theta}(r_1, \ldots, r_{t-1}) \quad (2)$$

and $H_{1,\theta}$ is initialized by a data-free method. The function $H_{\theta}(\cdot)$ returns a $N \times N$ positive definite, symmetric matrix function. As an example, consider the BEKK parameterization for $H_{t,\theta}$ (Engle and Kroner, 1995):

$$H_{t,\theta} = C'C + \sum_{j=1}^q \sum_{k=1}^K A'_{kj} r_{t-j} r'_{t-j} A_{kj} + \sum_{j=1}^p \sum_{k=1}^K B'_{kj} H_{t-j,\theta} B_{kj}, \quad (3)$$
where $A_{kj}, B_{kj}$ and $C$ are $N \times N$ parameter matrices and $C$ is upper triangular.

As a data-free initialization method for the BEKK model, Lucchetti (2002) proposes to set $H_{1,\theta}$ either to $C'C$ or, if the $H_{t,\theta}$ process is covariance stationary, to its unconditional expectation

$$E[H_{t,\theta}] = (I_N^2 - \sum_{j=1}^{q} \sum_{k=1}^{K} A'_{kj} \otimes A'_{kj} - \sum_{j=1}^{p} \sum_{k=1}^{K} B'_{kj} \otimes B'_{kj})^{-1}\text{vec}(C'C). \quad (4)$$

The operator $\otimes$ denotes the Kronecker product and vec is the operator that transforms a matrix into a vector by stacking the columns of the matrix one underneath the other.

We further require that the $u_t$’s are independent and identically distributed realizations from an elliptically symmetric distribution.\(^1\) This means that we assume that there exists a function $g_* : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that the $N$-dimensional density function of $r_t$ can be written in the form

$$p(r_t|I_{t-1}; \theta) = (\det H_{t,\theta})^{-1/2} g_* (d_{t,\theta}^2), \quad (5)$$

where $d_{t,\theta}^2$ is the inner product of $r_t$ standardized by its MGARCH volatility

$$d_{t,\theta}^2 = (H_{t,\theta}^{-1/2} r_t)'(H_{t,\theta}^{-1/2} r_t) = u_t'u_t. \quad (6)$$

It is common to call $d_{t,\theta}^2$ the squared Mahalanobis Distance (MD) between $r_t$ and zero, measured in terms of its conditional covariance matrix $H_{t,\theta}$. It

\(^1\)While there exist elliptic distributions without finite second moment, we consider here only distributions for which the covariance matrix exists.
measures the outlyingness of a return relatively to its MGARCH volatility prediction. The larger the MD, the more extreme the corresponding return is. Returns with a large MD are suspected to be potential outliers. Under the elliptical model (1)-(5) the squared MD in (6) has density function

\[ f_\star(z) = \frac{\pi^{N/2}}{\Gamma(N/2)} z^{N/2-1} g_\star(z). \]  

(7)

In the special case where \( g_\star(\cdot) \) is the \( N \)-variate normal density function, \( f_\star(\cdot) \) is the chi-squared density function with \( N \) degrees of freedom.

In (5) we assume the function \( g_\star(\cdot) \) to be normalized such that \( H_{t,\theta} \) is the conditional covariance matrix of \( r_t \). For the Gaussian distribution, \( g_\star(z) \) equals the standard multivariate Gaussian density function \( \phi(\cdot) \), while for the Student \( t \) density function with \( \nu > 2 \), the function \( g_\star(\cdot) \) is given by

\[ t_\nu(z) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [\pi(\nu-2)]^{\frac{N}{2}}} \left[ 1 + \frac{z}{\nu-2} \right]^{-\frac{N+\nu}{2}}, \]  

(8)

where \( \Gamma(\cdot) \) is the gamma function. The assumption of elliptical symmetry is standard in the robustness literature (e.g. Bilodeau and Brenner, 1998, Chapter 13) and has also been previously made in the literature on semiparametric analysis of MGARCH models by Hodgson and Vorkink (2003) and Brown and Hodgson (2007). Hodgson et al. (2002) and Hodgson and Vorkink (2003) show that this framework is suitable for the estimation and testing of conditional asset pricing models such as the conditional CAPM.
2.2 M-estimators

Denote \( \theta^* \) the true, unknown parameter vector belonging to the parameter space \( \Theta \) and \( \mathcal{P}_{\theta^*} \) the corresponding model. It is common to estimate \( \theta^* \) by the value of \( \theta \in \Theta \) that maximizes the log-likelihood of the observations \( r_1, ..., r_T \) assuming that the innovations have density function \( g(\cdot) \), which may in fact be different from the true function \( g_*(\cdot) \). This estimation approach yields the QML estimator

\[
\hat{\theta}_{QML} = \arg\max_{\theta \in \Theta} -\frac{1}{T} \sum_{t=1}^{T} \left[ \log \det H_{t,\theta} - 2 \log g \left( r_t' H_{t,\theta}^{-1} r_t \right) \right].
\]

(9)

If \( g(\cdot) = g_*(\cdot) \) the QML estimator is called the ML estimator. The ML and QML estimator belongs to the broader class of M-estimators defined as follows.

**Definition 1** The M-estimate based on a sample \( S_T \) of returns \( r_1, ..., r_T \) is the value of \( \theta \in \Theta \) for which the M-function

\[
M(S_T; \theta, \rho) = \frac{1}{T} \sum_{t=1}^{T} \left[ \log \det H_{t,\theta} + \sigma_* \rho \left( r_t' H_{t,\theta}^{-1} r_t \right) \right],
\]

(10)

is minimized, with \( \rho(\cdot) \) a continuous, positive and nondecreasing scalar function on the real positive numbers.

The scalar \( \sigma_* \) is a correction factor needed to ensure consistency of the M-estimator at the distribution \( g_*(\cdot) \). The function \( \rho(\cdot) \) in the above definition is called the loss function associated to the M-estimator. If \( \rho(\cdot) = -2 \log g(\cdot) \) and \( \sigma_* = 1 \), with \( g(\cdot) \) a specified density function, we obtain a QML estimator.
This choice of $\sigma^*$ only yields consistency for the Gaussian QML estimator. In the next subsection we discuss how to select $\sigma^*$ in an appropriate way. The Gaussian $\phi$ and standardized Student $t_\nu$ density functions are the most popular elliptical density functions used to describe financial return series. The corresponding Gaussian and Student $t_\nu$ ($\nu > 2$) loss functions are given by

$$
\rho_\phi(z) = z ; \quad \rho_{t_\nu}(z) = (N + \nu) \log \left[ 1 + \frac{z^\nu - 2}{\nu - 2} \right].
$$

(11)

### 2.3 Correction factor

The correction factor in the definition of the M-estimator is needed to ensure that the M-estimator estimates the conditional covariance matrix of the returns and not a multiple of the conditional covariance matrix. Without this correction factor, the M-estimate of $H_{t,\theta}$ will still estimate a multivariate measure of dispersion, called the scatter matrix.

This correction factor $\sigma^*$ depends on the true density function $g_\star(\cdot)$ and on the selected loss function. Let $\psi(\cdot) = \rho'(\cdot)$ be the first derivative of the

\footnote{As shown by Linton (1993) the parameters of a GARCH process concern both the overall scale of the GARCH process and the relative scale of subsequent values of the GARCH process. The correction factor $\sigma^*$ is needed to ensure that the overall scale of the process is identifiable. Newey and Steigerwald (1997) do not need this correction factor because they estimate the ratio between the parameters of a GARCH model that is linear in its parameters, which is independent of the overall scale of the GARCH process.}
function $\rho(\cdot)$. Denote $E_{\theta^*}$ the expectation with respect to $\mathcal{P}_{\theta^*}$. The correction factor $\sigma^*$ equals

$$
\sigma^* = N/E_{\theta^*}[\psi(u_t'u_t)u_t'u_t] = N/[2\pi^{N/2}\Gamma(N/2)] \int_0^\infty \psi(t^2)t^{N+1}g_*(t^2)dt].
$$

(12)

For the M-estimator with Gaussian loss function (being the Gaussian QML estimator) the correction factor is one for all elliptical density functions $g_*(\cdot)$, whatever be the dimension $N$ of the series. When the loss function is proportional to $\log g_*(\cdot)$, the M-estimator coincides with the ML estimator and $\sigma^* = 1$.

If the density function $g_*(\cdot)$ is not specified, the correction factor $\sigma^*$ needs to be estimated. For this, we recommend to first select a subset of the data for which there is no visual evidence of outliers and to use the Gaussian QML estimator to obtain an estimate of the innovations for this subset of the data. Let $\hat{u}_1, \ldots, \hat{u}_T$ be the series of such estimated innovations. Then the correction factor can be either estimated by

$$
N/[\frac{1}{T} \sum_{t=1}^{T} \psi(\hat{u}_t'\hat{u}_t)\hat{u}_t'\hat{u}_t],
$$

(13)

or by

$$
N/[2\pi^{N/2}\Gamma(N/2)] \int_0^\infty \psi(t^2)t^{N+1}\hat{g}_*(t^2)dt].
$$

(14)

where $\hat{g}_*(\cdot)$ is the Hodgson et al. (2002)’s Kernel estimate of $g_*(\cdot)$ on the basis of the $\hat{u}_t'\hat{u}_t$’s. The M-estimator in (10) can then be computed, yielding
a new set of estimated innovations that can be used to update the estimated correction factor $\sigma_*$ using (13) or (14).

### 3 Asymptotic properties

Jeantheau (1998) and Comte and Lieberman (2003) establish conditions under which the Gaussian QML-estimator is consistent and asymptotically normally distributed. In this section, we give conditions under which the M-estimator with loss functions other than the Gaussian one, is consistent and asymptotically normal.

**A1** The loss function $\rho(z)$, for $z > 0$, has the following properties

1. Its derivative $\psi(z) = \rho'(z)$ is nonnegative, nonincreasing and continuous.
2. The function $\psi(z)z$ is bounded. Let $K = \sup_{z \geq 0} \psi(z)z$. The function $\psi(z)z$ is nondecreasing, and is strictly increasing in the interval where $\psi(z)z < K$. Furthermore, there exists a $z_0$ such that $\psi(z_0^2)z_0^2 > N$ and $\psi(z)z > 0$ for $z \leq z_0$.
3. There exists $b > 0$ such that for every hyperplane $H$, the probability under $P_{\theta_*}$ for an observation to lie in $H$ is at most $1 - N/K - b$.

Assumption A1 is similar to the conditions required by Maronna (1976) for the consistency of M-estimators of scale. Assumptions A1.1 and A1.2 put conditions on the loss function. It is immediate to see that the Student $t$
loss function satisfies these assumptions. Furthermore, if the density function of the innovations is continuous, condition A1.3 always holds. We also need assumptions on the MGARCH model. Let us denote $\partial H_{t,\theta}/\partial \theta_i$ the matrix containing the $i$-th partial derivative of the corresponding elements of $H_{t,\theta}$.

**A2** The conditional covariance matrix $H_{t,\theta}$ is continuously differentiable w.r.t. $\theta$ and is uniquely identified: $H_{t,\theta} = H_{t,\theta_*} \Rightarrow \theta = \theta_*$. Jeantheau (1998) and Comte and Lieberman (2003) make the same identifiability assumption on the MGARCH specification. Jeantheau (1998) gives sufficient conditions for this to hold for MGARCH models with constant conditional correlation, while Engle and Kroner (1995) do this for the BEKK model in which conditional correlations can be time varying.

Finally, we also need an assumption on the parameter space and on the function $m(r_t; \theta, \rho) = \log \det H_{t,\theta} + \sigma_* \rho (r_t' H_{t,\theta}^{-1} r_t)$.

**A3** The parameter space $\Theta$ is a compact set with $E_{\theta_*} \sup_{\theta \in \Theta} |m(r_t; \theta, \rho)| < \infty$. The assumption of a compact parameter space is standard (see e.g. Jeantheau, 1998, and Comte and Lieberman, 2003). For bounded loss functions, the condition that $E_{\theta_*} \sup_{\theta \in \Theta} |m(r_t; \theta, \rho)|$ is finite will be satisfied if $E_{\theta_*} \sup_{\theta \in \Theta} |\log \det H_{t,\theta}| < \infty$, which is an assumption that is verified for all covariance stationary MGARCH processes.
In the Appendix we show that under model $\mathcal{P}_{\theta^*}$ given by (1), (2) and (5) and under the assumptions A1-A3, $\theta^*$ is a local minimum of the probability limit of the M-function in (10). To obtain consistency, we also need the following assumption.

**A4** The matrix $E_{\theta^*}[I_N - \sigma^* \psi(r_i' H_t^{-1}_{t,\theta^*} r_i) r_i' H_t^{-1} | I_{t-1}] \frac{\partial H_t}{\partial \theta^i}$ is positive semidefinite for all parameter components $\theta_i$.

This assumption is verified for $\theta = \theta^*$ since, due to the choice of $\sigma^* = N/E_{\theta^*}[\psi(u'_i u_i) u'_i u_i]$, we have that $E_{\theta^*}[\sigma^* \psi(r_i' H_t^{-1}_{t,\theta^*} r_i) r_i' r_i | I_{t-1}] = H_t,\theta^*$.

**Proposition 1** Under model $\mathcal{P}_{\theta^*}$ given by (1), (2) and (5) and under the assumptions A1-A4, we have that $\hat{\theta}_M \to \theta^*$ for $T \to \infty$.

In the Appendix we prove that the M-estimator is asymptotically normal at $\mathcal{P}_{\theta^*}$, if in addition to the assumptions ensuring the M-estimator to be consistent, also the following assumptions are satisfied.

**A5** The loss function is three times continuously differentiable w.r.t. $\theta$ and these derivatives are bounded.

**A6** The function $H_\theta(\cdot)$ is three times continuously differentiable w.r.t. $\theta$.

**A7** The process $r_t$ admits bounded moments of order 8.
The Student $t$ loss function satisfies A5. Assumptions A6 and A7 are the same as in Theorem 3 of Comte and Lieberman (2003) establishing asymptotic normality of the Gaussian QML estimator for MGARCH models.

**Proposition 2** Under model $P_{\theta^*}$ given by (1), (2) and (5) and under the Assumptions A1-A7 the M-estimator $\hat{\theta}_M$ is asymptotically normal

$$\sqrt{T}(\hat{\theta}_M - \theta^*) \overset{d}{\to} N(0, C_0^{-1}C_1^{-1})$$

for $T \to \infty$. \hspace{1cm} (15)

The matrices $C_0$ and $C_1$ equal

$$C_0 = E_{\theta^*}[\frac{\partial m(r_t; \theta, \rho)}{\partial \theta} \frac{\partial m(r_t; \theta, \rho)}{\partial \theta^*}]_{\theta=\theta^*}, \quad C_1 = E_{\theta^*}[\frac{\partial^2 m(r_t; \theta, \rho)}{\partial \theta \partial \theta^*}]_{\theta=\theta^*}. \hspace{1cm} (16)$$

These matrices can be estimated by their sample counterparts.\footnote{Analytical expressions for the first and second derivative of the BEKK model are reported in Hafner and Herwartz (2008) and Lucchetti (2002).}

## 4 Robust M-estimation

We are interested in the estimation of the parameters of a MGARCH model in the presence of outliers in the data. We suppose that there exists a central parametric model denoted $P_{\theta}$ (imposed by the econometrician) and a parameter vector $\theta^*$ such that $P_{\theta^*}$ explains well the large majority (and not necessarily all) of the return observations. In order to estimate $\theta^*$, we need an estimator that cannot be highly influenced by one single observation and that yields a good fit for the bulk of the data (but not the outliers).
We identify outliers as the observations with a large squared Mahalanobis Distance (MD). In order to safeguard our analysis against outliers, we propose to use (i) M-estimators with a loss function that downweights observations with a large MD and (ii) a bounded innovation propagation MGARCH model, i.e. a MGARCH model for which the effect of innovations on future volatility is bounded.

4.1 The loss function

In (26) of the Appendix, we show that the M-estimator of $H_{t,\theta}$ equals a weighted covariance matrix, with weights proportional to $\rho'(d_{t,\theta}^2)$. Consequently, the M-estimator will be robust to outliers if the derivative of the loss function gives a lower weight to observations with a large MD. This will be the case if the derivative of the loss function decreases sufficiently fast to zero. More precisely, we need condition A1.2 to ensure that outliers have a bounded effect on the estimator. The derivatives of the Gaussian and Student $t_\nu$ loss functions equal

$$
\rho'_\phi(z) = 1 ; \quad \rho'_{t_\nu}(z) = \frac{N + \nu}{\nu - 2 + z}.
$$

(17)

We see that the M-estimator with Gaussian loss function is not robust because it gives the same weight to all realizations irrespective of their degree of outlyingness. However, the derivative of the Student $t_\nu$ loss function tends to zero
at infinity at the rate $\varepsilon^{-1}$. The smaller $\nu$ is, the more robust the M-estimator with Student $t_{\nu}$ loss function will be. In the simulation study and empirical application of Sections 5 and 6, we use the Student $t_4$ loss function as a more robust alternative to the Gaussian loss function.

### 4.2 Impact of news on subsequent volatility predictions

There is a rich empirical evidence (e.g. Hamilton and Susmel, 1994) that extremely large shocks, such as the October 1987 crash, have a proportionally smaller effect on subsequent volatility than do small and moderate shocks. Most MGARCH models do not account for this and model the effect of past returns on future volatility as quadratic, neglecting the difference in reactivity of volatility to extremely large shocks. This leads to overly pessimistic volatility predictions in periods subsequent to extreme return realizations. Because of the persistency of the MGARCH process, this overestimation of volatility dies out only slowly.

We propose to remediate for this by modeling the conditional covariance matrix by the function $H_\theta(\cdot)$ in (2) of the returns weighted in function of their squared Mahalanobis Distance

$$\tilde{H}_{t,\theta} = H_\theta(\tilde{r}_{1,\theta}, \ldots, \tilde{r}_{t-1,\theta}),$$

(18)

with $\tilde{r}_{t,\theta} = r_t \sqrt{w(r_t' \tilde{H}_{t,\theta}^{-1} r_t)}$. The weight function must be such that the effect
of \( r_t \) on \( \tilde{H}_{t,\theta} \) is bounded. Henceforth, we will use the following weight function

\[
w_c(z) = \begin{cases} 
1 & \text{if } z \leq c_1 \\
1 - (1 - c_1/z)^3 & \text{if } c_1 < z \leq c_2 \\
(c_2/z)(1 - (1 - c_1/c_2)^3) & \text{else.}
\end{cases}
\]  

(19)

The parameters \( c_1 \) and \( c_2 \) equal the 99% and 99.9% quantile of the distribution in (7) of the squared Mahalanobis Distances.\(^4\) We plot this weight function in Figure 1. Note that only the observations with an extremely large MD are downweighted and that the weighting heavily depends on the distribution assumption. In Figure 1 we also plot the function \( w(z)z \), which is of interest since, if \( z \) is the squared MD of \( r_t \), then \( w(z)z \) will be the squared MD of \( \tilde{r}_t \).

Note that the downweighting is such that the function \( w_c(z)z \) is nondecreasing and bounded by \( w_c(c_2)c_2 \). The smoothness of the weight function is needed to avoid numerical problems in the parameter estimation.

The idea of robustifying the conditional covariance matrix against outliers in the data by downweighting past observations of \( r_t \) in the GARCH specification has already been pursued by Muler and Yohai (2002,2007) in the univariate case. We follow them in calling this the \textit{Bounded Innovation Propagation} (BIP) MGARCH model. When \( H_\theta(\cdot) \) has the BEKK specification

\(^4\)In the empirical application, we assume a MGARCH model with Student \( t \) innovations. We choose the number of degrees of freedom such that the correction factor \( \sigma_* \) in (12) equals the one obtained using the semiparametric estimator in (13) or (14).
Figure 1: Plot of the functions $w_c(z)$ and $w_c(z)z$ used in the BIP MGARCH model. The parameters $c_1$ and $c_2$ equal to the 99% and 99.9% quantiles of the squared MD under Gaussian (upper panel) or Student $t_4$ innovations (lower panel).

(3), we obtain the following BIP-BEKK model for the conditional covariance matrix:

$$
\hat{H}_{t,\theta} = C'C + \sum_{j=1}^{q} \sum_{k=1}^{K} A'_{kj} \hat{\epsilon}_{t-j} \hat{\epsilon}_{t-j} A_{kj} + \sum_{j=1}^{p} \sum_{k=1}^{K} B'_{kj} \hat{H}_{t-j,\theta} B_{kj}.
$$

(20)

Under the assumption that the MD is an indicator of outlyingness, the BIP MGARCH model can be interpreted as the MGARCH model of the return series from which the outlying component has been filtered away. Consequently, it is possible that in the presence of outliers, the use of the misspecified BIP
MGARCH model can produce more accurate parameter estimates and volatility predictions than the correctly specified MGARCH model of the raw returns. Similarly as in Muler and Yohai (2002, 2007), we therefore propose that, if there is a suspicion of outliers in the data, one estimates the parameter $\theta^*$ of the MGARCH model using the misspecified BIP MGARCH model. This estimator is called the BIP M-estimator. It is defined as the value of $\theta \in \Theta$ that minimizes the BIP M-function

$$\tilde{M}(S_T; \theta, \rho) = \frac{1}{T} \sum_{t=1}^{T} \left[ \log \det \tilde{H}_{t,\theta} + \sigma_t^* \rho \left( r_t' \tilde{H}_{t,\theta}^{-1} r_t \right) \right].$$

(21)

In the next section, we show that, in the presence of additive outliers, the BIP M-estimator of $\theta^*$ is more accurate than the M-estimator based on the correctly specified MGARCH model.

5 Simulation study

In this section we show through a simulation study that the BIP M-estimator with Student $t_4$ loss function offers an attractive trade-off between robustness to additive outliers and efficiency under the model assumptions.

We consider the bivariate symmetric BEKK volatility model of order one, proposed by Engle and Kroner (1995), and assume the innovations to be Student $t$ distributed. Under this model, the data generating process is given by
the following set of equations:

\[
\begin{cases}
    r_t | I_{t-1} \overset{\text{i.i.d.}}{\sim} t_{\nu}(0, H_{t,\theta}) \ ; \ \nu > 2 \\
    H_{t,\theta} = C' C + A' r_{t-1} r_{t-1}' A + B' H_{t-1,\theta} B.
\end{cases}
\]  

(22)

The parameter matrices $C$, $A$ and $B$ all denote $2 \times 2$ coefficient matrices. The matrix $C$ is upper triangular and the matrices $A$ and $B$ are symmetric.\(^5\) Let $\theta$ be the vector of length 9 that stacks the upper diagonal elements of the matrices $C$, $A$ and $B$ one on top of the other. Engle and Kroner (1995) and Altay-Salih et al. (2003) show that under the symmetric BEKK specification, the estimates of $H_{t,\theta}$ will be positive definite and covariance stationary if $\theta$ satisfies the following constraints:

\[
C_{11} > 0, C_{22} > 0, A_{11} > 0, B_{11} > 0.
\]  

(23)

\[
I_4 - A' \otimes A' - B' \otimes B' \text{ is positive semidefinite}.
\]  

(24)

We compute the M-estimator in (10) using a sequential quadratic programming algorithm.

In order to study the effect of outlier contamination on (BIP) M-estimators with Gaussian or Student $t_4$ loss function, we consider 200 replications of simulated series of length 1000 from model (22) with Gaussian innovations

\(^5\)We assume the matrices $A$ and $B$ to be symmetric such that, in the computations, the covariance stationarity constraint on $H_{t,\theta}$ can be imposed as inequality constraints on quadratic functions of the parameter components (see Altay-Salih et al., 2003).
and with parameter matrices$^6$

$$C'C = \begin{pmatrix} 0.16 & 0.20 \\ 0.20 & 0.34 \end{pmatrix}, \quad A = \begin{pmatrix} 0.25 & 0.04 \\ 0.04 & 0.30 \end{pmatrix}, \quad B = \begin{pmatrix} 0.80 & 0.05 \\ 0.05 & 0.80 \end{pmatrix}.$$ 

For each realization, there is a probability $\varepsilon$ that the observed value is not $r_t$ but $r_t + z_t$, with $z_t$ an additive outlier whose magnitude is 5 times the conditional standard deviation of the corresponding element of $r_t$.\textsuperscript{7} Note that the observations affected by additive outliers are, after standardization by their MGARCH volatility, extremely large and that they affect future volatility less than predicted by the MGARCH model. Since the returns corresponding to the stock market crash in October 1987, share the same properties, one can interpret the additive outliers $z_t$ as the component of large shocks $r_t$ that does not affect future volatility.

In our Monte Carlo study, we consider three levels of outlier contamination ($\varepsilon = 0$, $\varepsilon = 0.01$ and $\varepsilon = 0.05$) and four M-estimators: the (BIP) M-estimators with Gaussian or Student $t_4$ loss function. For each M-estimator and for each

\textsuperscript{6}The dependence parameter matrices $A$ and $B$ are similar to those in Chapter 18 of Lütkepohl (2005), while the matrix $C$ is chosen such that it has a similar scale as $A$ and $B$.

\textsuperscript{7}More precisely, $z_t = 5\xi_t d_t$, where $d_t$ is a sequence of i.i.d. draws from a Bernoulli distribution such that the occurrence of an outlier ($d_t = 1$) has probability $\varepsilon$. The 2-dimensional i.i.d. vector process $\xi_t$ is such that, when an outlier occurs, it will be either in the first, second or both components with probability 0.3, 0.3 and 0.4, respectively, and its magnitude will be 5 times the conditional standard deviation of the corresponding element of $r_t$. 

22
type of simulated series, we report in Table 1 the root mean squared error (RMSE), computed as follows,

$$\text{RMSE} = \sqrt{\frac{1}{J} \sum_{j=1}^{J} \| \hat{\theta}_j - \theta_* \|^2},$$

(25)

where $J = 200$ and $\| \cdot \|$ is the Euclidian norm operator.

We find that in the absence of outliers ($\varepsilon = 0$) and for Gaussian innovations, the robust estimators have a RMSE similar to the Maximum Likelihood estimator. The estimation of the BEKK model on the basis of the BIP-BEKK model only leads to a slightly higher RMSE. Note also the loss in efficiency when the Gaussian QML estimator is used to estimate the BEKK model with Student $t_4$ innovations. The RMSE is 0.40 for the Gaussian QML estimator and 0.25 for the ML estimator.

If we add outlier contamination to the data, the RMSE of all M-estimators considered in Table 1 increases, but this increase is much larger for the M-estimator with Gaussian loss function than for the M-estimator with Student $t_4$ loss function. For the BEKK model with Gaussian innovations, we find that in the absence of outlier contamination the RMSE is 0.28 and 0.31 for the M-estimators with Gaussian and Student $t_4$ loss functions respectively. In the presence of 5% of outlier contamination, the RMSE of the M-estimator with Gaussian loss function is 1.02, while it is only 0.66 for the M-estimator with Student $t_4$ loss function.
In Table 1 we also see that the RMSE of the M-estimator with Gaussian loss function is always smaller than the RMSE of its BIP counterpart. We thus find that the estimation of the BEKK model using the BIP-BEKK model is not sufficient to obtain a robust estimation procedure. However, among the considered M-estimators, the BIP M-estimator with Student \( t_4 \) loss function is the most robust estimator since it offers a relatively high efficiency under the model without outliers and it has the lowest RMSE in the presence of additive outlier contamination. It seems thus that a high robustness to outliers is only achieved if the BIP-BEKK model is used in combination with a robust loss function.

The convergence frequency of the BIP M-estimator is slightly lower than the convergence frequency of the M-estimator. This is to be expected, since the BIP M-estimator is based on a misspecified model. For the BIP M-estimator with Student \( t_4 \) loss function (the robust estimator we recommend), the convergence frequency is the highest (92%) for the BEKK model with Student \( t_4 \) innovations and 0 or 1% of outlier contamination and the lowest (72%) for the BEKK model with Gaussian innovations and 5% of outlier contamination.
Table 1: Root mean squared error of the (BIP) M-estimators with Gaussian and Student $t_4$ loss function, when a proportion $\varepsilon$ of the Gaussian and Student $t_4$ BEKK simulated data are outlier contaminated.

<table>
<thead>
<tr>
<th>Loss function</th>
<th>Gaussian loss</th>
<th>Student $t_4$ loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIP</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>$\varepsilon = 0$</td>
<td>Gaussian innov.</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>Student $t_4$ innov.</td>
<td>0.40</td>
</tr>
<tr>
<td>$\varepsilon = 0.01$</td>
<td>Gaussian innov.</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Student $t_4$ innov.</td>
<td>0.64</td>
</tr>
<tr>
<td>$\varepsilon = 0.05$</td>
<td>Gaussian innov.</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>Student $t_4$ innov.</td>
<td>0.92</td>
</tr>
</tbody>
</table>

6 Empirical application

In this section we compare the forecasting performance of the BEKK and BIP-BEKK model for the 1980-2006 daily return series of the Nasdaq and NYSE stock market indices. The data were obtained from datastream and consist of 7043 observations. The series are plotted in Figure 2. Note the large negative return corresponding to the stock market crash of October 19, 1987. Interestingly, the Nasdaq index had returns of similar magnitude in the year 2000. In an unconditional framework, these returns are qualified as extreme tail realizations, whereas under a conditionally heteroscedastic time series model, some of these extreme returns correspond to innovations from the
Figure 2: Daily returns on the Nasdaq and NYSE composite index for the period 1980-2006, together with the log of squared Mahalanobis Distances under the BIP-BEKK model with Student $t_7$ innovations.

center of the distribution. Observe that not only there is volatility clustering in both series, but that there is also a lot of commonality in volatility across the two series.

We assume that the very large majority of these observations are generated by the symmetric BEKK time series process specified in (22), with elliptical innovations. Return realizations such as the 1987 stock market crash are clearly outliers under this parametric framework. Hence a robust estimator is needed. We use the M-estimator with Student $t_4$ loss function. For estimating the
correction factor $\sigma_*$, we use the estimator in (13) based on the Gaussian QML estimated innovations for the subperiod January 1990-December 1999 in which there is no visual evidence of outliers. We obtain $\hat{\sigma}_* = 0.897$. Because this is the correction factor needed under a MGARCH model with Student $t_7$ innovations, we compute the thresholds $c_1$ and $c_2$ of the BIP-BEKK model in (20) under the assumption of Student $t_7$ innovations. Then we re-estimate the model using the BIP M-estimator with Student $t_4$ loss function based on the complete time series. The third panel in Figure 2 plots the log of the estimated squared Mahalanobis Distances and their 99% and 99.9% quantile under the BIP-BEKK model with Student $t_7$ innovations. Recall that the MD is based on the returns standardized by their volatility estimate, such that one has only a large MD if the return is extreme with respect to its volatility prediction. We see that the 1987 stock market crash is clearly detected as an outlier, while there are no extreme exceedances in the MDs of the returns observed in the high volatility period of the year 2000.

In Figure 3 we report the estimates of the conditional standard deviation and correlation of the daily returns on the Nasdaq and NYSE composite index. It is of particular interest to compare the BEKK and BIP-BEKK volatility predictions for the period of the 1987 stock market crash and the period of the burst of the internet bubble in 2000. In each of these periods, large returns of similar magnitude occur but the persistence in volatility is different. The
Figure 3: Conditional BEKK and BIP-BEKK standard deviation and correlation of the daily returns on the Nasdaq and NYSE composite index for the period 1980-2006.

Stock market crash in 1987 was not followed by returns of similar size whereas the burst of the internet bubble corresponds to a prolonged period of high volatility. As can be seen in Figure 3, the BIP-BEKK model captures well this difference in reaction to shocks. In contrast, the BEKK model overestimates the persistence in the realized return variability in the period subsequent to the 1987 crash. The same interpretation holds for the patch of extreme returns on the Nasdaq index on March 27-28, 1980. At all other times, the BEKK and BIP-BEKK conditional standard deviation forecasts are similar.
In the bottom panel of Figure 3 we see that historically the BEKK and BIP-BEKK conditional correlation between the daily returns on the Nasdaq and NYSE indices, has been stable around 0.79. This period of almost constant correlation has been temporarily disrupted by the large asymmetric and persistent volatility shock due to the burst of the internet bubble in the year 2000. We also find that the daily BIP-BEKK conditional correlation prediction is more smooth and has less spikes than its BEKK counterpart. Consider e.g. the reaction of the BEKK and BIP-BEKK conditional correlations to stock market crash on October 18, 1987. The predicted correlation by the BEKK and BIP-BEKK model for the day of the crash is the same, namely 0.93, but according to the BEKK model, the conditional correlation for the days after the crash first jumps to 0.99 on October 19 and then falls back to 0.77 on October 20 and remains below 0.8 in the month following the crash. Under the BIP-BEKK model, the conditional correlation increases on October 19 to 0.96, but instead of falling back immediately to the historical average of 0.79 it stays above 0.9 in the month following the crash.

7 Concluding remarks

In this paper we introduce a class of M-estimators for multivariate GARCH time series models and provide asymptotic theory for these estimators. We
study the effect of the shape of the loss function and of bounding the news impact curve on the robustness of the M-estimator to outliers. The popular Gaussian quasi-maximum likelihood estimator is shown to be very sensitive to outliers in the data. If the data is suspected to be contaminated by outliers, we recommend the use of a volatility model that has the property of bounded innovation propagation and to use a M-estimator with Student $t_4$ loss function. The Monte Carlo study and empirical application document the good robustness properties of this approach.

This research can be extended by looking at other classes of robust estimators and by considering other multivariate GARCH models. In recent research on these models, the tendency has been to reduce the dimensionality of the problem using principal component analysis or by specifying separate univariate GARCH models for the volatility of each series and a time-varying process for the conditional correlations. Further research could show how well robust methods work for these types of models.
8 Appendix

Proof of Proposition 1.

To obtain consistency of the M-estimator, we need to establish the following three intermediary results (Proposition 7.3 in Hayashi, 2000).

R1 The M-function $M(S_t; \theta, \rho)$ in (10) converges in probability to a function that no longer depends on the data, say $m_\infty(\theta, \rho)$.

R2 The true parameter solves the limit problem of minimizing $m_\infty(\theta, \rho)$.

R3 The minimum of $m_\infty(\theta, \rho)$ is the probability limit of the minimum of the M-function $M(S_t; \theta, \rho)$.

Under our model $\mathcal{P}_{\theta^*}$ given by (1), (2) and (5), the process $r_t$ is strictly stationary and ergodic. Since $m(r_t; \theta, \rho) = \log \det H_t, \theta + \sigma_s \rho \left( r_t'H_t^{-1}r_t \right)$ is a measurable function of $r_t$, it is also strictly stationary and ergodic. By the Ergodic Theorem the M-function $M(S_t; \theta, \rho)$ in (10), being the time series average of $m(r_1; \theta, \rho), \ldots, m(r_T; \theta, \rho)$, converges a.s. to $m_\infty(\theta, \rho) = E_{\theta^*}[m(r_t; \theta, \rho)]$ for $T \to \infty$. We thus find that under $\mathcal{P}_{\theta^*}$ condition R1 is always satisfied. Assumption A3 is a sufficient condition for R3 to be satisfied.

The proof of R2 is longer. We first note that by the law of iterated expectations $m_\infty(\theta, \rho) = E_{\theta^*}[E_{\theta^*}[m(r_t; \theta, \rho) | I_{t-1}]]$ and that the partial derivative of
\(m(r_t; \theta, \rho)\) can be written as follows

\[
\frac{\partial m(r_t; \theta, \rho)}{\partial \theta} = \text{Tr} \left[ I_N - \sigma_* \psi(r'_t H^{-1}_{t, \theta} r_t) r_t r'_t H^{-1}_{t, \theta} \right] \frac{\partial H_{t, \theta}}{\partial \theta} H^{-1}_{t, \theta},
\]

where \(\text{Tr}\) is the trace operator (see e.g. Comte and Lieberman, 2003). By Lemmas 1 and 2 stated below, we have that under model \(\mathcal{P}_{\theta_*}\) given by (1), (2) and (5) and under the assumptions A1-A3, \(\theta_*\) will be a local minimum of the \(m_\infty(\theta, \rho)\). To prove consistency under the additional assumption that \(E_{\theta_*} \left[ I_N - \sigma_* \psi(r'_t H^{-1}_{t, \theta} r_t) r_t r'_t H^{-1}_{t, \theta} | I_{t-1} \right] \frac{\partial H_{t, \theta}}{\partial \theta} \) is positive semidefinite, we use the property that if \(A\) is positive semidefinite and \(B\) is positive definite, then the \(\text{Tr}(AB^{-1}) = 0\) iff \(A = 0\). Because in (26) \(H_{t, \theta}\) is positive definite, we must have thus that

\[
E_{\theta_*} \left[ I_N - \sigma_* \psi(r'_t H^{-1}_{t, \theta} r_t) r_t r'_t H^{-1}_{t, \theta} | I_{t-1} \right] \frac{\partial H_{t, \theta}}{\partial \theta} = O_N.
\]

Similarly as in Lemma 4.2 in Ling and McAleer (2003), one can show that this implies a.s. that \(E_{\theta_*} \left[ I_N - \sigma_* \psi(r'_t H^{-1}_{t, \theta} r_t) r_t r'_t H^{-1}_{t, \theta} | I_{t-1} \right] = O_N\), which is equivalent to

\[
\sigma_* E_{\theta_*} \left[ \psi(r'_t H^{-1}_{t, \theta} r_t) r_t r'_t | I_{t-1} \right] = H_{t, \theta}.
\]

Since \(r_t = H^{1/2}_{t, \theta} u_t, H_{t, \theta}\) solves the system in (27) when \(\sigma_* = N/E_{\theta_*} [\psi(u'_t u_t) u'_t u_t]\). Maronna (1976) shows that under \(\mathcal{P}_{\theta_*}\) and A1-A2, the solution to (27) is unique. We thus obtain that under \(\mathcal{P}_{\theta_*}\) and A1-A4, the probability limit of the M-function is uniquely minimized at \(H_{t, \theta}\) and under the no observational...
equivalence assumption on \( H_\theta(\cdot) \) in A2, \( \theta_* \) will be the unique minimizer of \( m_\infty(\theta, \rho) \).

\[\square\]

**Lemma 1** Under \( P_{\theta_*} \), the true value \( \theta_* \) is a critical value of \( m_\infty(\theta, \rho) \).

Proof. Due to the choice of \( \sigma_* = N/E_{\theta_*}[\psi(u_t' u_t) u_t' u_t] \), we have that \( \sigma_* E_{\theta_*}[\psi(u_t' u_t) u_t' u_t] = I_N \). Furthermore, since \( r_t = H_{t, \theta_*}^{1/2} u_t \), it directly follows from taking conditional expectations of (26) that \( \theta_* \) solves the first order condition for \( E_{\theta_*}[m(r_t; \theta, \rho)|I_{t-1}] \) and hence also for \( m_\infty(\theta, \rho) \).

\[\square\]

**Lemma 2** Under \( P_{\theta_*} \) and under A1.1, the Hessian of \( m_\infty(\theta, \rho) \) is positive definite at \( \theta_* \).

Proof. Consider the following decomposition of \( \partial^2 m(r_t; \theta, \rho)/\partial \theta_j \partial \theta_i = m_{a,i,j}(\theta) + m_{b,i,j}(\theta) \), with

\[
m_{a,i,j}(\theta) = \text{Tr}\left\{ \frac{\partial^2 H_{t,\theta}^{-1}}{\partial \theta_j \partial \theta_i} H_{t,\theta}^{-1} - \frac{\partial H_{t,\theta}^{-1}}{\partial \theta_j} H_{t,\theta}^{-1} \frac{\partial H_{t,\theta}^{-1}}{\partial \theta_i} H_{t,\theta}^{-1} + [\sigma_* \psi(r_t' H_{t,\theta}^{-1} r_t) r_t' H_{t,\theta}^{-1}] \right\}
\]

\[
m_{b,i,j}(\theta) = -\sigma_* \psi(r_t' H_{t,\theta}^{-1/2} r_t) r_t' \text{Tr}(\frac{\partial H_{t,\theta}^{-1}}{\partial \theta_j} H_{t,\theta}^{-1}) \text{Tr}(r_t' H_{t,\theta}^{-1} \frac{\partial H_{t,\theta}^{-1}}{\partial \theta_i} H_{t,\theta}^{-1}).
\]

Note that \( \sigma_* E_{\theta_*}[\psi(r_t' H_{t,\theta_*}^{-1} r_t) r_t' H_{t,\theta_*}^{-1} | I_{t-1}] \) equals the identity matrix and thus,

\[
E_{\theta_*}[m_{a,i,j}(\theta_*) | I_{t-1}] = \text{Tr}(\frac{\partial H_{t,\theta}^{-1}}{\partial \theta_j} H_{t,\theta}^{-1} \frac{\partial H_{t,\theta}^{-1}}{\partial \theta_i} H_{t,\theta}^{-1}).
\] (28)

Denote by \( H_{t,\theta_*}^{-1/2} \) the symmetric root of \( H_{t,\theta_*}^{-1} \). Using that for any matrices
A, B, Tr(AB) = Tr(BA) and Tr(AB) = vec(A′)vec(B), we further find that

\[ E_{\theta^*}[m^h_{i,j}(\theta_*)|I_{t-1}] = k_*vec'(H^{-1/2}_t,\theta_* \frac{\partial H_{t,\theta_*}}{\partial \theta_j} H^{-1/2}_t)vec(H^{-1/2}_t,\theta_* \frac{\partial H_{t,\theta_*}}{\partial \theta_i} H^{-1/2}_t), \]  

(29)

where \( k_* = E_{\theta^*}[-\sigma_* \psi'(u'u)(u'u)^2] > 0 \) by A1.1. It follows from the results in Comte and Lieberman (2003, p.77-78) that the \( N \times N \) matrices with \( i,j \)-th element \( E_{\theta^*}[m^i_{i,j}(\theta_*)|I_{t-1}] \) and \( E_{\theta^*}[m^h_{i,j}(\theta_*)|I_{t-1}] \), are positive definite. Since the Hessian of \( m_\infty(\theta, \rho) \) is the unconditional expectation of the sum of these two matrices, it will also be positive definite.

\[ \square \]

Proof of Proposition 2.

Since \( \partial m(r_t;\theta^*, \rho) / \partial \theta \) and \( \partial^2 m(r_t;\theta^*, \rho) / \partial \theta \partial \theta' \) are ergodic processes under our model \( P_{\theta^*} \) given by (1), (2) and (5), it follows from the theorem of Basawa et al. (1976) that asymptotic normality holds if the loss function and conditional covariance matrix function are three times differentiable and if the following conditions are verified (see also Comte and Lieberman, 2003).

R1' \[ C_0 = E_{\theta^*}\left[ \frac{\partial m(r_t;\theta, \rho)}{\partial \theta} \frac{\partial m(r_t;\theta, \rho)}{\partial \rho} \right]_{\theta=\theta^*} \text{ is finite.} \]

R2' \[ C_1 = E_{\theta^*}\left[ \frac{\partial^2 m(r_t;\theta, \rho)}{\partial \theta \partial \rho} \right]_{\theta=\theta^*} \text{ is finite and positive definite.} \]

R3' For all \( i, j, k \), \( E_{\theta^*}[\sup_{\|\theta-\theta^*\| \leq \delta} |\frac{\partial^3 m(r_t;\theta, \rho)}{\partial \theta_i \partial \theta_j \partial \theta_k}|] \) is bounded for all \( \delta > 0 \).

Comte and Lieberman (2003) prove that for the loss function \( \rho(z) = z \), R1’-R3’ are satisfied if the \( r_t \) process admits bounded moments of order 8. The following lemma extends this result to M-estimators.  

\[ \square \]
Lemma 3  If Assumptions A1-A7 hold, then R1’-R3’ are satisfied.

Proof. By Lemma 2 the positive definiteness requirement in R2’ is satisfied. Write \( m = m(r_t; \theta, \rho) \), \( a = \log \det H_{t,\theta} \), \( b = r_t' H_{t,\theta}^{-1} r_t \) and express the partial derivatives of a function \( z \) as \( z_i = \partial z / \partial \theta_i \), \( z_{ij} = \partial^2 z / \partial \theta_i \partial \theta_j \), \( z_{ijk} = \partial^3 z / \partial \theta_i \partial \theta_j \partial \theta_k \). Then \( m = a + \sigma \rho(b) \) and

\[
\begin{align*}
m_i m_j &= a_i a_j + \sigma \rho'(b)(a_i b_j + a_j b_i) + \sigma^2 \rho'(b)b_i b_j \\
m_{ij} &= a_{ij} + \sigma \rho''(b)b_j b_i + \sigma \rho'(b)b_{ij} \\
m_{ijk} &= a_{ijk} + \sigma \rho'''(b)b_k b_i b_j + \sigma \rho''(b)(b_{jk} b_i + b_j b_{ik} + b_k b_{ij}) + \sigma \rho'(b)b_{ijk}.
\end{align*}
\]

From these formulas it can be directly verified that, if the requirements of finite and bounded moments in R1’-R3’ are satisfied for \( \rho(z) = z \), they will also be satisfied for loss functions with continuous and bounded first, second and third derivative. \( \square \)
References


