CREATES Research Paper 2008-46

Semiparametric Inference in a GARCH-in-Mean Model

Bent Jesper Christensen, Christian M. Dahl and Emma M. Iglesias

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Semiparametric Inference in a GARCH-in-Mean Model*

Bent Jesper Christensen† — University of Aarhus and CREATES

Christian M. Dahl‡ — University of Aarhus and CREATES

Emma M. Iglesias§ — Michigan State University

This version: August 2008.

Abstract

A new semiparametric estimator for an empirical asset pricing model with general nonparametric risk-return tradeoff and a GARCH process for the underlying volatility is introduced. The estimator does not rely on any initial parametric estimator of the conditional mean function, and this feature facilitates the derivation of asymptotic theory under possible nonlinearity of unspecified form of the risk-return tradeoff. Besides the nonlinear GARCH-in-mean effect, our specification accommodates exogenous regressors that are typically used as conditioning variables entering linearly in the mean equation, such as the dividend yield. Using the profile likelihood approach, we show that our estimator under stated conditions is consistent, asymptotically normal, and efficient, i.e. it achieves the semiparametric lower bound. A sampling experiment provides evidence on finite sample properties as well as comparisons with the fully

*We are grateful to Valentina Corradi, Enno Mammen, Joon Y. Park and participants at the 2008 SETA Conference in Seoul and at the 2008 European Meeting of the Econometric Society in Milan for very useful comments. We also thank the Center for Research in Econometric Analysis of TimE Series (CREATES) funded by the Danish National Research Foundation, the Danish Social Science Research Council, and the MSU Intramural Research Grants Program for financial and research support.

†Address: School of Economics and Management and CREATES, University of Aarhus, Building 1322, DK-8000 Aarhus C. Phone: +45 8942 1547. E-mail: bjchristensen@creates.au.dk.

‡Address: School of Economics and Management and CREATES, University of Aarhus, Building 1322, DK-8000 Aarhus C. E-mail: cdahl@econ.au.dk.

§Address: Department of Economics, Michigan State University, 101 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA. E-mail: iglesia5@msu.edu.

parametric approach and the iterative semiparametric approach using a parametric initial estimate proposed by Conrad and Mammen (2008). An empirical application to the daily S&P 500 stock market returns suggests that the linear relation between conditional expected return and conditional variance of returns from the literature is misspecified, and this could be the reason for the disagreement on the sign of the relation.

Key words: Efficiency bound; GARCH-M model; Profile likelihood; Risk-return relation; Semiparametric inference.

JEL Codes: C13, C14, C22, G12.

1 Introduction

The relation between risk and return is of central importance in asset pricing, hedging, derivative pricing, and risk management. The fundamental proposition that asset prices and hence conditional expected returns should reflect investors' willingness to bear risk has been the object of extensive research. Early theoretical and empirical contributions on the risk-return relation were due to Merton (1973, 1980). Merton's intertemporal capital asset pricing model predicts a positive and linear relation between the expectation and the variance of returns. Essentially, investors must be compensated for taking on additional risk. Perhaps surprisingly, both significance and even the sign of the linear relation between expected return and variance of return has proved elusive in empirical work.

In the present paper, we explore the possibility that the mixed empirical evidence may be due to misspecification of the functional form of the risk-return relation. In fact, the conditions for linearity are rather restrictive, and probably unlikely to hold in practice. Instead, we allow for a general nonparametric risk-return tradeoff, and model the conditional variances as a GARCH process. Besides the nonlinear GARCH-in-mean effect, our specification accommodates exogenous regressors that are typically used as conditioning variables entering linearly in the mean equation, such as the dividend yield. We introduce a new semiparametric estimation procedure for the resulting GARCH-in-mean (or GARCH-M) type model that does not rely on an initial parametric (linear) estimate of the risk-return relation, which would necessarily be inconsistent if the true relation is indeed nonlinear, and this feature is the key to establishing the asymptotic properties of our estimator. Using the profile likelihood approach, we prove that our semiparametric estimator is consistent, asymptotically normal, and achieves the semiparametric efficiency bound.

The literature on the risk-return tradeoff is massive. Motivated by Merton's linear relation between expected return and variance, the original GARCH-M model proposed by Engle, Lilien, and Robins (1987) allows for the direct effect of conditional variance on asset prices through required returns in a GARCH-type model by introducing conditional variance into the conditional mean return equation in a linear fashion. Empirical studies of the risk-return tradeoff applying GARCH-type models to stock returns have obtained mixed results regarding both the sign and the significance of this variance-in-mean effect, see e.g. Bollerslev, Engle, and Wooldridge (1988), Chou (1988), Glosten, Jagannathan, and Runkle (1993), Nelson (1991), Campbell and Hentschel (1992), Chou, Engle, and Kane (1992), Backus and Gregory (1993), and Harrison and Zhang (1999). Poterba and Summers (1986) show that the stock market level is determined by the risk-return tradeoff in conjunction with the degree of serial correlation in volatility. Indeed, recent work in asset pricing focussing on volatility innovations examines cross-sectional risk premia induced by covariance between volatility changes and stock returns and finds negative premia, e.g., Ang, Hodrick, Xing, and Zhang (2006). The idea is that since innovations in volatility are higher during recessions, stocks which co-vary with volatility pay off in bad states, and so should require smaller risk premia. Christensen and Nielsen (2007) consider aggregate time series data on returns and innovations in realized as well as option-implied volatility and obtain results consistent with the cross-sectional findings. On the other hand, a positive risk-return tradeoff has been indicated by Brandt and Kang (2004), who use a latent VAR methodology, and Ghysels, Santa-Clara, and Valkanov (2005) using weighted rolling sample windows in the variance measurements. For a survey of these and related studies, see Lettau and Ludvigson (forthcoming).

Thus, there is mixed evidence on the risk-return relation in the literature. One possible source of the disagreement is misspecification of the way in which conditional variance enters the conditional mean return equation. Indeed, already Merton (1980) in his early empirical study regressed returns not only on sample variances, but also on realized sample standard deviations (volatilities) of returns over subintervals, to determine which relation was more stable. The coefficient in the first regression would be interpreted as the rate of relative risk aversion of the representative investor in the intertemporal capital asset pricing if this relation were most stable, whereas the coefficient on volatility in the alternative (square root) version of the regression would be interpreted as the slope of the capital market line or Sharpe ratio if this proved more stable. Considerations and debates of this type have lead to the interest in specifying a more flexible model encompassing these and other alternative risk-return relations. Flexibility, however, comes at the cost of more complicated statistical properties. Hodgson and Vorkink (2003) estimate the density function of a multivariate GARCH-M model in a semiparametric fashion, but do not provide a formal asymptotic theory. Linton and Perron (2003) use a mean equation given by

$$y_t = \mu\left(\sigma_t^2\right) + \varepsilon_t \sigma_t,$$

where y_t is the daily return, $E\left(\varepsilon_t^2 \sigma_t^2 | I_{t-1}\right) = \sigma_t^2$ is the conditional return variance, I_{t-1} denotes the sigma field generated by the information available up to time $t-1$, $\varepsilon_t \sim iid\,(0,1)$, and $\mu(\cdot)$ is a smooth mean function determining the functional form of the risk-return relation. The specification is estimated semiparametrically, using an EGARCH process for the conditional variance, but again no asymptotic theory is provided. As the authors state, "...unfortunately, in our model, we cannot define the corresponding profile quantity $\widehat{\mu}_\phi(\sigma_t^2)$ so easily, since σ_t^2 depends, in addition to the parameters, on lagged $\varepsilon's$, which in turn depend on lagged $\mu's$. Therefore, we need to know the entire function $\mu(.)$ (or at least its values at the T sample points) to construct $\widehat{\mu}_\phi(\sigma_t^2)$." Conrad and Mammen (2008) propose a specification test for GARCH-M models, but they do not show the asymptotic theory for the quasi maximum likelihood (QML) estimator they use as a starting point of their iterative procedure. In fact, this initial estimator would be inconsistent if the true risk-return relation were nonlinear. Sun and Stengos (2006) propose yet another type of semiparametric GARCH-M model that like our specification is closely related to the original GARCH-M of Engle, Lilien, and Robins (1987).

The model of the present paper is an extension of the double autoregressive model of Ling (2004) and is amenable to asymptotic analysis based on the profile likelihood methodology, along the lines of Severini and Wong (1992). We provide an estimator that reaches the semiparametric lower bound. Our estimation procedure is easy to apply and readily allows calculation of consistent standard errors.

Section 2 below presents our new model and semiparametric estimator, and we state conditions under which our estimator is consistent, asymptotically normal, and attains the semiparametric lower bound. Section 3 describes in detail our semiparametric estimation algorithm. In a sampling experiment we explore finite sample accuracy and compare with the parametric approach and the iterative approach with parametric initial estimate proposed by Conrad and Mammen (2008). Finally, an empirical application to the daily S&P 500 stock market returns suggests that the linear relation between conditional expected return and conditional variance of returns from the literature is misspecified, and this could be the reason for the disagreement on the significance and sign of the relation. Section 4 concludes. Appendix 1 collects the proofs of lemmas and theorems. Appendix 2 contains Figures and Tables.

2 Asymptotic Theory: The Semiparametric Lower Bound

We extend the model of Ling (2004) to include a general risk-return relation. Writing y_t for the daily returns, the model we specify is

$$y_t = \mu\left(\sigma_t^2\right) + \varepsilon_t \sigma_t, \tag{1a}$$

$$\sigma_t^2 = \omega + \gamma y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{1b}$$

where $E\left(\varepsilon_t^2\sigma_t^2|I_{t-1}\right) = \sigma_t^2$ is the conditional variance of return, I_{t-1} denotes the sigma field generated by the information available up to time $t-1$, and $\varepsilon_t \sim iid\,(0,1)$ (see assumption A1 below). We write $\phi = (\omega, \gamma, \beta)'$, and $\mu(\cdot)$ is a smooth conditional mean function. In contrast to Linton and Perron (2003) and Conrad and Mammen (2008), we follow Ling (2004) and use the squared lagged return y_{t-1}^2 in (1b), rather than the squared lagged innovation ε_{t-1}^2 as in the traditional GARCH(1,1) equation $\sigma_t^2 = \omega + \gamma\varepsilon_{t-1}^2\sigma_{t-1}^2 + \beta\sigma_{t-1}^2$. Linton and Perron (2003) do not show any asymptotic results, and although Conrad and Mammen (2008) propose an algorithm and a test for GARCH-M effects using QML to get starting values, they actually require a consistent estimator for the starting values (e.g., in their assumption 5), and the QML estimator is necessarily inconsistent if $\mu(\cdot)$ is indeed nonlinear. Further, the main tool of Conrad and Mammen (2008) is empirical process theory, and this involves high level assumptions such as $E\left[\exp\left(\rho\left|\varepsilon_t\right|\right)\right] < \infty$ for $\rho > 0$. The two main differences between our approach and that of Conrad and Mammen (2008) are that (1) we do not use QML or any other inconsistent estimator as starting value, and (2) instead of empirical process theory we use a profile likelihood approach.

The conditional quasi log-likelihood corresponding to (1a)-(1b) is given by

$$L_T\left(\phi, \mu(\cdot)\right) = \ln l_T\left(\phi, \mu(\cdot)\right) = -\frac{1}{2}\sum_{t=1}^{T}\ln\left(\sigma_t^2\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{\left(y_t - \mu\left(\sigma_t^2\right)\right)^2}{\sigma_t^2},$$

with l_T denoting the conditional quasi likelihood. Note that as ϕ varies, $L_T\left(\phi, \mu(\cdot)\right)$ varies because by (1b) σ_t^2 depends on $\phi = (\omega, \gamma, \beta)'$, for each t. If it were feasible to compute the function, say $\widehat{\mu}_\phi(\cdot)$, that for given ϕ maximizes $L_T\left(\phi, \mu(\cdot)\right)$, then we might consider the profile quasi log-likelihood function $L_T\left(\phi, \widehat{\mu}_\phi(\cdot)\right)$. Maximizing this with respect to ϕ would yield an asymptotically efficient estimator (indeed, the semiparametric MLE). The main idea in the profile likelihood approach to obtaining an efficient semiparametric estimator (see Severini and Wong (1992)) is to first get a consistent estimator $\widehat{\mu}_\phi(\cdot)$ of a least favorable curve $\mu_\phi(\cdot)$. The efficient semiparametric estimator of ϕ is then obtained by maximizing the quasi log-likelihood function $L_T\left(\phi, \widehat{\mu}_\phi(\cdot)\right)$ (indeed, the generalized profile quasi log-likelihood function) with respect to ϕ. Here, the least favorable curve $\mu_\phi(\cdot)$ has the properties that (i) maximizing $L_T\left(\phi, \mu_\phi(\cdot)\right)$ with respect to ϕ yields an asymptotically efficient estimator, and (ii) if $\widehat{\mu}_\phi(\cdot)$ is a consistent estimator of the least favorable curve $\mu_\phi(\cdot)$, the estimators of ϕ obtained by maximizing $L_T\left(\phi, \mu_\phi(\cdot)\right)$ and $L_T\left(\phi, \widehat{\mu}_\phi(\cdot)\right)$ are asymptotically equivalent.

The marginal Fisher information for ϕ in this semiparametric setting is defined as

$$\begin{aligned} i_\phi &= E_0\left[\left(\frac{\partial L_T}{\partial\phi}\left(\phi_0, \mu_{\phi_0}(\cdot)\right) + \frac{\partial L_T}{\partial\mu_\phi(\cdot)}\left(\phi_0, \mu_{\phi_0}(\cdot)\right)\mu_{\phi_0}'(\cdot)\right)\right. \\ &\quad \left.\times\left(\frac{\partial L_T}{\partial\phi}\left(\phi_0, \mu_{\phi_0}(\cdot)\right) + \frac{\partial L_T}{\partial\mu_\phi(\cdot)}\left(\phi_0, \mu_{\phi_0}(\cdot)\right)\mu_{\phi_0}'(\cdot)\right)'\right], \end{aligned}$$

where $\partial L_T/\partial\mu_\phi(\cdot)$ is the Frechet derivative of $L_T(\phi,\mu_\phi(\cdot))$ with respect to the function $\mu_\phi(\cdot)$, whereas $\mu'_\phi(\cdot)$ is the ordinary derivative $d\mu_\phi(\cdot)/d\phi$, and subscript 0 throughout denotes evaluation at the true parameter. With this, i_ϕ^{-1} gives the semiparametric lower bound on the asymptotic variance of estimators of ϕ, just like the inverse Fisher information provides the Cramer-Rao lower bound in a parametric model.

Following Severini and Wong (1992, Lemma 8) and Su and Jin (2008), we will work with $T_\phi(y_t)$, a scalar function of y_t depending on a ϕ and such that $\mu_\phi(\sigma_t^2)=E[T_\phi(y_t)|I_{t-1}]$. Initially, we will have $T_\phi(y_t)=y_t$, but this will be modified later. We also denote from now on by $f_{\phi j}(y_t|I_{t-1})$ for $j=0,1,2$ the conditional density of

$$T_\phi^{(j)}(y_t)=\frac{\partial^j}{\partial\phi^j}T_\phi(y_t).$$

Finally $f(\sigma_t^2)$ denotes the marginal density of σ_t^2.

For clarity of exposition, we provide our results in four stages. First, we consider the case where $\mu(\cdot)$ is known function and provide the asymptotic theory of the QMLE. Secondly, we consider the case of an unknown least favorable curve $\mu_\phi(\cdot)$ that we estimate nonparametrically. Third, we provide the asymptotic theory of the semiparametric estimator when $\mu_\phi(\cdot)$ is estimated nonparametrically and QML is used in the second stage. The fourth stage allows for the introduction of correctly specified exogenous variables in the mean equation.

2.1 Asymptotic theory: the parametric case

We start by assuming that in (1a)-(1b) we know the mean function $\mu(\cdot)$, e.g., $\mu(\sigma_t^2)$ may be given by $\lambda\sigma_t^2$ or $\lambda\sigma_t$, for known λ.

We now introduce our assumptions. For ease in referring to these in later sections, we carry a subscript ϕ on μ in the notation. Assumption A1 is a collection of conditions related to the disturbance and the conditional variance. They are quite standard in the literature of QML estimation of GARCH processes (see e.g. Berkes and Horváth (2004), Jensen and Rahbek (2004a, 2004b), Straumann and Mikosch (2006) and Dahl and Iglesias (2007)). Assumption A2 gives conditions for $\mu_\phi(\cdot)$ to take values inside the interior of the parameter space. Assumption A3 is a typical assumption needed in the GARCH context as in Jensen and Rahbek (2004a, 2004b) to get local identification.

Assumption A

A1 *σ_t is a strictly stationary process satisfying $\varepsilon_t\sim i.i.d.(0,1)$ with $E(\varepsilon_t^4-1)=\zeta<\infty$, and $E(|\varepsilon_t|^{2r})<\infty$ for some $r>2$. Furthermore, σ_t is a sequence of strong mixing random variables* *with mixing numbers* α_m, $m = 1, 2, \ldots$, *that satisfy* $\alpha_m \leq Cm^{-(4r-2)/(2r-2)-\delta}$ *for positive* C *and* δ, *as* $m \to \infty$. *Moreover,* ω, γ *and* β *are strictly positive,* σ_0^2 *is a finite positive constant,* $\phi \in \Phi$ *and* $\phi_0 \in int\,(\Phi)$. *Finally,* $E\left[\ln\left(\gamma\varepsilon_t^2 + \beta\right)\right] < 0$, *and* $\sigma_0^2(\phi)$ *is a drawing from the stationary distribution.*

A2 *M denotes a compact subset of the real line such that* $\mu_\phi\left(\sigma_t^2\right) \in int\,(M)$ *for all* σ_t^2 *and all* t.

A3 $\mu_\phi(\cdot)$ *is three times differentiable with*

$$E\left|\frac{\left(y_t - \mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}}{\sigma_t^2}\right| < \infty, \quad E\left|\frac{\left(y_t - \mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}\right| < \infty,$$

$$E\left|\frac{\left(y_t - \mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^2}\right| < \infty.$$

We also assume that there are constants C_1, C_2, C_3, C_4, C_5 *and* C_6 *so that*

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\right)^2}{\sigma_t^2}\Big|_{\phi=\phi_0} \xrightarrow{p} C_1 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)^2}{\sigma_t^2}\Big|_{\phi=\phi_0} \xrightarrow{p} C_2 < \infty,$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\right)^2}{\sigma_t^2}\Big|_{\phi=\phi_0} \xrightarrow{p} C_3 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}\Big|_{\phi=\phi_0} \xrightarrow{p} C_4 < \infty,$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}\Big|_{\phi=\phi_0} \xrightarrow{p} C_5 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^2}\Big|_{\phi=\phi_0} \xrightarrow{p} C_6 < \infty.$$

Also, there exists a neighborhood $N(\phi_0)$ *given by (12) in Definition 1 for which*

$$(a) \quad \sup_{\phi\in N(\phi_0)}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}}{\sigma_t^2}\right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{11t},$$

$$(b) \quad \sup_{\phi\in N(\phi_0)}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial i}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4}\right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{12t},$$

$$(c) \quad \sup_{\phi\in N(\phi_0)}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial^2\sigma_t^2}{\partial i\partial j}}{\sigma_t^4}\right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{13t},$$

$$(d) \qquad \sup_{\phi\in N(\phi_0)}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial\sigma_t^2}{\partial i}\frac{\partial\sigma_t^2}{\partial j}}{\sigma_t^6}\right|\leq\frac{1}{T}\sum_{t=1}^{T}w_{14t},$$

$$(e) \qquad \sup_{\phi\in N(\phi_0)}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4}\right|\leq\frac{1}{T}\sum_{t=1}^{T}w_{15t},$$

$$(f) \qquad \sup_{\phi\in N(\phi_0)}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial^3\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j\partial k}}{\sigma_t^2}\right|\leq\frac{1}{T}\sum_{t=1}^{T}w_{16t},$$

where $w_{11t},...,w_{16t}$ *are stationary and have finite moments,* for any $i,j,k=\omega,\gamma,\beta$, $E\left(w_{lt}\right)=M_l<\infty,\forall l=11,...,16$. *Finally,* $\frac{1}{T}\sum_{t=1}^{T}w_{lt}\xrightarrow{a.s.}M_l,\forall l=11,...,16$.

Let $\widehat{\phi}$ be the quasi-maximum likelihood (QML) estimator of ϕ. The following theorems provide the first asymptotic theory for parametric GARCH-M models. It should be noted that this is for our type of variance specification (1b).

THEOREM A1. *Let* $\mu_\phi(\cdot)$ *denote a known correctly specified curve. Further, for each for each* T *define* $\widehat{\phi}\equiv\widehat{\phi}_T$ *to be any element of* $int(\Phi)$ *satisfying*

$$L_T\left(\widehat{\phi},\mu_{\widehat{\phi}}(\cdot)\right)=\sup_{\phi\in\Phi}L_T\left(\phi,\mu_\phi(\cdot)\right),$$

then under assumption A

$$\widehat{\phi}\xrightarrow{p}\phi_0$$

as $T\to\infty$.

Theorem A1 shows consistency. Asymptotic normality for known mean function is established next.

THEOREM B1. *Suppose that Assumption A holds. Then (a)*

$$\sqrt{T}(\widehat{\phi}-\phi_0)\xrightarrow{d}N(0,i_\phi^{-1}),$$

where i_ϕ *is given in Lemmas 2-3 in Appendix 1. Finally, let*

$$\widehat{i}_\phi=-T^{-1}\frac{\partial^2L_T}{\partial\phi\partial\phi'}\left(\phi,\mu_\phi(\cdot)\right)|_{\phi=\widehat{\phi}},$$

then (b)

$$\widehat{i}_\phi\xrightarrow{p}i_\phi \text{ as } T\to\infty.$$

Proofs of Theorems A1 and B1. The proofs proceed by the classic Cramér type conditions for consistency and asymptotic normality (central limit theorem for the score, convergence of the Hessian and uniformly bounded third-order derivatives, see Lehmann (1999)) established e.g. as in Jensen and Rahbek (2004a, 2004b). Detailed proofs are provided in Appendix 1.■

2.2 Nonparametric estimation of least favorable curve

We now relax the assumption that $\mu(\cdot)$ is known and we estimate this function nonparametrically. We need to add Assumption B below. Assumptions B1-B6 give conditions for the nonparametric estimator of the least favorable curve given ϕ, along the lines of the smoothness and nuisance parameter conditions of Severini and Wong (1992). Note that in our case, $T_\phi(y_t) = y_t$.

Assumption B

B1 $E\left\{\sup_\phi \left|T_\phi^{(j)}(y_t)\right|\right\} < \infty$, $j = 0, 1, 2, 3$ *for any* t.

B2 *For some even integer* $v \geq 10$, $\sup_\phi E\left\{\left|T_\phi^{(j)}(y_t)\right|^v\right\} < \infty$, $j = 0, 1, 2$ *for any* t.

B3 $\sup_\phi \sup_{y_t, y_{t-1}} \left|f_{\phi j}^{(r)}(y_t|I_{t-1})\right| < \infty$, $j = 0, 1, 2$; $r = 0, 1, 2, 3, 4$ *for any* t. $f_{\phi j}(y_t|I_{t-1})$ *belongs conditionally to the exponential family.*

B4 $\sup_{y_t} \left|f^{(r)}(\sigma_t^2)\right| < \infty$, $r = 0, 1, 2, 3, 4$ *for any* t.

B5 $0 < \inf_{y_t} f(\sigma_t^2) \leq \sup_{y_t} f(\sigma_t^2) < \infty$ *for any* t.

B6 $K(\cdot)$ *satisfies* $\int K(u)\,du = 1$, $\int uK(u)\,du = 0$, $\int u^2 K(u)\,du < \infty$, $\sup_u \left|K^{(r)}(u)\right| < \infty$, $r = 0, ..., 4$ *and* h_T *is a sequence of constants satisfying* $h_T \to 0$ *as* $T \to \infty$. $h_T = O\left(T^{-\overline{\alpha}}\right)$ *with* $1/8 < \overline{\alpha} < (v-2)/(4(v+4))$.

Lemma 1 provides the result on nonparametric estimation of a least favorable curve. The proof is given in Appendix 1.

LEMMA 1. *Suppose that assumptions A2 and B hold. Define* $\widehat{\mu}_\phi(\cdot)$ by

$$\widehat{\mu}_\phi\left(\sigma^2\right) = \frac{\sum_j y_j K\left(\left(\sigma^2 - \sigma_j^2\right)/h_T\right)}{\sum_j K\left(\left(\sigma^2 - \sigma_j^2\right)/h_T\right)}.$$

Then $\widehat{\mu}_\phi(\cdot)$ *is a consistent estimator of a least favorable curve.*

Note that $\widehat{\mu}_\phi(\cdot)$ depends on $\sigma_1^2, ..., \sigma_T^2$, and hence on ϕ, by (1b). The result is elegant since the estimator takes the simple Nadaraya-Watson form, yet it is consistent for a least favorable curve in the profile likelihood approach.

2.3 Asymptotic theory: the semiparametric case

We consider now the case that in (1a)-(1b) we know neither the parameter ϕ nor the function $\mu(\cdot)$. We combine the results of the previous subsections on estimation of ϕ given $\mu(\cdot)$, and the estimation of $\mu(\cdot)$ given ϕ. We need to add an extra assumption C. Note that in this case, assumptions B1-B6 can be relaxed in case we are not interested in reaching the semiparametric efficiency lower bound, but only in asymptotic normality and consistency. Assumption C1 is a standard condition corresponding to the global identifiability condition of Severini and Wong (1992). It is needed in order to get a consistent estimator of the least favorable curve as well as asymptotic normality in the second stage, and also to reach the semiparametric lower bound.

Assumption C

C1 *(a) For fixed but arbitrary $\phi_1 \in int(\Phi)$ and $\mu_{1\phi}(\cdot)$* with $\mu_{1\phi}(\sigma_t^2) \in int(M)$ *for all σ_t^2 we define*

$$\varphi(\phi, \mu_\phi(\cdot)) = \int l_T(y, \phi, \mu_\phi(\cdot)) f\left(y; \phi_1, \mu_{1\phi}\left(\sigma_t^2\right)\right) dy,$$

for all $\phi \in int(\Phi)$, all $\mu_\phi(\cdot)$ with $\mu_\phi(\sigma_t^2) \in int(M)$. *If $\phi \neq \phi_1$, then*

$$\varphi(\phi, \mu_\phi(\cdot)) < \varphi(\phi_1, \mu_{1\phi}(\cdot)).$$

(b) Let $\widetilde{i}_\phi(\phi, \mu_\phi(\cdot))$ denote the marginal Fisher information for ϕ in the parametric model, that is

$$\begin{aligned}
\widetilde{i}_\phi(\phi, \mu_\phi(\cdot)) \quad = \quad & E_{\phi,\mu_\phi(\cdot)}\left(\frac{\partial L_T}{\partial \phi}(y; \phi, \mu_\phi(\cdot))\left(\frac{\partial L_T}{\partial \phi}(y; \phi, \mu_\phi(\cdot))\right)'\right) \\
& -E_{\phi,\mu_\phi(\cdot)}\left[\left(\frac{\partial L_T}{\partial \phi}(y; \phi, \mu_\phi(\cdot))\frac{\partial L_T}{\partial \mu_\phi(\cdot)}(y; \phi, \mu_\phi(\cdot))\right)\right. \\
& \left.\times\left(\frac{\partial L_T}{\partial \phi}(y; \phi, \mu_\phi(\cdot))\frac{\partial L_T}{\partial \mu_\phi(\cdot)}(y; \phi, \mu_\phi(\cdot))\right)'\right] \\
& E_{\phi,\mu_\phi(\cdot)}\left(\frac{\partial L_T}{\partial \phi}(y; \phi, \mu_\phi(\cdot))\frac{\partial L_T}{\partial \phi}(y; \phi, \mu_\phi(\cdot))'\right)^{-1}.
\end{aligned}$$

Assume that $\widetilde{i}_\phi(\phi, \mu_\phi(\cdot)) > 0$ for all $\phi \in int(\Phi)$,all $\mu_\phi(\cdot)$ with $\mu_\phi(\sigma_t^2) \in int(M)$.

We now provide the asymptotic theory combining the estimators $\widehat{\phi}$ and $\widehat{\mu}_\phi(\cdot)$, with the latter estimating consistently a least favorable curve in the profile likelihood framework.

Theorem A2. *Let $\widehat{\mu}_{\phi}(\cdot)$ be a consistent estimator of a least favorable curve $\mu_{\phi}(\cdot)$, as given in Lemma 1. Further, for each T, define $\widehat{\phi} \equiv \widehat{\phi}_T$ to be any element of $int(\Phi)$ satisfying*

$$L_T\left(\widehat{\phi}, \widehat{\mu}_{\widehat{\phi}}(\cdot)\right) = \sup_{\phi\in\Phi} L_T\left(\phi, \widehat{\mu}_{\phi}(\cdot)\right).$$

T*hen, under assumptions A, B, and C*

$$\widehat{\phi} \xrightarrow{p} \phi_0$$

as $T \to \infty$.

Theorem A2 shows consistency of the estimator of ϕ obtained by maximizing the generalized profile log-likelihood.

Theorem B2. *Suppose that assumptions A, B, and C hold. Then (a)*

$$\sqrt{T}(\widehat{\phi} - \phi_0) \xrightarrow{d} N(0, i_{\phi}^{-1}),$$

where i_{ϕ} is given in Lemmas 2-3 in Appendix 1. Finally, let

$$\widehat{i}_{\phi} = -T^{-1}\frac{\partial^2 L_T}{\partial\phi\partial\phi'}\left(\phi, \widehat{\mu}_{\phi}(\cdot)\right)|_{\phi=\widehat{\phi}},$$

where $\widehat{\mu}_{\phi}(\cdot)$ is given in Lemma 1, then (b)

$$\widehat{i}_{\phi} \xrightarrow{p} i_{\phi} \text{ as } T \to \infty.$$

Proofs of Theorems A2 and B2. The proofs proceed as those for Theorems A1 and B1, but with $\mu_{\phi}(\cdot)$ replaced by $\widehat{\mu}_{\phi}(\cdot)$, based on the asymptotic equivalence result of Severini and Wong (1992) stated in Corollary 1 in Appendix 1.■

Theorem B2 shows that the semiparametric estimator is asymptotically normal and achieves the semiparametric lower bound. In addition, part (b) provides for the consistent estimation of standard errors using the second derivative matrix of the generalized profile log-likelihood.

2.4 Exogenous conditioning variables in the mean equation

Here, we generalize the semiparametric GARCH-M model to accommodate exogenous regressors since conditioning variables typically are entered linearly into the mean equation, e.g., the dividend yield. We extend the theory from (1a)-(1b) to the model given by

$$y_t = x_t'\alpha + \mu\left(\sigma_t^2\right) + \varepsilon_t\sigma_t, \tag{2a}$$
$$\sigma_t^2 = \omega + \gamma y_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{2b}$$

where x_t is a vector of pre-determined regressors that are correctly specified in the mean equation and α is $p \times 1$ vector of parameters. We avoid identification issues by not allowing lagged dependent variables or an intercept in x_t. Note that now $\phi = (\alpha', \omega, \gamma, \beta)'$ and $T_\phi(y_t) = y_t - x_t'\alpha$. We now extend Theorems A2 and B2 to allow for exogenous variables, and to this end we strengthen our assumptions as follows. As in Assumption A, we carry a subscript ϕ on μ in the notation.

Assumption D

D1 Let $\mu_\phi(\cdot)$ *denote a smooth mean function that is three times differentiable with*

$$E\left|\frac{\left(y_t - \mu_\phi\left(\sigma_t^2\right) - x_t'\alpha\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}}{\sigma_t^2}\right| < \infty, \quad E\left|\frac{\left(y_t - \mu_\phi\left(\sigma_t^2\right) - x_t'\alpha\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}\right| < \infty,$$
$$E\left|\frac{\left(y_t - \mu_\phi\left(\sigma_t^2\right) - x_t'\alpha\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^2}\right| < \infty, \quad E\left|\frac{\left(y_t - \mu_\phi\left(\sigma_t^2\right) - x_t'\alpha\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\alpha}}{\sigma_t^2}\right| < \infty.$$

We also assume that there are constants C_1, C_2,...,C_{10} *so that*

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\right)^2}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_1 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)^2}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_2 < \infty,$$
$$\frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\right)^2}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_3 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_4 < \infty,$$
$$\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_5 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_6 < \infty,$$
$$\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\alpha}\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\alpha}\right)'}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_7 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\alpha}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_8 < \infty,$$
$$\frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\alpha}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_9 < \infty, \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\alpha}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^2}\Big|_{\phi=\phi_0}\xrightarrow{p} C_{10} < \infty.$$

Also, there exists a neighborhood $N(\phi_0)$ *given by (12) in Definition 1 for which*

$$(a) \quad \sup_{\phi\in N(\phi_0)} \left| \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}}{\sigma_t^2} \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{11t},$$

$$(b) \quad \sup_{\phi\in N(\phi_0)} \left| \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial i}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4} \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{12t},$$

$$(c) \quad \sup_{\phi\in N(\phi_0)} \left| \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial^2\sigma_t^2}{\partial i\partial j}}{\sigma_t^4} \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{13t},$$

$$(d) \quad \sup_{\phi\in N(\phi_0)} \left| \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial\sigma_t^2}{\partial i}\frac{\partial\sigma_t^2}{\partial j}}{\sigma_t^6} \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{14t},$$

$$(e) \quad \sup_{\phi\in N(\phi_0)} \left| \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4} \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{15t},$$

$$(f) \quad \sup_{\phi\in N(\phi_0)} \left| \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{\partial^3\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j\partial k}}{\sigma_t^2} \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{16t},$$

where $w_{11t}, ..., w_{16t}$ are stationary and have finite moments, for any $i, j, k = \omega, \gamma, \beta, \alpha'$, $E(w_{lt}) = M_l < \infty, \forall l = 11, ..., 16$. *Finally,* $\frac{1}{T}\sum_{t=1}^{T} w_{lt} \xrightarrow{a.s.} M_l, \forall l = 11, ..., 16$.

D2 *Define $\widehat{\mu}_\phi(\cdot)$ by*

$$\widehat{\mu}_\phi\left(\sigma^2\right) = \frac{\sum_j \left(y_j - x_j'\alpha\right) K\left(\left(\sigma^2 - \sigma_j^2\right)/h_T\right)}{\sum_j K\left(\left(\sigma^2 - \sigma_j^2\right)/h_T\right)}.$$

We now provide the asymptotic theory for the generalized semiparametric model.

Theorem A3. *Let $\widehat{\mu}_\phi(\sigma_t^2)$ be the consistent estimator from D2 of a least favorable curve $\mu_\phi(\cdot)$ with $\mu_\phi(\sigma_t^2) \in int(M)$ for all σ_t^2 and all t, and define for each T $\widehat{\phi} \equiv \widehat{\phi}_T = \left(\widehat{\alpha}_T', \widehat{\omega}_T, \widehat{\gamma}_T, \widehat{\beta}_T\right)'$ in (2a)-(2b) to be any element of $int(\Phi)$ satisfying*

$$L_T\left(\widehat{\phi}, \widehat{\mu}_{\widehat{\phi}}(\cdot)\right) = \sup_{\phi\in\Phi} L_T\left(\phi, \widehat{\mu}_\phi(\cdot)\right).$$

Then under assumptions A, B, C, and D

$$\widehat{\phi} \xrightarrow{p} \phi_0$$

as $T \to \infty$.

By Theorem A3, the maximizer of the generalized profile log-likelihood remains consistent in the presence of exogenous regressors.

Theorem B3. *Suppose that assumptions A, B, C, and D hold in (2a)-(2b). Then (a)*

$$\sqrt{T}(\widehat{\phi}-\phi_0) \xrightarrow{d} N(0, i_\phi^{-1}),$$

where i_ϕ is given in Lemmas 2-3 in Appendix 1. Finally, let

$$\widehat{i}_\phi = -T^{-1}\frac{\partial^2 L_T}{\partial\phi\partial\phi'}\left(\phi, \widehat{\mu}_\phi(\cdot)\right)|_{\phi=\widehat{\phi}},$$

where $\widehat{\mu}_\phi(\cdot)$ is an estimator of $\mu_\phi(\cdot)$ as given in D2, then (b)

$$\widehat{i}_\phi \xrightarrow{p} i_\phi \text{ as } T \to \infty.$$

Proofs of Theorems A3 and B3. The proofs proceeds as those for Theorems A2 and B2 except that we need the extra derivatives with respect to the α vector to take into account the introduction of exogenous variables. See Appendix 1 for details.■

Theorem B3 shows that asymptotic normality carries over to the case with exogenous regressors. Again, the semiparametric efficiency bound i_ϕ^{-1} is achieved, and part (b) provides consistent standard errors based on the second derivative matrix of the generalized profile log-likelihood function.

3 Illustrations

We begin this section by describing in detail our semiparametric estimation algorithm. Secondly, we carry out a simulation study on the accuracy of the new approach and provide comparisons to the parametric approach as well as the iterative approach with parametric initial estimate proposed by Conrad and Mammen (2008). Finally, we present an empirical illustration, applying our method to the daily S&P 500 stock index return series in order to analyze the shape of the classical risk-return relation.

3.1 Estimation of conditional mean and variance

In the general setting, we wish to estimate the unknown function $\mu(\cdot)$ as well as all the unknown parameters $\phi = (\alpha', \omega, \gamma, \beta)'$ in the representation of our generalized semiparametric GARCH(1,1)-M model given by

$$\begin{aligned} y_t - x_t'\alpha &= \mu(\sigma_t^2) + e_t, & (3)\\ e_t &= \varepsilon_t\sigma_t, & (4)\\ \sigma_t^2 &= \omega + \gamma y_{t-1}^2 + \beta\sigma_{t-1}^2, & (5)\\ z_t &= y_t - x_t'\alpha, & (6) \end{aligned}$$

corresponding to (2a)-(2b), where x_t is a vector of pre-determined regressors without an intercept or lagged dependent variables. Consider first the case where σ_t^2 is observed (i.e., the parameters $\theta = (\omega, \gamma, \beta)'$ are known) and we are interested in estimating $\mu(\cdot)$ and α, only. The kernel based estimator of $\mu(\cdot)$ given α in this case is given by

$$\begin{aligned} \widehat{\mu}(\sigma^2) &= \sum_j w_j\left(\sigma^2\right) z_j \\ &= \sum_j w_j\left(\sigma^2\right) y_j - \sum_j w_j\left(\sigma^2\right) x_j' \alpha, \end{aligned} \tag{7}$$

where

$$w_j\left(\sigma^2\right) = \frac{K\left(\left(\sigma^2 - \sigma_j^2\right)/h_T\right)}{\sum_{s=1}^T K\left(\left(\sigma^2 - \sigma_s^2\right)/h_T\right)} \tag{8}$$

and again $K(\cdot)$ and h_T denote the kernel function and bandwidth, respectively. Plugging the kernel estimator back into (3) yields

$$y_t = x_t'\alpha + \widehat{\mu}(\sigma_t^2) + e_t, \tag{9}$$

which may conveniently be written as

$$y_t - \sum_j w_j\left(\sigma_t^2\right) y_j = \left(x_t - \sum_j w_j\left(\sigma_t^2\right) x_j\right)' \alpha + \widehat{e}_t. \tag{10}$$

From (10), an estimator of α is easily obtained by QMLE (or simply WLS - weighted least squares).

Consider next the case where we do not observe σ_t^2 and we need to estimate $\phi = (\alpha', \omega, \gamma, \beta)'$. The derivations above suggest the following iterative estimation procedure.

Step 1: Provide a set of initial parameters ($i = 0$) $\widehat{\theta}^{(i)}$ and compute $\widehat{\sigma}_t^{2(i)} = \sigma_t^2\left(\widehat{\theta}^{(i)}\right)$ for $t = 1, 2, ..., T$ by iterating on equation (5).

Step 2: Based on the sequence $\left\{\widehat{\sigma}_t^{2(i)}\right\}_{t=1}^T$, compute first $\widehat{\alpha}^{(i)}$ based on (10), and secondly $\widehat{\mu}_t^{(i)} = \widehat{\mu}(\widehat{\sigma}_t^{2(i)})$ for $t = 1, 2, ..., T$ from (7).

Step 3: Update $\widehat{\theta}^{(i)}$ and $\left\{\widehat{\sigma}_t^{2(i)}\right\}_{t=1}^T$ (i.e., find $\widehat{\theta}^{(i+1)} = \left(\widehat{\omega}, \widehat{\gamma}, \widehat{\beta}\right)'$ and consequently $\left\{\widehat{\sigma}_t^{2(i+1)}\right\}_{t=1}^T$) by performing QML on the GARCH(1,1) model

$$\begin{aligned} \widehat{y}_t^{(i)} &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \gamma y_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

where

$$\widehat{y}_t^{(i)} \equiv y_t - x_t' \widehat{\alpha}^{(i)} - \widehat{\mu}_t^{(i)}.$$

Step 4: Repeat Steps 2 and 3 for a finite fixed number of iterations or until convergence.

In the simulations we used a fixed number of iterations. For a convergence criterion one could use the suggestions by Linton and Perron (2004, page 357) or Conrad and Mammen (2008, page 19). Alternatively, Steps 2 and 3 could be replaced by the following.

Step 2': Conditional on the sequence $\left\{\widehat{\sigma}_t^{2(i)}\right\}_{t=1}^{T}$ and

$$\begin{aligned}\widetilde{y}_t^{(i)} &= y_t - \sum_j w_j\left(\widehat{\sigma}_t^{2(i)}\right) y_j,\\ \widetilde{x}_t^{(i)} &= \left(x_t - \sum_j w_j\left(\widehat{\sigma}_t^{2(i)}\right) x_j\right),\end{aligned}$$

update $\widehat{\theta}^{(i)}$ and $\widehat{\alpha}^{(i)}$ (i.e., find $\widehat{\theta}^{(i+1)} = \left(\widehat{\omega}, \widehat{\gamma}, \widehat{\beta}\right)'$ and $\widehat{\alpha}^{(i+1)} = \widehat{\alpha}$) by performing QML on the GARCH(1,1) model

$$\begin{aligned}\widetilde{y}_t^{(i)} &= \widetilde{x}_t^{(i)\prime}\alpha + \varepsilon_t\sigma_t,\\ \sigma_t^2 &= \omega + \gamma y_{t-1}^2 + \beta\sigma_{t-1}^2.\end{aligned}$$

Step 3': Based on $\widehat{\theta}^{(i+1)}$, update the sequences $\left\{\widehat{\sigma}_t^{2(i+1)}\right\}_{t=1}^{T}$ and $\left\{\widetilde{y}_t^{(i+1)}, \widetilde{x}_t^{(i+1)}\right\}_{t=1}^{T}$.

Under the alternative algorithm, $\widehat{\mu}(\widehat{\sigma}_t^2)$ is computed in the final iteration ($i = I$) simply as $\widehat{\mu}(\widehat{\sigma}_t^2) = \widehat{\mu}(\widehat{\sigma}_t^{2(I)})$. For the choice of bandwidth parameter we recommend using cross validation, following Linton and Perron (2003). The quasi log-likelihood function is maximized at each point on a grid and the leave-one-out maximizer is chosen.

In order to compute nonparametric confidence bands, we suggest the following wild bootstrap algorithm, which is similar to the one described in Linton and Perron (2003).

Step 1: Based on the estimated parameters/functions $\widehat{\phi} = (\widehat{\omega}, \widehat{\gamma}, \widehat{\beta})'$, $\widehat{\sigma}_t^2$, $\widehat{\mu}(\widehat{\sigma}_t^2)$, and $\widehat{\varepsilon}_t = \left(y_t - \widehat{\mu}(\widehat{\sigma}_t^2)\right)/\widehat{\sigma}_t$, calculate $\varepsilon_t^c = \widehat{\varepsilon}_t - T^{-1}\sum\widehat{\varepsilon}_t$.

Step 2: Let z_t be a discrete variable taking the values -1 and 1. Draw a random sample $(z_1, z_2, ..., z_T)$ and construct the sequence $\varepsilon_t^* = \varepsilon_t^c z_t$ for $t = 1, 2, ..., T$.

Step 3: Given initial starting values for y_0 and $\widehat{\sigma}_0^2$, define recursively

$$\begin{aligned}(\widehat{\sigma}_t^*)^2 &= \widehat{\omega} + \widehat{\gamma}\left(y_{t-1}^*\right)^2 + \widehat{\beta}\left(\widehat{\sigma}_{t-1}^*\right)^2, \\ y_t^* &= \widehat{\mu}((\widehat{\sigma}_t^*)^2 ; \{\widehat{\sigma}_s^2\}_{s=1}^T, h) + \widehat{\sigma}_t^* \varepsilon_t^*,\end{aligned}$$

for $t = 1, 2, ..., T$.

Step 4: Given the bootstrapped sequence $\{y_t^*\}_{t=1}^T$, calculate $\widehat{\phi}$ and $\{\widehat{\mu}_t^*\}_{t=1}^T$ by the the proposed semiparametric estimation algorithm.

Step 5: Repeat steps 2-4 m times. The pointwise α and $1-\alpha$ percent confidence bands around $\widehat{\mu}(\widehat{\sigma}_t^2)$ are then constructed as the α and $1-\alpha$ empirical quantiles of the m bootstrapped estimates of $\widehat{\mu}(\widehat{\sigma}_t^2)$. As a by-product, standard errors of $\widehat{\phi}$ are estimated from the sample standard deviation of the m bootstrapped estimates $\widehat{\phi}^*$.

3.2 Monte Carlo Simulation

In this section we examine the finite sample properties of our semiparametric GARCH(1,1)-M estimation procedure. We compare the performance of our approach to the performance of the parametric GARCH(1,1)-M model, as well as to the semiparametric approach suggested recently by Conrad and Mammen (2008), and for comparison purposes we work with the data generating processes suggested by these authors, given as

$$\begin{aligned}N1 &: \ \mu(\sigma_t^2) = 0.05\sigma_t^2, \\ N2 &: \ \mu(\sigma_t^2) = 0.5\sigma_t^2, \\ N3 &: \ \mu(\sigma_t^2) = \sigma_t^2, \\ A1 &: \ \mu(\sigma_t^2) = \sigma_t^2 + 0.5\sin(10\sigma_t^2), \\ A2 &: \ \mu(\sigma_t^2) = 0.5\sigma_t^2 + 0.1\sin(0.5 + 20\sigma_t^2), \\ A3 &: \ \mu(\sigma_t^2) = \sigma_t^2 + 0.12\sin(3 + 30\sigma_t^2),\end{aligned}$$

with $\theta = \left(\omega, \gamma, \beta^{(CM)}\right)' = (0.01, 0.1, 0.85)'$, i.e.,

$$\sigma_t^{2(CM)} = 0.01 + 0.1\varepsilon_{t-1}^2 + 0.85\sigma_{t-1}^2.$$

Since our volatility process (1b) is defined differently from that in Conrad and Mammen (2008), we cannot use the exact same value for θ and obtain processes with the same properties as theirs. In order to obtain comparable processes we fix $\omega = 0.01, \gamma = 0.1$, and set β such that

$$\begin{aligned}E(\sigma_t^2) &= E(\sigma_t^{2(CM)}), \\ E(\mu(\sigma_t^2)^2) &= E(\mu(\sigma_t^{2(CM)})^2).\end{aligned}$$

This yields the relationship

$$\begin{aligned}\beta &= \beta^{(CM)} - \left(1-\gamma-\beta^{(CM)}\right)\omega^{-1}\gamma E(\mu(\sigma_t^{2(CM)})^2)) \\ &= \beta^{(CM)} - \gamma\frac{E(\mu(\sigma_t^{2(CM)})^2)}{E(\sigma_t^{2(CM)})},\end{aligned}$$

where $E(\mu(\sigma_t^{2(CM)})^2)$ is obtained by simulation. By this procedure, our GARCH-M specifications are based on the following values of $\beta : \beta(N1) = 0.85$, $\beta(N2) = 0.84$, $\beta(N3) = 0.82$, $\beta(A1) = 0.68$, $\beta(A2) = 0.84$, $\beta(A3) = 0.82$. The residual ϵ_t is drawn form a standard normal distribution.

[Table 1 about here]

Table 1 presents the medians as well as the 25% and 75% quantiles of the estimated parameters in 200 replications based on the parametric GARCH(1,1)-M approach with $\mu(\sigma_t^2) = \mu + \lambda\sigma_t^2$. As expected, the median parameter estimates under the data generating processes N1-N3 are very close to the true parameter values. As noted by Conrad and Mammen (2008), the in-mean parameter λ is relative well estimated when the model is correctly specified (N1-N3), particularly when the true value is relatively large (as in N3 where $\lambda = 1$).

[Table 2 about here]

Table 2 presents the corresponding results based on our semiparametric GARCH(1,1)-M estimator. From the table it is clear that the semiparametric estimator yields very precise estimates of the conditional variance function parameters $\theta = (\omega, \gamma, \beta)'$ under N1-N3. The semiparametric estimate of γ is slightly lower than the parametric estimate, while the estimates of ω and β are practically identical. Furthermore, the ranges between the 25% and 75% quantiles are approximately the same, indicating that the precisions of the semiparametric estimates are comparable to the parametric maximum likelihood estimates.

[Figure 1 about here]

Figure 1 shows the true mean functions and the pointwise median of the parametric and semiparametric estimates, along with the pointwise 25% and 75% quantiles of the semi parametric estimate. Under the models N1, N2, and N3, both estimation procedures seem to do equally well in recovering the true structure of the conditional mean function.

Table 1 also presents the mean and variance parameter estimates from the parametric GARCH(1,1)-M approach with $\mu(\sigma_t^2) = \mu + \lambda\sigma_t^2$ applied to the data generating processes A1, A2, and A3, i.e., the model is misspecified. While the estimates of the variance function parameters are perhaps surprisingly accurate, the mean function parameters are clearly inconsistently estimated. In particular, using the parametric GARCH(1,1)-M model to approximate A1 would lead to a significantly negative estimate of λ, thus falsely indicating a negative risk–return relationship.

In Table 2 similar results are presented based on the semiparametric approach. Again the estimated variance function parameters are close to their true values and quite accurately estimated.

[Figure 2 about here]

Figure 2 reveals that the semiparametric estimate of the mean function again performs very well in uncovering the true mean function. The parametric estimate which is restricted to be linear fails to do so.

[Table 3 about here]

Finally, we use the Conrad and Mammen (2008) semiparametric GARCH(1,1)-M approach adapted to our model (1a)-(1b). This approach is essentially identical to our approach, except that it is based on an initial estimate of σ_t^2 from a parametric GARCH(1,1)-M specification. The results are presented in Table 3 and show that if the Conrad and Mammen (2008) approach is iterated until convergence, it provides exactly the same results as our approach. This feature is expected asymptotically, but is perhaps a bit surprising in finite samples. The result is a consequence of the fact that the parametric variance function estimates are quite robust to mean function misspecification, as illustrated in Table 1. However, the extent to which this result generalizes to other in-mean functional forms is unknown. To be sure, the asymptotic theory, which we establish for our approach, has not been established for the case of an initial estimate of σ_t^2 based on a parametric specification.

3.3 Empirical Illustration

As a further illustration, we apply the semiparametric GARCH(1,1)-M estimator to the daily returns on the Standard and Poors (S&P) 500 stock market index from January 2, 1990 to December 31, 2007, for a total of $T = 4,537$ daily returns in the time series. The continuously compounded cum dividend returns are defined as $y_t = \log(p_t + d_t) - \log(p_{t-1})$, where p_t is the closing index level and d_t the dividend paid to the index on day t, obtained from CRSP.

In addition to the stock returns, we also collect daily data for the same period on a set of four explanatory variables, denoted x_t, that have been used in the literature as state variables for conditional mean returns, namely, the dividend yield, term spread, default spread, and momentum. The dividend yield or dividend-price ratio has been used for modelling conditional mean returns by

Campbell and Shiller (1988a, 1988b), Fama and French (1988, 1989), and many others. We compute it as the sum $\sum_{j>0} d_{t-j}$ of dividends to the index over the 12 month period up to and including date $t-1$ divided by p_{t-1}, so that it is known at the beginning of the time interval over which the return y_t is realized. The term spread or yield curve slope has been used for conditioning mean returns by Campbell (1987) and Fama and French (1988, 1989), among others. It was also used by Campbell (1987) to condition return variances, and in this capacity it may through the risk-return tradeoff be a candidate regressor in the conditional mean equation, too, as an alternative to the GARCH-variance σ_t^2 that enters through $\mu(\cdot)$. The term spread is computed as the difference between the yields on the 10-year Treasury bond and the 12-month T-bill from CRSP. The credit or default spread has been used for conditional mean returns by Fama and French (1988, 1989) and others, and for conditional variances by Schwert (1989) and others. It is calculated as the difference between Moody's seasoned Baa-rated and Aaa-rated corporate bond yields from the homepage of the Federal Reserve Bank of St. Louis. Finally, trend or momentum factors have been used by Keim and Stambaugh (1986), Carhart (1997) and others. We define momentum as $\log p_{t-1} - \log \overline{p}$, where $\overline{p}$ is the average index level across the 12-month period ending on date $t-1$.

[Table 4 about here]

For benchmarking purposes, we first estimate the parametric models, then turn to the new semi-parametric estimator. Table 4 shows estimation results for the GARCH(1,1) model without variance-in-mean effects but with x_t entering linearly in the conditional mean specification. Model 1 in the first column includes all explanatory variables in x_t. The estimates of the parameters $\theta = (\omega, \gamma, \beta)$ in the variance equation are strongly significant, with β large (at .94) and the sum of γ and β close to unity, which is standard. Dividend yield, default spread, and momentum all enter the conditional mean return with significantly positive coefficients, whereas the term spread is insignificant. Models 2 through 6 in the remaining columns are for the cases where each regressor enters alone, or none is used. The GARCH parameters in the variance equation are relatively robust to these reductions. None of the regressors is significant if entered alone, consistent with the notion that the model is misspecified if any of the regressors (except possibly the term spread) is left out.

[Table 5 about here]

Results from estimation of the parametric GARCH(1,1) model obtained by adding the conditional variance to the mean equation in a linear fashion appear in Table 5, which is laid out as Table 4. The variance equation parameters are largely unaltered by changing the mean specification. The risk-return tradeoff parameter λ, the new parameter compared to Table 4, is significantly positive throughout the table, except in Model 3 (only term spread included from x_t) which is not the preferred model in the table. In Model 1 including all regressors (first column) dividend yield and momentum enter significantly as in Table 4, but the default spread drops out when adding σ_t^2 to the mean specification. Again, Model 1 is preferred within the table, and the model with M-effect (Table 5) is preferred over the pure GARCH (Table 4). The point estimate of λ in Model 1, at .13, would suggest a moderate degree of relative risk aversion in the representative investor.

[Table 6 about here]

Table 6 shows the similar results when σ_t replaces σ_t^2 in the mean equation, and the findings are similar: The risk-return tradeoff is positive and significant, the variance parameters θ are robust to changes in the mean specification, and the preferred model retains in addition dividend yield and momentum in the mean equation, whereas any regressor from x_t is insignificant if entered individually. Various information criteria (not reported) did not yield much guidance regarding whether it is the conditional variance or its square root (i.e., conditional volatility) that enters the conditional mean return.

[Table 7 about here]

Our new semiparametric estimator is applied next, and the results are shown in Table 7. The model with all explanatory variables included (first column) is clearly preferred over that without x_t in the mean equation, thus verifying the empirical relevance of our generalization of the semiparametric model to include exogenous linear regressors. In the preferred specification (Model 1), momentum enters positively as in the parametric models, whereas the default spread now regains significance (it was lost in the parametric case when entering the M-effect, moving from Table 4 to Table 5), and dividend yield now gets a negative coefficient. The results make sense, i.e., higher credit or default spread indicates increased risk and so required return is up, and similarly momentum is expected to enter positively, whereas expected returns should be down if more is paid out in the form of dividends, consistent with dividends being valued positively by investors. The empirical findings are that the semiparametric approach makes a difference for inferences, and the received estimates appear plausible. As before, the variance equation parameters are robust to changes in mean specification, but using the semiparametric estimator we do recover significance of the trend or momentum variable also when entered alone (Model 6 in Table 7).

[Figures 3 + 4 about here]

Figure 3 shows the estimated in-mean risk-return tradeoff function $\mu(\cdot)$ from Model 1, Table 7, along with 25% and 75% quantile curves (the argument σ^2 is on the first axis). From the figure, the estimated effect is first flat, then convex. Also shown are the parametric fits (Model 1 from Tables 5 and 6) with either variance or volatility in-mean, producing linear or even concave curves. This is strong evidence against the parametric specifications. The results using the new semiparametric estimator show that it is large variance events that matter for conditional mean returns. Also shown is the estimated curve using the approach proposed by Conrad and Mammen (2008), and this hovers slightly above ours, in particular for medium to moderately high variance levels, although we do not know whether the proposed estimator is consistent or if so how to draw inference.

[Table 8 about here]

For comparison Table 8 shows the parameter estimates using their method adapted to our model specification, and here reported standard errors are based on the wild bootstrap as in Tables 4 through 7, again for comparison purposes. The results in Tables 7 and 8 are similar. Finally, dropping x_t (Model 2 in Tables 5 through 8) produces the alternative estimated curves shown in Figure 4. It is clear that while the shape is similar, and the parametric specifications again are clearly rejected, the level does change. Furthermore, the Conrad and Mammen (2008) estimated curve is now below ours, showing that there is not always a simple systematic ordering of the two estimators.

All in all, the empirical illustration verifies the usefulness of our new semiparametric estimator. The parametric alternatives are misspecified, the risk-return relation appears positive and convex in shape, and certain economic conditioning variables are relevant in the mean specification.

4 Conclusions

We have proposed a new semiparametric estimator for an empirical asset pricing model with general nonparametric risk-return tradeoff and a GARCH process for the underlying volatility. The estimator does not rely on any initial parametric estimator of the conditional mean function, and this feature facilitates the derivation of asymptotic theory under possible nonlinearity of unspecified form of the risk-return tradeoff. Using the profile likelihood approach, we have shown that our estimator under stated conditions is consistent, asymptotically normal, and achieves the semiparametric lower bound, including in our generalized semiparametric model where we also allow for the presence of exogenous variables in the mean equation. The sampling experiment provides evidence on finite sample properties and comparisons with the parametric approach and the iterative approach with parametric initial estimate proposed by Conrad and Mammen (2008). Note that in Theorems A1 and B1, we also provide the first order asymptotic theory for a GARCH-M-type model where we know the correctly specified mean function. To the best of our knowledge, this is a separate contribution, since the result is not available in the literature. Finally, the empirical application to daily stock market returns suggests that the linear relation between expected return and variance from the literature is misspecified, and this could be the reason for the disagreement on the sign and significance of the relation.

Appendix 1

Proof of Lemma 1 The proof of Lemma 1 follows Lemmas 7-9 in Severini and Wong (1992). Define

$$\widehat{g}_{\phi}\left(\sigma^{2}\right)=\frac{1}{Th_{T}}\sum_{j}y_{j}K\left(\frac{\left(\sigma^{2}-\sigma_{j}^{2}\right)}{h_{T}}\right);\; g_{\phi}\left(\sigma^{2}\right)=\mu_{\phi}\left(\sigma^{2}\right)f\left(\sigma^{2}\right),$$

and

$$\begin{aligned}\widehat{f}\left(\sigma^{2}\right) &= \frac{1}{Th_{T}}\sum_{j}K\left(\frac{\left(\sigma^{2}-\sigma_{j}^{2}\right)}{h_{T}}\right);\quad f_{\phi}\left(y_{t},\sigma^{2}\right)=f_{\phi_{0}}\left(y_{t}|\sigma^{2}\right)f\left(\sigma^{2}\right),\\ \widehat{g}_{\phi}^{(r)}\left(z\right) &= \frac{\partial^{r}}{\partial z^{r}}\widehat{g}_{\phi}\left(z\right);\; g_{\phi}^{(r)}\left(z\right)=\frac{\partial^{r}}{\partial z^{r}}g_{\phi}\left(z\right).\end{aligned}$$

Then, following Lemma 8 of Severini and Wong (1992), we have that

$$\begin{aligned}\left|E\left\{\widehat{g}_{\phi}^{(r)}\left(z\right)\right\}-g_{\phi}^{(r)}\left(z\right)\right| &= O\left(h_{T}^{2}\right),\\ \widehat{\mu}_{\phi}\left(\sigma^{2}\right) &= \frac{\sum_{j}y_{j}K\left(\left(\sigma^{2}-\sigma_{j}^{2}\right)/h_{T}\right)}{\sum_{j}K\left(\left(\sigma^{2}-\sigma_{j}^{2}\right)/h_{T}\right)}\end{aligned}$$

for any $\gamma>0$ and

$$\begin{aligned}&\sup_{\phi}\left\|\frac{\partial^{k}}{\partial\left(\sigma^{2}\right)^{k}}\frac{\partial^{j}}{\partial\phi^{j}}\widehat{\mu}_{\phi}\left(\sigma^{2}\right)-\frac{\partial^{k}}{\partial\left(\sigma^{2}\right)^{k}}\frac{\partial^{j}}{\partial\phi^{j}}\mu_{\phi}\left(\sigma^{2}\right)\right\|\\ &= O_{p}\left(T^{-v/(2v+4)}h_{T}^{-(k+(v+4)/(v+2))}T^{\gamma}+h_{T}^{2}\right)\end{aligned}$$

as $T\longrightarrow\infty$ and for $j=0,1,2$ and $k=0,1$.■

Proof of Theorem A1 The proof follows from Lemmas 2-4 that are given in the proof of Theorem B1 below.■

Proof of Theorem B1 The proof of part (a) follows from Lemmas 2-4 below. The proof technique for the QMLE in the parametric part (as in Jensen and Rahbek (2004a, 2004b)) utilizes the classic Cramér type conditions for consistency and asymptotic normality (central limit theorem for the score, convergence of the Hessian and uniformly bounded third-order derivatives); see e.g. Lehmann (1999).

4.1 First order derivatives

The first order derivatives are given by

$$\frac{\partial L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)}{\partial\omega} = -\frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}-\frac{1}{2}\sum_{t=1}^{T}\frac{-2\sigma_t^2\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}-\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^4}$$

$$= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}+\sum_{t=1}^{T}\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}}{\sigma_t^2}=\sum_{t=1}^{T}s_{1t},$$

$$\frac{\partial L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)}{\partial\gamma} = -\frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}-\frac{1}{2}\sum_{t=1}^{T}\frac{-2\sigma_t^2\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}-\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^4}$$

$$= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}+\sum_{t=1}^{T}\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^2}=\sum_{t=1}^{T}s_{2t},$$

$$\frac{\partial L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)}{\partial\beta} = -\frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}-\frac{1}{2}\sum_{t=1}^{T}\frac{-2\sigma_t^2\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}-\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^4}$$

$$= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}+\sum_{t=1}^{T}\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^2}=\sum_{t=1}^{T}s_{3t},$$

where

$$\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}=_{t=1}^{T}\beta^{j-1}\frac{1}{\sigma_t^2};\quad \frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}=_{t=1}^{T}\beta^{j-1}\frac{y_{t-j}^2}{\sigma_t^2};\quad \frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}=_{t=1}^{T}\beta^{j-1}\frac{\sigma_{t-j}^2}{\sigma_t^2}.$$

Also, note that it is possible to prove, following Lumsdaine (1996, Lemma 6, page 587) the asymptotic negligibility of the initial value σ_0^2. Then we have the following

Lemma 2 *Under Assumption A, and with the expressions of the first order derivatives evaluated at the true parameters,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{1t}\xrightarrow{d}N\left(0,\frac{\zeta}{4\omega_0^2}+C_1\right),$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{2t}\xrightarrow{d}N\left(0,\frac{\zeta}{4\gamma_0^2}+C_2\right),$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{3t}\xrightarrow{d}N\left(0,\frac{\zeta\left(1+\mu_1\right)\mu_2}{4\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)}+C_3\right),$$

with $\mu_i = E\left(\beta_0/\left(\gamma_0\varepsilon_t^2+\beta_0\right)\right)^i$, $i=1,2$ *as* $T \longrightarrow \infty$.

Proof of Lemma 2 We have by the law of iterated expectations that $E\left(s_{1t}|I_{t-1}\right) = E\left(s_{2t}|I_{t-1}\right) = E\left(s_{3t}|I_{t-1}\right) = 0$. Also, we can show straightforwardly that

$$E\left|s_{1t}\right| < \infty,\ E\left|s_{2t}\right| < \infty,\ E\left|s_{3t}\right| < \infty. \tag{11}$$

Moreover, using the ergodic theorem,

$$\frac{1}{T}\sum_{t=1}^{T} E\left(s_{1t}^2|I_{t-1}\right) = \frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\right)^2}{\sigma_t^2} \xrightarrow{p} \frac{\zeta}{4\omega_0^2} + C_1 > 0,$$

$$\frac{1}{T}\sum_{t=1}^{T} E\left(s_{2t}^2|I_{t-1}\right) = \frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)^2}{\sigma_t^2} \xrightarrow{p} \frac{\zeta}{4\gamma_0^2} + C_2 > 0,$$

$$\frac{1}{T}\sum_{t=1}^{T} E\left(s_{3t}^2|I_{t-1}\right) = \frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\frac{\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\right)^2}{\sigma_t^2} \xrightarrow{p} \frac{\zeta\left(1+\mu_1\right)\mu_2}{4\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)} + C_3 > 0.$$

We use the central limit theorem in Brown (1971) and the ergodic theorem (as in Lumsdaine (1996, page 594, proof of Theorem 3). The Lindeberg type condition also follows as in Jensen and Rahbek (2004a, 2004b). This completes the proof of Lemma 2.■

4.2 Second order derivatives

The second order derivatives are given by

$$\frac{\partial^2}{\partial\omega^2} L_T(\phi, \mu_\phi(\cdot)) = -\frac{1}{2}\sum_{t=1}^{T} \frac{\sigma_t^2 \frac{\partial^2\sigma_t^2}{\partial\omega^2} - \left(\frac{\partial\sigma_t^2}{\partial\omega}\right)^2}{\sigma_t^4} + \frac{1}{2}\sum_{t=1}^{T} \frac{-\frac{\partial\sigma_t^2}{\partial\omega}\left(y_t - \mu_\phi(\sigma_t^2)\right)^2 2\sigma_t^2 \frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^8}$$

$$+\frac{1}{2}\sum_{t=1}^{T} \frac{\sigma_t^4\left[\left(\frac{\partial^2\sigma_t^2}{\partial\omega^2}\left(y_t - \mu_\phi(\sigma_t^2)\right)^2 - \frac{\partial\sigma_t^2}{\partial\omega} 2\left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}\right)\right]}{\sigma_t^8}$$

$$+\sum_{t=1}^{T} \frac{\sigma_t^2\left[-\left(\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}\right)^2 + \left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\partial^2\mu_\phi(\sigma_t^2)}{\partial\omega^2}\right] - \left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^4}$$

$$= -\frac{1}{2}\sum_{t=1}^{T} \frac{\frac{\partial^2\sigma_t^2}{\partial\omega^2} - 2\left[-\left(\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}\right)^2 + \left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\partial^2\mu_\phi(\sigma_t^2)}{\partial\omega^2}\right]}{\sigma_t^2}$$

$$+\frac{1}{2}\sum_{t=1}^{T} \frac{\left(\frac{\partial\sigma_t^2}{\partial\omega}\right)^2 + \left[\left(\frac{\partial^2\sigma_t^2}{\partial\omega^2}\left(y_t - \mu_\phi(\sigma_t^2)\right)^2 - \frac{\partial\sigma_t^2}{\partial\omega} 2\left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}\right)\right]}{\sigma_t^4}$$

$$-\sum_{t=1}^{T} \frac{\frac{\partial\sigma_t^2}{\partial\omega}\left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}}{\sigma_t^4} - \sum_{t=1}^{T} \frac{\frac{\partial\sigma_t^2}{\partial\omega}\left(y_t - \mu_\phi(\sigma_t^2)\right)^2 \frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^6}$$

$$= -\frac{1}{2}\sum_{t=1}^{T}\left(1 - \frac{\left(y_t - \mu_\phi(\sigma_t^2)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\omega^2}}{\sigma_t^2} + \sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi(\sigma_t^2)}{\partial\omega^2}\left(y_t - \mu_\phi(\sigma_t^2)\right) - \left(\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}\right)^2\right)\frac{1}{\sigma_t^2}$$

$$-\frac{1}{2}\sum_{t=1}^{T}\left(2\frac{\left(y_t - \mu_\phi(\sigma_t^2)\right)^2}{\sigma_t^2} - 1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}\right)^2 - 2\sum_{t=1}^{T}\left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\frac{\partial\sigma_t^2}{\partial\omega}\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}}{\sigma_t^4},$$

$$\frac{\partial^2}{\partial\omega\partial\gamma} L_T(\phi, \mu_\phi(\cdot)) = -\frac{1}{2}\sum_{t=1}^{T}\left(1 - \frac{\left(y_t - \mu_\phi(\sigma_t^2)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\gamma\partial\omega}}{\sigma_t^2} - \frac{1}{2}\sum_{t=1}^{T}\left(2\frac{\left(y_t - \mu_\phi(\sigma_t^2)\right)^2}{\sigma_t^2} - 1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}\right)$$

$$+\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi(\sigma_t^2)}{\partial\gamma\partial\omega}\left(y_t - \mu_\phi(\sigma_t^2)\right) - \left(\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\gamma}\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}\right)\right)\frac{1}{\sigma_t^2} - 2\sum_{t=1}^{T}\left(y_t - \mu_\phi(\sigma_t^2)\right)\frac{\frac{\partial\sigma_t^2}{\partial\gamma}\frac{\partial\mu_\phi(\sigma_t^2)}{\partial\omega}}{\sigma_t^4},$$

$$\frac{\partial^2}{\partial\omega\partial\beta}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)=-\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\beta\partial\omega}}{\sigma_t^2}-\frac{1}{2}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}\right)$$

$$+\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\beta\partial\omega}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\right)\right)\frac{1}{\sigma_t^2}-2\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}}{\sigma_t^4},$$

$$\frac{\partial^2}{\partial\gamma\partial\beta}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)=-\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\beta\partial\gamma}}{\sigma_t^2}-\frac{1}{2}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}\right)$$

$$+\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\beta\partial\gamma}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)\right)\frac{1}{\sigma_t^2}-2\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^4},$$

$$\frac{\partial^2}{\partial\gamma^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)=-\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\gamma^2}}{\sigma_t^2}-\frac{1}{2}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}\right)^2$$

$$+\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma^2}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)^2\right)\frac{1}{\sigma_t^2}-2\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\gamma}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^4},$$

$$\frac{\partial^2}{\partial\beta^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)=-\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\beta^2}}{\sigma_t^2}-\frac{1}{2}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\right)^2$$

$$+\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\beta^2}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\right)^2\right)\frac{1}{\sigma_t^2}-2\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^4}.$$

Lemma 3 *Under Assumption A, and with the expressions of the second order derivatives evaluated at the true parameters*

(a) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\omega^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)|_{\phi=\phi_0}\right)\xrightarrow{p}\frac{1}{2\omega_0^2}+C_1>0,$

(b) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\gamma^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)|_{\phi=\phi_0}\right)\xrightarrow{p}\frac{1}{2\gamma_0^2}+C_2>0,$

(c) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)|_{\phi=\phi_0}\right)\xrightarrow{p}\frac{(1+\mu_1)\mu_2}{2\beta_0^2(1-\mu_1)(1-\mu_2)}+C_3>0,$

(d) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\gamma\partial\beta}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)|_{\phi=\phi_0}\right)\xrightarrow{p}\frac{\mu_1}{2\gamma_0\beta_0(1-\mu_1)}+C_4,$

(e) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\omega\partial\gamma}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)|_{\phi=\phi_0}\right)\xrightarrow{p}\frac{1}{2\omega_0\gamma_0}+C_5,$

(f) $\frac{1}{T}\left(-\frac{\partial^2}{\partial w\partial\beta}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)|_{\phi=\phi_0}\right)\xrightarrow{p}\frac{\mu_1}{2\omega_0\beta_0(1-\mu_1)}+C_6,$

as $T\longrightarrow\infty$.

Proof of Lemma 3 By the ergodic theorem (for simplicity we drop the subscript of the parameters evaluated at the true values),

$$\frac{1}{T}\left(-\frac{\partial^2}{\partial\omega^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)\right)=\frac{1}{2T}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\omega^2}}{\sigma_t^2}$$
$$-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\omega^2}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\right)^2\right)\frac{1}{\sigma_t^2}+\frac{1}{2T}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}\right)^2$$
$$+\frac{2}{T}\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\omega}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}}{\sigma_t^4}\xrightarrow{p}\frac{1}{2\omega_0^2}+C_1,$$

$$\frac{1}{T}\left(-\frac{\partial^2}{\partial\gamma^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)\right)=\frac{1}{2T}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\gamma^2}}{\sigma_t^2}$$
$$-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma^2}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)^2\right)\frac{1}{\sigma_t^2}+\frac{1}{2T}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\partial\sigma_t^2/\partial\gamma}{\sigma_t^2}\right)^2$$
$$+\frac{2}{T}\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\gamma}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^4}\xrightarrow{p}\frac{1}{2\gamma_0^2}+C_2,$$

$$\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta^2}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)\right)=\frac{1}{2T}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\beta^2}}{\sigma_t^2}$$
$$-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\beta^2}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\right)^2\right)\frac{1}{\sigma_t^2}$$
$$+\frac{1}{2T}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial^2\sigma_t^2}{\partial\beta^2}}{\sigma_t^2}\right)^2$$
$$+\frac{2}{T}\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}}{\sigma_t^4}\xrightarrow{p}\frac{\left(1+\mu_1\right)\mu_2}{2\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)}+C_3.$$

$$
\begin{aligned}
\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\gamma}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)\right) &= \frac{1}{2T}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\beta\partial\gamma}}{\sigma_t^2} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\beta\partial\gamma}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)\right)\frac{1}{\sigma_t^2} \\
&\quad +\frac{1}{2T}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\frac{\frac{\partial\sigma_t^2}{\partial\gamma}}{\sigma_t^2}\right) \\
&\quad +\frac{2}{T}\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^4}\xrightarrow{p}\frac{\mu_1}{2\gamma_0\beta_0\left(1-\mu_1\right)}+C_4,
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{T}\left(-\frac{\partial^2}{\partial\omega\partial\gamma}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)\right) &= \frac{1}{2T}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\omega\partial\gamma}}{\sigma_t^2} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\beta\partial\gamma}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}\right)\right)\frac{1}{\sigma_t^2} \\
&\quad +\frac{1}{2T}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\partial\sigma_t^2/\partial\omega}{\sigma_t^2}\frac{\partial\sigma_t^2/\partial\gamma}{\sigma_t^2}\right) \\
&\quad +\frac{2}{T}\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\omega}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\gamma}}{\sigma_t^4}\xrightarrow{p}\frac{1}{2\omega_0\gamma_0}+C_5,
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\omega}L_T\left(\phi,\mu_\phi\left(\cdot\right)\right)\right) &= \frac{1}{2T}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^2\sigma_t^2}{\partial\beta\partial\gamma}}{\sigma_t^2} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial\beta\partial\omega}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}\right)\right)\frac{1}{\sigma_t^2} \\
&\quad +\frac{1}{2T}\sum_{t=1}^{T}\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\frac{\frac{\partial\sigma_t^2}{\partial\omega}}{\sigma_t^2}\right) \\
&\quad +\frac{2}{T}\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\frac{\partial\sigma_t^2}{\partial\beta}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial\omega}}{\sigma_t^4}\xrightarrow{p}\frac{\mu_1}{2\omega_0\beta_0\left(1-\mu_1\right)}+C_6,
\end{aligned}
$$

and this concludes the proof of Lemma 3.■

4.3 Third order derivatives

With $\sigma_t^2 = \sigma_t^2(\phi)$ and for any $i, j, k = \omega, \beta, \gamma$, the third order derivatives are given by

$$\frac{\partial^3}{\partial i \partial j \partial k} L_T(\phi, \mu_\phi(\cdot)) = -\frac{1}{2}\sum_{t=1}^{T} \frac{\sigma_t^2 \frac{\partial^3 \sigma_t^2}{\partial i \partial j \partial k} - \frac{\partial^2 \sigma_t^2}{\partial i \partial j}\frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^4}$$

$$+\frac{1}{2}\sum_{t=1}^{T} \frac{\sigma_t^4\left[\left(-2\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial \mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial^2\sigma_t^2}{\partial i\partial j}+\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2\frac{\partial^3\sigma_t^2}{\partial i\partial j\partial k}\right)\right]-2\sigma_t^2\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2\frac{\partial^2\sigma_t^2}{\partial i\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^8}$$

$$+\sum_{t=1}^{T}\frac{\sigma_t^2\left(\left(\frac{\partial^3\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j\partial k}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)-\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\right)\right)-\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4}$$

$$-\sum_{t=1}^{T}\frac{\sigma_t^2\left[2\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\right)\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\right]-\left(\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\right)^2\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4}+\frac{1}{2}\sum_{t=1}^{T}\frac{\sigma_t^4\left[2\left(\frac{\partial\sigma_t^2}{\partial k}\right)\left(\frac{\partial^2\sigma_t^2}{\partial i\partial j}\right)\right]-\frac{\partial\sigma_t^2}{\partial i}\frac{\partial\sigma_t^2}{\partial j}\frac{\partial\sigma_t^2}{\partial k}2\sigma_t^2}{\sigma_t^8}$$

$$-\sum_{t=1}^{T}\frac{\sigma_t^6\left[-2\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial\sigma_t^2}{\partial i}\frac{\partial\sigma_t^2}{\partial j}+2\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2\left(\frac{\partial\sigma_t^2}{\partial k}\right)\left(\frac{\partial^2\sigma_t^2}{\partial i\partial j}\right)\right]-3\sigma_t^4\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2\frac{\partial\sigma_t^2}{\partial i}\frac{\partial\sigma_t^2}{\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^{12}}$$

$$-2\sum_{t=1}^{T}\frac{\sigma_t^4\left[-\frac{\partial\sigma_t^2}{\partial k}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial i}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial j}+\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\left[\frac{\partial^2\sigma_t^2}{\partial i\partial j}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}+\frac{\partial\sigma_t^2}{\partial k}\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\right]\right]}{\sigma_t^8}$$

$$=-\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\left(\frac{\frac{\partial^3\sigma_t^2}{\partial i\partial j\partial k}}{\sigma_t^2}\right)-\frac{1}{2}\sum_{t=1}^{T}3\left(2\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}-1\right)\frac{\frac{\partial^2\sigma_t^2}{\partial i\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4}$$

$$-\frac{1}{2}\sum_{t=1}^{T}2\left(1-3\frac{\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial i}\frac{\partial\sigma_t^2}{\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^6}-3\sum_{t=1}^{T}\left(\frac{\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}}{\sigma_t^2}-\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial i}\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4}\right)$$

$$-\sum_{t=1}^{T}\left(y_t-\mu_\phi\left(\sigma_t^2\right)\right)\left(\frac{3\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial^2\sigma_t^2}{\partial i\partial j}}{\sigma_t^4}-2\frac{\frac{\partial\mu_\phi\left(\sigma_t^2\right)}{\partial k}\frac{\partial\sigma_t^2}{\partial i}\frac{\partial\sigma_t^2}{\partial j}}{\sigma_t^6}+\frac{3\frac{\partial^2\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j}\frac{\partial\sigma_t^2}{\partial k}}{\sigma_t^4}-\frac{\frac{\partial^3\mu_\phi\left(\sigma_t^2\right)}{\partial i\partial j\partial k}}{\sigma_t^2}\right).$$

We have now

Definition 1 *Following Jensen and Rahbek (2004b), we introduce bounds on each parameter in* ϕ_0,

$$\omega_L < \omega_0 < \omega_U, \quad \beta_L < \beta_0 < \beta_U, \quad \gamma_L < \gamma_0 < \gamma_U,$$

and we define the neighborhood $N(\phi_0)$ around ϕ_0 as

$$N(\phi_0) = \{\phi : \omega_L \leq \omega \leq \omega_U, \beta_L \leq \beta \leq \beta_U, \gamma_L < \gamma < \gamma_U\}. \tag{12}$$

Then

Lemma 4 *Under Assumption A, there exists a neighborhood $N(\phi_0)$ of the type given by (12) in Definition 1 for which*

(a) $\sup_{\phi \in N(\phi_0)} \left| \frac{\partial^3}{\partial \omega^3} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{1t}$,

(b) $\sup_{\phi \in N(\phi_0)} \left| \frac{\partial^3}{\partial \beta^3} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{2t}$,

(c) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \gamma^3} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{3t}$,

(d) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \omega^2 \partial \beta} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{4t}$,

(e) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \omega^2 \partial \gamma} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{5t}$,

(f) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \beta^2 \partial \gamma} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{6t}$,

(g) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \omega \partial \beta^2} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{7t}$,

(h) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \omega \partial \gamma^2} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{8t}$,

(i) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \beta \partial \gamma^2} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{9t}$,

(j) $\sup_{\phi \in N(\phi_0)} \left| \frac{1}{T} \frac{\partial^3}{\partial \omega \partial \beta \partial \gamma} L_T(\phi, \mu_\phi(\cdot)) \right| \leq \frac{1}{T} \sum_{t=1}^T w_{10t}$,

where $w_{1t}, ..., w_{9t}$ and w_{10t} are stationary and have finite moments, $E(w_{it}) = M_i < \infty, \forall i = 1, ..., 10$. Furthermore $\frac{1}{T} \sum_{t=1}^T w_{it} \xrightarrow{a.s.} M_i, \forall i = 1, ..., 10$.

Proof of Lemma 4 Following Jensen and Rahbek (2004b, Lemma 10), we start with part (a), where $i, j, k = \omega$

$$\begin{aligned} w_{1t}(\phi) &= \left(1 + \varepsilon_t^2 \frac{\sigma_t^2(\phi_0)}{\sigma_t^2}\right) \left(\frac{\frac{\partial^3 \sigma_t^2}{\partial i \partial j \partial k}}{\sigma_t^2}\right) + 3\left(2\varepsilon_t^2 \frac{\sigma_t^2(\phi_0)}{\sigma_t^2} + 1\right) \frac{\frac{\partial^2 \sigma_t^2}{\partial i \partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^4} + 2\left(1 + 3\varepsilon_t^2 \frac{\sigma_t^2(\phi_0)}{\sigma_t^2}\right) \frac{\frac{\partial \sigma_t^2}{\partial i} \frac{\partial \sigma_t^2}{\partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^6} \\ &\quad + 3\left(\frac{\frac{\partial^2 \mu_\phi(\sigma_t^2)}{\partial i \partial j} \frac{\partial \mu_\phi(\sigma_t^2)}{\partial k}}{\sigma_t^2} + \frac{\frac{\partial \mu_\phi(\sigma_t^2)}{\partial i} \frac{\partial \mu_\phi(\sigma_t^2)}{\partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^4}\right) \\ &\quad + \varepsilon_t \sigma_t(\phi_0) \left(\frac{3 \frac{\partial \mu_\phi(\sigma_t^2)}{\partial k} \frac{\partial^2 \sigma_t^2}{\partial i \partial j}}{\sigma_t^4} + 2 \frac{\frac{\partial \mu_\phi(\sigma_t^2)}{\partial k} \frac{\partial \sigma_t^2}{\partial i} \frac{\partial \sigma_t^2}{\partial j}}{\sigma_t^6} + \frac{3 \frac{\partial^2 \mu_\phi(\sigma_t^2)}{\partial i \partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^4} + \frac{\frac{\partial^3 \mu_\phi(\sigma_t^2)}{\partial i \partial j \partial k}}{\sigma_t^2}\right). \end{aligned}$$

Then the quantities $\frac{\frac{\partial^3 \sigma_t^2}{\partial i \partial j \partial k}}{\sigma_t^2}$, $\frac{\frac{\partial^2 \sigma_t^2}{\partial i \partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^4}$ and $\frac{\frac{\partial \sigma_t^2}{\partial i} \frac{\partial \sigma_t^2}{\partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^6}$ are bounded by functions that have the desired moments. By assumption A3,

$\frac{\frac{\partial^2 \mu_\phi\left(\sigma_t^2\right)}{\partial i \partial j} \frac{\partial \mu_\phi\left(\sigma_t^2\right)}{\partial k}}{\sigma_t^2}$, $\frac{\frac{\partial \mu_\phi\left(\sigma_t^2\right)}{\partial i} \frac{\partial \mu_\phi\left(\sigma_t^2\right)}{\partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^4}$, $\frac{3\frac{\partial \mu_\phi\left(\sigma_t^2\right)}{\partial k} \frac{\partial^2 \sigma_t^2}{\partial i \partial j}}{\sigma_t^4}$, $\frac{\frac{\partial \mu_\phi\left(\sigma_t^2\right)}{\partial k} \frac{\partial \sigma_t^2}{\partial i} \frac{\partial \sigma_t^2}{\partial j}}{\sigma_t^6}$, $\frac{3\frac{\partial^2 \mu_\phi\left(\sigma_t^2\right)}{\partial i \partial j} \frac{\partial \sigma_t^2}{\partial k}}{\sigma_t^4}$ and $\frac{\frac{\partial^3 \mu_\phi\left(\sigma_t^2\right)}{\partial i \partial j \partial k}}{\sigma_t^2}$ also are bounded by functions that have the desired moments. Parts (b)-(j) of Lemma 4 follow by the same arguments.■

Part (b) of Theorem B1 follows directly from Theorem A1.■

Proofs of Theorems A2 and B2 It follows from the asymptotic equivalence result provided in Severini and Wong (1992) and stated in the following Corollary

Corollary 1 *Under Assumptions A, B and C, maximizing the quasi log-likelihood function* $L_T\left(\phi, \widehat{\mu}_\phi(\cdot)\right)$ *evaluated at* $\widehat{\mu}_\phi(\cdot)$, *is asymptotically equivalent to maximizing* $L_T\left(\phi, \mu_\phi(\cdot)\right)$, *since it reaches the same maximum as* $L_T\left(\phi, \mu_\phi(\cdot)\right)$.

Proof of Corollary 1 It follows from the asymptotic equivalence result of Severini and Wong (1992, page 1775).■

This completes the proof of Theorems A2 and B2.■

Proofs of Theorems A3 and B3 It follows from the proofs of Theorems A2 and B2 and with the introduction of Assumption D.■

Appendix 2: Figures and Tables

Figure 1: Parametric and semiparametric estimates for models N1, N2 and N3.

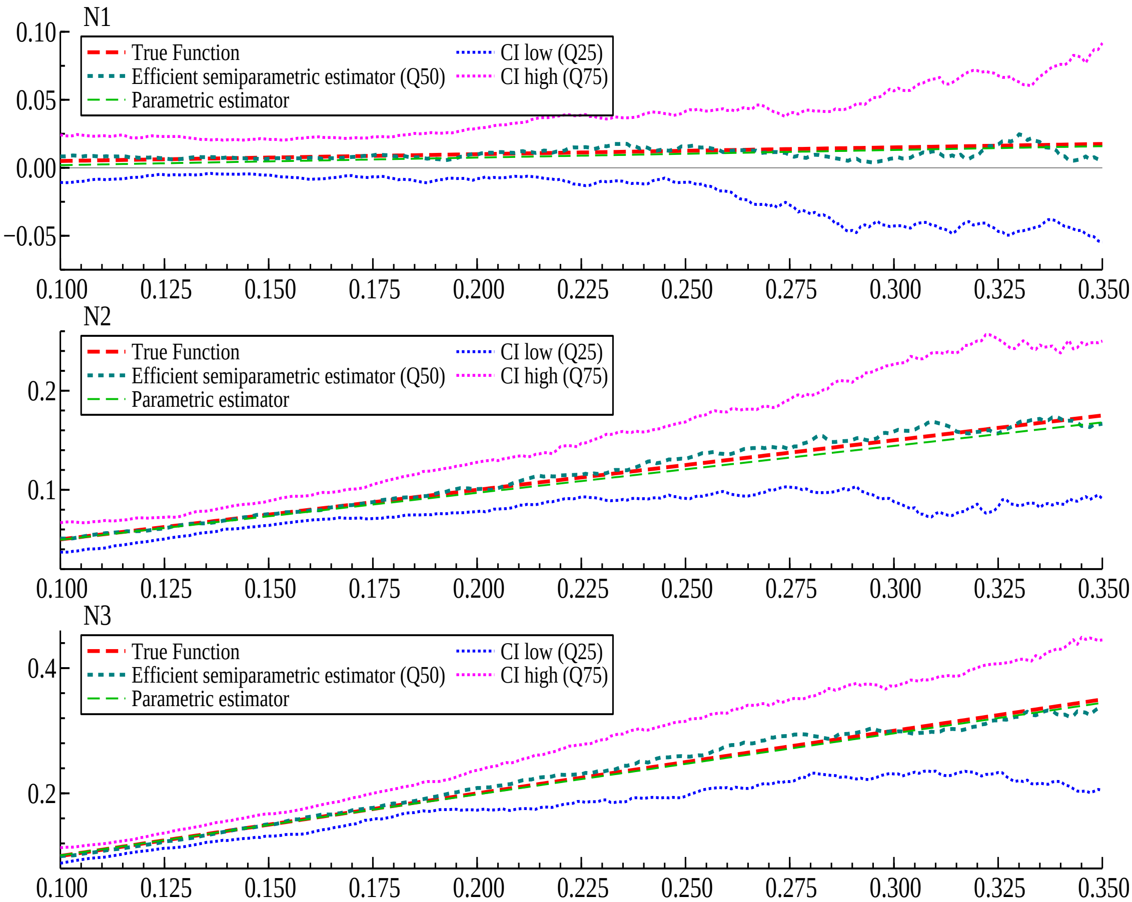

Figure 2: Parametric and semiparametric estimates for models A1, A2 and A3.

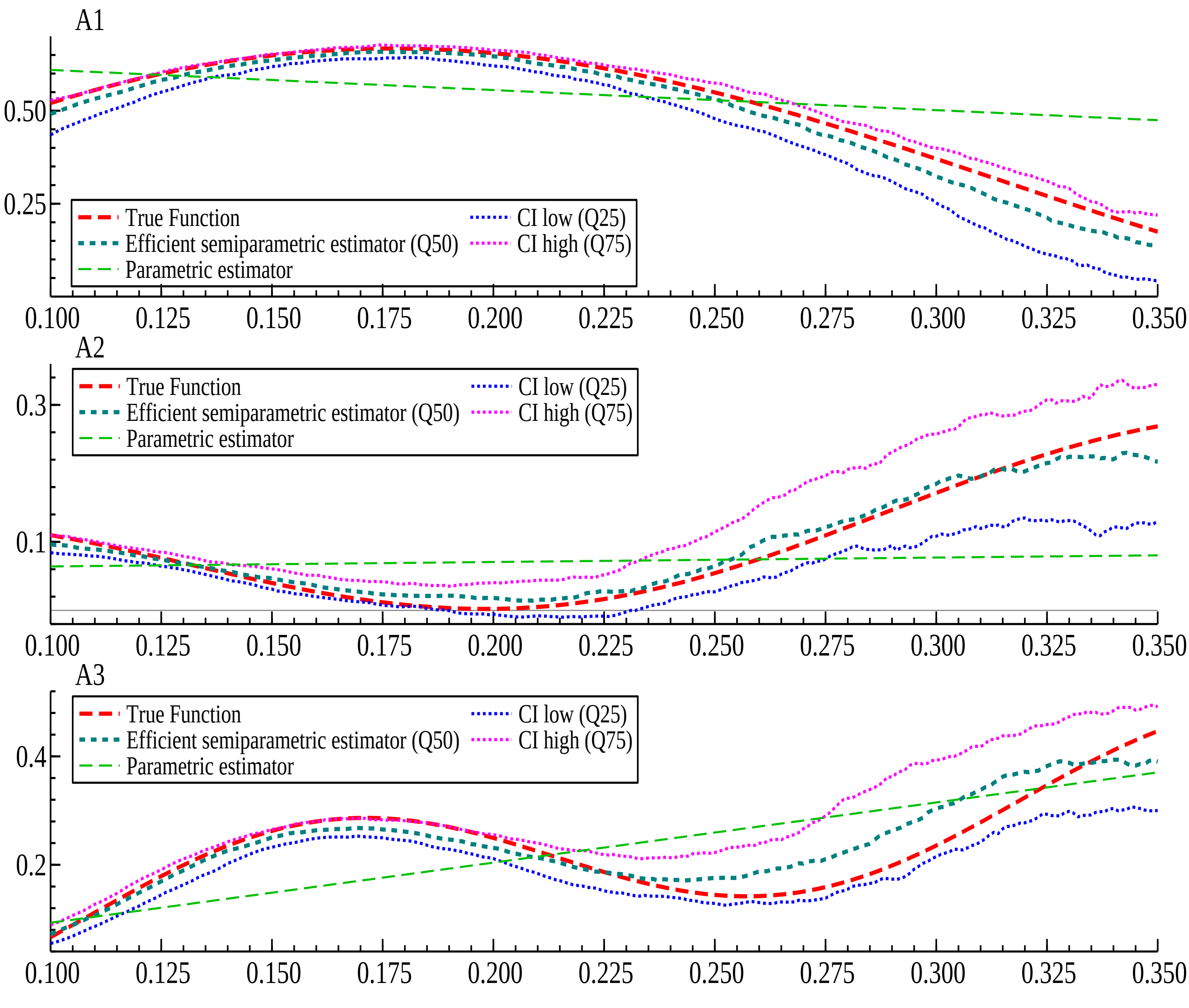

Figure 3: Parametric and semiparametric estimates of the M-effects based on the model for S&P 500 with covariates included for the period Jan. 1, 1990 - Dec 31, 2007.

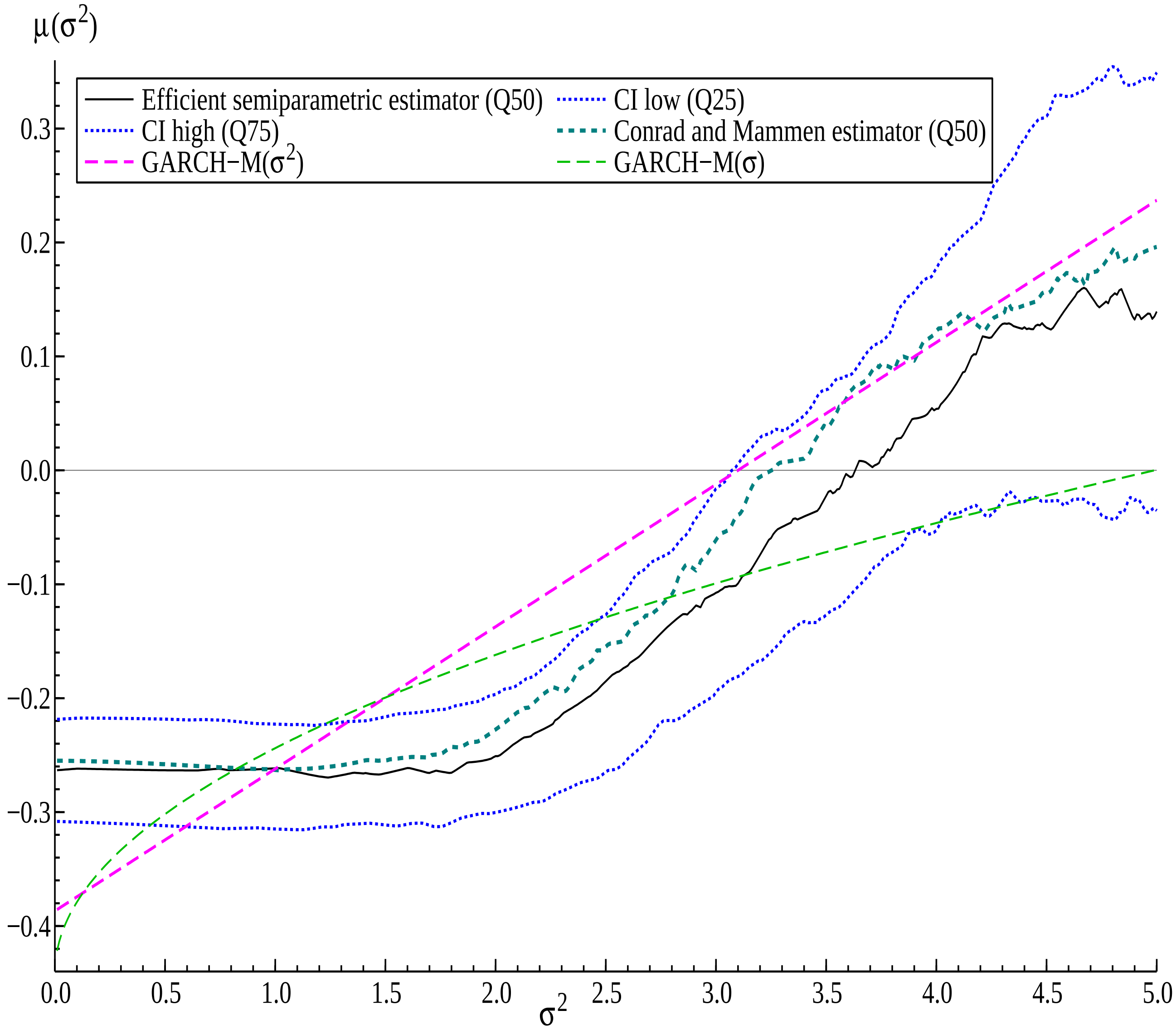

Figure 4: Parametric and semiparametric estimates of the M-effects based on the model for S&P 500 without covariates for the period Jan. 1, 1990 - Dec 31, 2007.

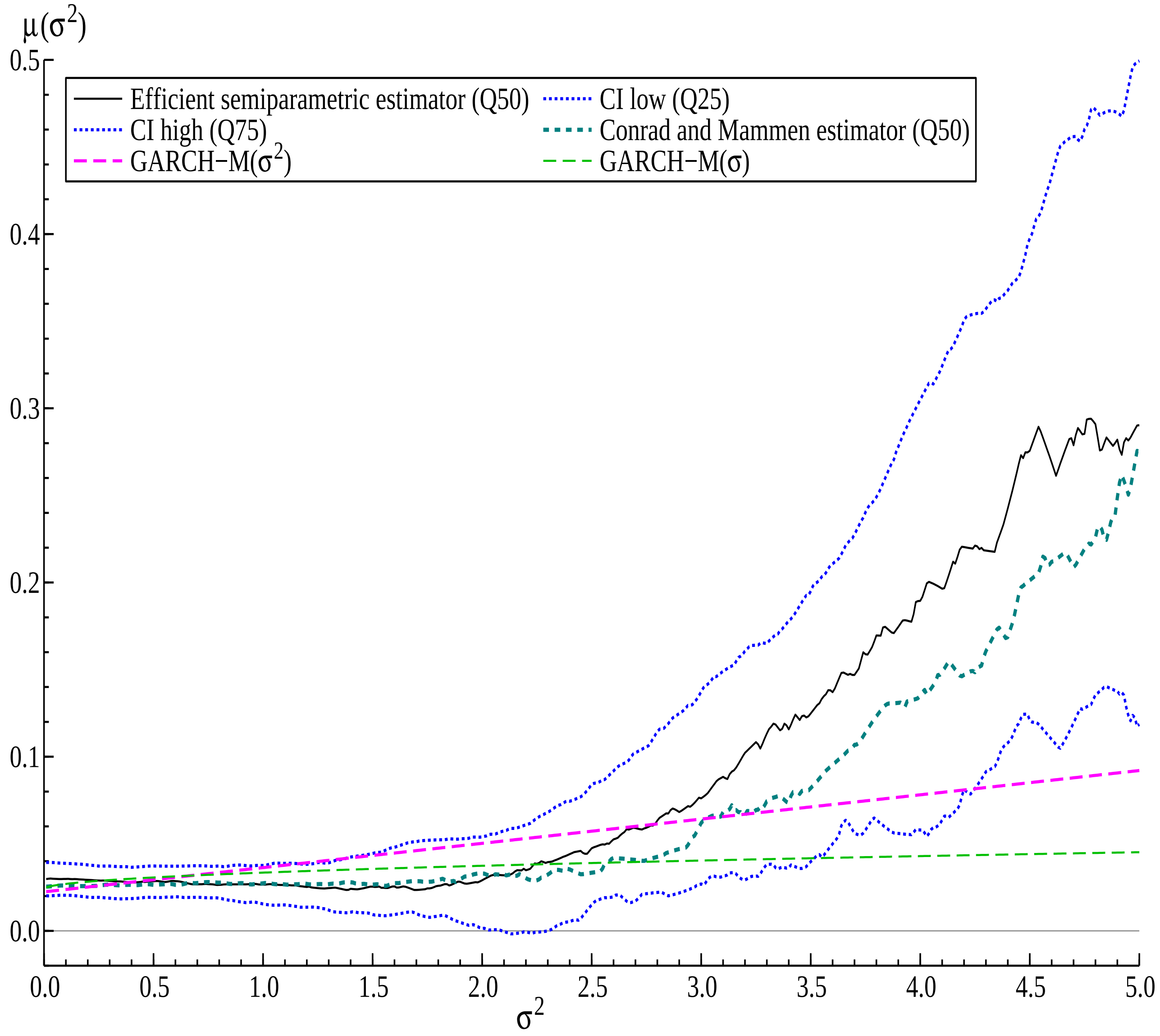

Table 1: MC results on precision of estimators: Parametric GARCH(1,1)-M approach

	Q25	ω	Q75	Q25	γ	Q75	Q25	β	Q75	Q25	μ	Q75	Q25	λ	Q75
N1	0.007	0.011	0.015	0.087	0.103	0.122	0.813	0.842	0.870	-0.023	-0.004	0.026	-0.094	0.056	0.181
N2	0.008	0.011	0.014	0.084	0.101	0.119	0.802	0.837	0.858	-0.019	0.002	0.026	0.342	0.475	0.625
N3	0.009	0.012	0.014	0.083	0.096	0.111	0.786	0.808	0.836	-0.020	0.003	0.025	0.813	0.976	1.107
A1	0.026	0.038	0.051	0.096	0.111	0.125	0.470	0.534	0.606	0.592	0.664	0.740	-0.899	-0.541	-0.225
A2	0.007	0.010	0.014	0.087	0.102	0.121	0.812	0.836	0.865	0.023	0.058	0.094	-0.168	0.065	0.279
A3	0.008	0.010	0.013	0.075	0.089	0.100	0.813	0.834	0.856	-0.049	-0.018	0.010	0.948	1.109	1.371

The number of Monte Carlo replications equals 200. Sample size is 1000

Table 2: MC results on precision of estimators: Efficient semiparametric GARCH(1,1)-M approach

	Q25	ω	Q75	Q25	γ	Q75	Q25	β	Q75
N1	0.008	0.011	0.015	0.081	0.097	0.114	0.820	0.845	0.872
N2	0.008	0.011	0.015	0.079	0.094	0.109	0.805	0.834	0.861
N3	0.009	0.012	0.014	0.077	0.087	0.104	0.788	0.811	0.845
A1	0.013	0.018	0.029	0.082	0.093	0.106	0.595	0.644	0.681
A2	0.008	0.011	0.014	0.077	0.092	0.109	0.814	0.841	0.868
A3	0.009	0.011	0.015	0.076	0.090	0.101	0.791	0.821	0.845

The number of Monte Carlo replications equals 200. Sample size is 1000

Table 3: MC results on precision of estimators: The Conrad and Mammen GARCH(1,1)-M approach.

	Q25	ω	Q75	Q25	γ	Q75	Q25	β	Q75
N1	0.008	0.011	0.015	0.080	0.097	0.114	0.819	0.845	0.872
N2	0.008	0.011	0.015	0.079	0.094	0.109	0.806	0.833	0.862
N3	0.009	0.012	0.014	0.077	0.087	0.104	0.788	0.811	0.846
A1	0.013	0.018	0.029	0.082	0.093	0.106	0.594	0.644	0.681
A2	0.008	0.011	0.014	0.077	0.092	0.109	0.814	0.839	0.868
A3	0.009	0.011	0.014	0.076	0.090	0.101	0.791	0.821	0.845

The number of Monte Carlo replications equals 200. Sample size is 1000

Table 4: Parametric GARCH(1,1)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
ω	0.006***	0.007***	0.005***	0.000	0.001	0.003*
	(0.002)	(0.002)	(0.002)	(0.000)	(0.001)	(0.002)
γ	0.055***	0.054***	0.056***	0.061***	0.078***	0.068***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.010)	(0.008)
β	0.939***	0.937***	0.940***	0.944***	0.931***	0.930***
	(0.010)	(0.009)	(0.009)	(0.007)	(0.008)	(0.008)
μ	-0.277***	0.041***	0.072***	-0.015	0.023	0.040***
	(0.067)	(0.012)	(0.018)	(0.054)	(0.039)	(0.016)
Term spread	0.007		-0.009			
	(0.012)		(0.010)			
Default spread	0.162**			0.090		
	(0.063)			(0.069)		
Dividend yield	3.229*				0.970	
	(1.894)				(1.705)	
Momentum	2.048***					0.004
	(0.241)					(0.201)

Numbers in parentheses are wild bootstrap standard errors.

Sample: Jan 2. 1990 - Dec 31. 2007

*: significant at 10 percent level, two-sided (normal dist.).

**: significant at 5 percent level, two-sided (normal dist.).

***: significant at 1 percent level, two-sided (normal dist.).

Table 5: Parametric GARCH(1,1)-M w. σ_t^2 in mean

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
ω	0.007***	0.007***	0.006***	0.004***	0.006***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
γ	0.056***	0.055***	0.056***	0.053***	0.060***	0.058***
	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
β	0.937***	0.936***	0.938***	0.943***	0.935***	0.935***
	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
μ	-0.387***	0.011	0.048**	-0.041	0.047	0.031
	(0.069)	(0.018)	(0.023)	(0.054)	(0.051)	(0.026)
λ	0.125***	0.048**	0.038	0.054**	0.086***	0.090***
	(0.028)	(0.022)	(0.024)	(0.025)	(0.029)	(0.028)
Term spread	0.009		-0.009			
	(0.012)		(0.010)			
Default spread	0.085			0.085		
	(0.066)			(0.075)		
Dividend yield	6.613***				-1.850	
	(2.065)				(1.748)	
Momentum	2.378***					-0.133
	(0.241)					(0.215)

Numbers in parentheses are wild bootstrap standard errors.

Sample: Jan 2. 1990 - Dec 31. 2007

*: significant at 10 percent level, two-sided (normal dist.).

**: significant at 5 percent level, two-sided (normal dist.).

***: significant at 1 percent level, two-sided (normal dist.).

Table 6: Parametric GARCH(1,1)-M w. σ_t in mean

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
ω	0.007***	0.007***	0.006***	0.004**	0.006***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
γ	0.055***	0.055***	0.057***	0.053***	0.060***	0.057***
	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)
β	0.937***	0.936***	0.938***	0.944***	0.935***	0.936***
	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
μ	-0.442***	-0.024	0.023	-0.078	-0.027	-0.058
	(0.078)	(0.035)	(0.040)	(0.054)	(0.074)	(0.047)
λ	0.198***	0.086**	0.065	0.097*	0.167***	0.191***
	(0.055)	(0.043)	(0.047)	(0.051)	(0.060)	(0.055)
Term spread	0.008		-0.010			
	(0.012)		(0.010)			
Default spread	0.089			0.083		
	(0.067)			(0.076)		
Dividend yield	6.030***				-1.749	
	(2.063)				(1.780)	
Momentum	2.268***					-0.114
	(0.240)					(0.213)

Numbers in parentheses are wild bootstrap standard errors.
Sample: Jan 2. 1990 - Dec 31. 2007
*: significant at 10 percent level, two-sided (normal dist.).
**: significant at 5 percent level, two-sided (normal dist.).
***: significant at 1 percent level, two-sided (normal dist.).

Table 7: Efficient semiparametric GARCH(1,1)-M

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
ω	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
γ	0.053***	0.055***	0.055***	0.054***	0.054***	0.054***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
β	0.940***	0.939***	0.940***	0.940***	0.940***	0.940***
	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Term spread	0.012		-0.011			
	(0.012)		(0.012)			
Default spread	0.288***			-0.065		
	(0.077)			(0.076)		
Dividend yield	-2.041***				-0.599	
	(0.270)				(2.018)	
Momentum	1.971***					1.897***
	(0.263)					(0.247)

Numbers in parentheses are wild bootstrap standard errors.
Sample: Jan 2. 1990 - Dec 31. 2007
*: significant at 10 percent level, two-sided (normal dist.).
**: significant at 5 percent level, two-sided (normal dist.).
***: significant at 1 percent level, two-sided (normal dist.).

Table 8: Conrad and Mammen semiparametric GARCH(1,1)-M

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
ω	0.006**	0.006**	0.007**	0.006***	0.006**	0.006**
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
γ	0.056***	0.056***	0.057***	0.056***	0.056***	0.057***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)
β	0.938***	0.938***	0.938***	0.938***	0.938***	0.938***
	(0.009)	(0.008)	(0.008)	(0.007)	(0.009)	(0.009)
Term spread	0.012		-0.011			
	(0.015)		(0.012)			
Default spread	0.288**			-0.066		
	(0.122)			(0.074)		
Dividend yield	-2.042***				-0.565	
	(0.327)				(1.853)	
Momentum	1.971***					1.893***
	(0.307)					(0.201)

Numbers in parentheses are wild bootstrap standard errors.

Sample: Jan 2. 1990 - Dec 31. 2007

*: significant at 10 percent level, two-sided (normal dist.).

**: significant at 5 percent level, two-sided (normal dist.).

***: significant at 1 percent level, two-sided (normal dist.).

References

[1] Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006), The Cross-Section of Volatility and Expected Returns, *Journal of Finance* 61, 259-299.

[2] Backus, D. K. and A. W. Gregory (1993), Theoretical Relations Between Risk Premiums and Conditional Variances, *Journal of Business and Economic Statistics* 11, 177-185.

[3] Berkes, I. and L. Horváth (2004), The Efficiency of the Estimators of the Parameters in GARCH Processes, *The Annals of Statistics* 32, 2, 633-655.

[4] Bollerslev, T., R. F. Engle and J. M. Wooldridge (1988), A Capital Asset Pricing Model with Time-Varying Covariances, *Journal of Political Economy* 96, 116-131.

[5] Brandt, M. W. and Q. Kang (2004), On the Relationship Between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach, *Journal of Financial Economics* 72, 217-257.

[6] Brown, B. M. (1971), Martingale Central Limit Theorems, *Annals of Mathematical Statistics* 42, 59-66.

[7] Campbell, J. Y. (1987), Stock Returns and the Term Structure, *Journal of Financial Economics* 18, 373-399.

[8] Campbell, J. Y. and L. Hentschel (1992), No News Is Good News: An Asymmetric Model of Changing Volatility in Stock Returns, *Journal of Financial Economics* 31, 281-318.

[9] Campbell, J. Y. and R. J. Shiller (1988a), The Dividend Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1, 195-228.

[10] Campbell, J. Y. and R. J. Shiller (1988b), Stock Prices, Earnings, and Expected Dividends, *Journal of Finance* 43, 661-676.

[11] Carhart, M. (1997), On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

[12] Chou, R. Y. (1988), Volatility Persistence and Stock Valuations: Some Empirical Evidence Using GARCH, *Journal of Applied Econometrics* 3, 279-294.

[13] Chou, R. Y., R. F. Engle and A. Kane (1992), Measuring Risk Aversion from Excess Returns on a Stock Index, *Journal of Econometrics* 52, 201-224.

[14] Christensen, B. J. and M. Ø. Nielsen (2007), The Effect of Long Memory in Volatility on Stock Market Fluctuations, *Review of Economics and Statistics* 89, 4, 684-700.

[15] Conrad, C. and E. Mammen (2008), Nonparametric Regression on Latent Covariates with an Application to Semiparametric GARCH-in-Mean Models, *Working paper, University of Mannheim.*

[16] Dahl, C. and E. M. Iglesias (2007), Asymptotic normality of the QMLE of stationary and nonstationary GARCH with serially dependent standardized disturbances, *Working Paper.*

[17] Fama, E. F. and K. R. French (1988), Dividend Yields and Expected Stock Returns, *Journal of Financial Economics* 22, 3-25.

[18] Fama, E. F. and K. R. French (1989), Business Conditions and Expected Returns on Stock and Bonds, *Journal of Financial Economics* 25, 23-49.

[19] 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.

[20] Ghysels, E., P. Santa-Clara and R. Valkanov (2005), There Is a Risk-Return Tradeoff After All, *Journal of Financial Economics* 76, 509-548.

[21] 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.

[22] Harrison, P. and H. H. Zhang (1999), An Investigation of the Risk and Return Relation at Long Horizons, *Review of Economics and Statistics* 81, 399-408.

[23] Hodgson, D. J. and K. P. Vorkink (2003), Efficient Estimation of Conditional Asset Pricing Models, *Journal of Business and Economic Statistics* 21, 2, 269-283.

[24] Jensen, S. T. and A. Rahbek (2004a), Asymptotic Normality of the QML Estimator of ARCH in the Nonstationary Case, *Econometrica* 72, 2, 641-646.

[25] Jensen, S. T. and A. Rahbek (2004b), Asymptotic Inference for Nonstationary GARCH, *Econometric Theory* 20, 6, 1203-1226.

[26] Keim, D. B. and R. F. Stambaugh (1986), Predicting Returns in the Stock and Bond Markets, *Journal of Financial Economics* 17, 357-390.

[27] Lehmann, E. L. (1999), *Elements of Large Sample Theory*, New York: Springer Verlag.

[28] Lettau, M. and S. C. Ludvigson (forthcoming), Measuring and Modeling Variation in the Risk-Return Tradeoff, in Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics* (Amsterdam: North-Holland).

[29] Ling, S. (2004), Estimation and Testing Stationarity for Double-Autoregressive Models, *Journal of the Royal Statistical Society Series B* 66, 1, 63-78.

[30] Linton, O. and B. Perron (2003), The Shape of the Risk Premium: Evidence from a Semiparametric GARCH Model, *Journal of Business and Economic Statistics* 21, 3, 354-367.

[31] Lumsdaine, R. L. (1996), Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models. *Econometrica* 64, 3, 575-596.

[32] Merton, R. (1973), An Intertemporal Capital Asset Pricing Model, *Econometrica* 41, 867-887.

[33] Merton, R. (1980), On Estimating the Expected Return on the Market, *Journal of Financial Economics* 8, 323-361.

[34] Nelson, D. B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.

[35] Poterba, J. M. and L. H. Summers (1986), The Persistence of Volatility and Stock Market Fluctuations, *American Economic Review* 76, 1142-1151.

[36] Schwert, G. W. (1989), Why Does Stock Market Volatility Change Over Time?, *Journal of Finance* 44, 1115-1153.

[37] Severini, T. A. and W. H. Wong (1992), Profile Likelihood and Conditionally Parametric Models, *The Annals of Statistics* 20, 4, 1768-1802.

[38] Straumann, D. and T. Mikosch, T. (2006), Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach, *The Annals of Statistics* 34, 5, 2449-2495.

[39] Su, L. and S. Jin (2008), Profile Quasi-Maximum Likelihood Estimation of Partially Linear Spatial Autoregressive Models, *Journal of Econometrics*, forthcoming.

[40] Sun, Y. and T. Stengos (2006), Semiparametric Efficient Adaptive Estimation of Asymmetric GARCH Models, *Journal of Econometrics* 133, 373-386.

Research Papers 2008

CREATES
Center for Research in Econometric Analysis of Time Series

2008-31: Ingmar Nolte and Valeri Voev: Estimating High-Frequency Based (Co-) Variances: A Unified Approach

2008-32: Martin Møller Andreasen: How to Maximize the Likelihood Function for a DSGE Model

2008-33: Martin Møller Andreasen: Non-linear DSGE Models, The Central Difference Kalman Filter, and The Mean Shifted Particle Filter

2008-34: Mark Podolskij and Daniel Ziggel: New tests for jumps: a threshold-based approach

2008-35: Per Frederiksen and Morten Ørregaard Nielsen: Bias-reduced estimation of long memory stochastic volatility

2008-36: Morten Ørregaard Nielsen: A Powerful Test of the Autoregressive Unit Root Hypothesis Based on a Tuning Parameter Free Statistic

2008-37: Dennis Kristensen: Uniform Convergence Rates of Kernel Estimators with Heterogenous, Dependent Data

2008-38: Christian M. Dahl and Emma M. Iglesias: The limiting properties of the QMLE in a general class of asymmetric volatility models

2008-39: Roxana Chiriac and Valeri Voev: Modelling and Forecasting Multivariate Realized Volatility

2008-40: Stig Vinther Møller: Consumption growth and time-varying expected stock returns

2008-41: Lars Stentoft: American Option Pricing using GARCH models and the Normal Inverse Gaussian distribution

2008-42: Ole E. Barndorff-Nielsen, Silja Kinnebrock and Neil Shephard: Measuring downside risk — realised semivariance

2008-43: Martin Møller Andreasen: Explaining Macroeconomic and Term Structure Dynamics Jointly in a Non-linear DSGE Model

2008-44: Christian M. Dahl, Henrik Hansen and John Smidt: The cyclical component factor model

2008-45: Christian M. Dahl and Yu Qin: The limiting behavior of the estimated parameters in a misspecified random field regression model

2008-46: Bent Jesper Christensen, Christian M. Dahl and Emma M. Iglesias: Semiparametric Inference in a GARCH-in-Mean Model