SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2009-59

Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary

Christian M. Dahl and Emma M. Iglesias

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary*

Christian M. Dahl†
CREATES, School of Economics and Management
University of Aarhus

Emma M. Iglesias
Department of Economics
Michigan State University and University of Essex

October 2, 2009

Abstract

In this paper a new GARCH–M type model, denoted the GARCH-AR, is proposed. In particular, it is shown that it is possible to generate a volatility-return trade-off in a regression model simply by introducing dynamics in the standardized disturbance process. Importantly, the volatility in the GARCH-AR model enters the return function in terms of relative volatility, implying that the risk term can be stationary even if the volatility process is nonstationary. We provide a complete characterization of the stationarity properties of the GARCH-AR process by generalizing the results of Bougerol and Picard (1992b). Furthermore, allowing for nonstationary volatility, the

*Comments from Tim Bollerslev, Lynda Khalaf, Nour Meddahi, Jeffrey Wooldridge and participants at seminars at University of Aarhus and at the conferences: 2005 *Symposium on Econometric Theory and Applications (Taipei, Taiwan)*; the 2006 *North American Summer meeting of the Econometric Society (Minnesota), 2006 IMS Annual Meeting in Statistics (Rio de Janeiro), 2006 Canadian Econometrics Group (Niagara Falls, Canada), 2006 European Meeting of the Econometric Society (Vienna)* and at the 2009 Conference on Periodicity, Non-stationarity and Forecasting of Economic and Financial Time Series (Aarhus) are gratefully acknowledged. The first author acknowledges the research support of CREATES (funded by the Danish National Research Foundation). The second author acknowledges support from the MSU Intramural Research Grants Program.

†Corresponding author. Address: School of Economics and Management, Room 225, Building 1326. Phone: +45 8942 1559. E-mail: cdahl@creates.au.dk.

asymptotic properties of the estimated parameters by quasi-maximum likelihood in the GARCH-AR process are established. Finally, we stress the importance of being able to choose correctly between AR-GARCH and GARCH-AR processes: First, it is shown, by a small simulation study, that the estimators for the parameters in an AR-GARCH model will be seriously inconsistent if the data generating process actually is a GARCH-AR process. Second, we provide an LM test for neglected GARCH-AR effects and discuss its finite sample size properties. Third, we provide an empirical illustration showing the empirical relevance of the GARCH-AR model based on modelling a wide range of leading US stock return series.

Key words: Quasi-Maximum Likelihood; GARCH-M Model; Asymptotic Properties; Risk-return Relation.

JEL Codes: C12, C13; C22, G12.

1 Introduction

Since Merton's (1973) pathbreaking article deriving the intertemporal Capital Asset Pricing Model, the idea that higher risk assets must have higher expected returns to attract investors has been a foundation of modern finance theory. The GARCH-M (Generalized Autoregressive Conditional Heteroskedastic-in-mean) model proposed by Engle, Lilien and Robins (1987) allows for the direct effect of volatility changes on asset prices through required returns in a short memory GARCH-type model, by introducing the conditional volatility function into the conditional mean return equation. Numerous empirical studies have explored this volatility-return trade-off in attempts to estimate the magnitude of the trade-off itself. Perhaps surprisingly, both statistical significance and even the sign of the linear relation between expected return and variance of return have proved elusive in empirical work. Just naming a few studies and confining the listing to studies using some form of GARCH-in-mean models, for example, confirmation of the positive volatility-return trade-off is supported by French, Schwert, and Stambaugh (1987), Chou (1988), Baillie and DeGennaro (1990), and Campbell and Hentschel (1992), while evidence of a negative risk-return trade-off was found by Nelson (1991) and Glosten, Jagannathan, and Runkle (1993). Indeed, Campbell and Hentschel (1992) argue that an alternative source of a negative volatility-return relation is the volatility feedback mechanism, that is, if volatility is increased, then so is the risk premium, in case of a positive trade-off between risk and conditional expected return. Hence, the discount rate is also increased, which in turn, for an unchanged dividend yield, lowers the stock price. Negative premia is also found in recent work in asset pricing focussing on volatility innovations that examines cross-sectional risk premia induced by covariance between volatility changes and stock returns, 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 they should require smaller risk premia. For a survey of these and related studies, see Lettau and Ludvigson (forthcoming).

One possible source of the disagreement is misspecification of the way in which conditional variance enters the conditional mean return equation. Considerations and debates of this type have lead to the interest in specifying a more flexible model encompassing these and other alternative volatility-return relations. Flexibility, however, comes at the cost of more complicated statistical properties and typically also more restrictive assumptions on the data generating process. Hodgson and Vorkink (2003) estimate the density function of a multivariate GARCH-M model in a semiparametric fashion. Linton and Perron (2003), Sun and Stengos (2006), Conrad and Mammen (2008) and Christensen, Dahl and Iglesias (2008) all propose to 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 $t-1$, $\varepsilon_t \sim i.i.d.\,(0,1)$, and $\mu\left(\cdot\right)$ is a smooth mean function determining the functional form of the volatility-return relation. Common for all these papers is that the specification is estimated semiparametrically, assuming stationary GARCH or exponential GARCH processes for the conditional variance. Indeed, based on this approach and using CRSP excess return data, a hump shaped (nonlinear) pattern of the risk premium is generally supported.

However, as pointed out by Christensen, Nielsen and Zhu (2009), a potential statistical problem with the traditional GARCH-M specification is that possible nonstationarity or long memory properties of the volatility process may not balance well with the short memory properties of the return series in the regression model. Whether estimating the GARCH-M parametrically or semiparametrically, this might give unwanted and spurious results. The empirical evidence supporting the long memory properties of the volatility is extensive and includes studies by Robinson (1991), Crato and de Lima (1994), Baillie et al (1996), Ding and Granger (1996), Breidt, Crato and de Lima (1998), Robinson (2001), Andersen, Bollerslev, Diebold and Labys (2003) and Christensen et al (2009). Consequently, in the recent literature, see, e.g., Ang et al (2006), Christensen and Nielsen (2007) and Christensen et al (2009), it has been suggested to use changes of volatility as the M-term in the GARCH-M type return equations.

In this paper, we propose a new GARCH-M model which allows for an alternative parametric relation across conditional means and variances relative to the traditional GARCH-M of Engle et al (1987). In particular, we show that by introducing simple dynamics in the standardized disturbance process of the return series (described in details later), it is possible to generate a volatility-return trade-off in a regression model setting. Importantly, the volatility will enter the return function in terms of relative volatility, implying that even if the volatility process is nonstationary, the risk term will maintain stationarity as required. For reasons that will be obvious later, and to differentiate our model from the traditional GARCH-M of Engle, Lilien and Robins (1987), we will refer to the new GARCH-M type model as the GARCH-AR model.

One unfortunate and potentially serious problem associated with all parametric GARCH-M models (with or without stationary volatility) and with semiparametric GARCH-M models (with nonstationary volatility) is the lack of asymptotic theory associated with the estimated parameters. Lending on the techniques of Jensen and Rahbek (2004a, 2004b), we are able to establish the asymptotic normality of the Quasi Maximum Likelihood Estimator (QMLE) in the proposed GARCH-M type specification, allowing for possible explosive and nonstationary behavior of the GARCH process. We provide a general asymptotic theory that holds both in the stationary and the nonstationary regions of the parameter space. We hereby fill in an important gap in the literature by facilitating simple quasi-likelihood based inference in presence of risk premiums.

Finally, we provide some empirical illustrations. First, it is shown, by a small simulation study, that the estimators for the parameters in an AR-GARCH model will be seriously inconsistent if the data generating process actually is a GARCH-AR process. Second, we propose an LM test for neglected GARCH-AR effects and discuss its finite sample size properties. Third, we provide an empirical illustration showing the empirical relevance of the GARCH-AR model based on modelling a wide range of leading US stock return series.

An important final remark is that the asymptotic theory applies to the stationary as well as the nonstationary regions of the parameter space. In financial data it is very common to be very close to the boundary of the stationary region and to have Integrated GARCH type effects. In this paper we show that regardless if we are in the stationary or the nonstationary region, the researcher can use the asymptotic normality of the QMLE in the new model.

The plan of the paper is as follows. Section 2 motivates our new GARCH-M type model in the risk-return trade-off literature. Section 3 establishes the strict stationarity of a GARCH(1,1) model with ergodic and strictly stationary standardized disturbances. Section 4 presents the simplest version of new GARCH-M type model and it provides results on consistency and asymptotic normality of the QML estimator. Section 5 generalizes Section 3. Section 6 presents some illustrations of our approach: first we provide a Monte Carlo simulation study on the consequences of dynamic misspecification of the GARCH model; second, we propose and illustrate an LM test, and third, we present an empirical application based on modelling US riskless rate of return series. Finally, Section 7 concludes. The proofs are collected in Appendices 1-3.

2 The new GARCH-M type representation

For illustrative purposes consider initially the process y_t, with the representation

$$y_t = \sigma_t \epsilon_t, \tag{1}$$

$$\epsilon_t = \mu + \rho \epsilon_{t-1} + \upsilon_t, \tag{2}$$

$$\upsilon_t \sim i.i.d.(0,1), \tag{3}$$

where σ_t^2 has, say, a GARCH(1,1) specification. Relative to the existing literature, this model contains one new term given by $\rho\epsilon_{t-1}$. The contribution of this term and the link to the traditional GARCH-M model can be seen more explicitly if we rewrite (1)-(3) as

$$y_t = \underbrace{\mu\sigma_t}_{\text{Term 1}} + \underbrace{\rho \frac{\sigma_t}{\sigma_{t-1}} y_{t-1}}_{\text{Term 2}} + \underbrace{\sigma_t \upsilon_t}_{\text{Term 3}}, \tag{4}$$

Note that Terms 1 and 3 sum up to the standard GARCH-M model. However, when $\rho \neq 0$, Term 2 becomes relevant. Some of the implications on the risk-return relationship of this new term can be seen from the following derivative

$$E\left(\frac{\partial y_t}{\partial \sigma_t} | I_{t-1}\right) = \mu + \frac{\rho}{\sigma_{t-1}} y_{t-1}.$$

In the standard GARCH-M model, if volatility increases by one unit we would expect the mean return (or excess return) to increase by μ. However from (4), in addition, we would expect the return to change by an additional amount as a result of the new term, namely $\frac{\rho}{\sigma_{t-1}} y_{t-1}$. In summary, the risk-return relationship becomes state dependent. Furthermore, notice that this expected change ($E\left(\frac{\partial y_t}{\partial \sigma_t} | I_{t-1}\right)$) would be larger if the increase in volatility came from a low level at t-1, i.e., σ_{t-1} is small when considering a change in σ_t. Another, important implication from Term 2 is that a negative (excess) return in the previous period ($y_{t-1} < 0$) could cause the risk premium to be smaller than μ and may even be negative if the negative return comes at a time where the volatility level in the market is low.

Equation (4) also illustrates that it is critical how to model the dynamics. Traditionally, lagged values of the return series have been included in the conditional mean equation as in the traditional Autoregresive-GARCH (AR-GARCH) model . However, if the dynamics should enter the process for ϵ_t, then the AR-GARCH will be misspecified. It is well known that misspecification of the conditional mean function can have very serious effects on the estimates in GARCH-type models. In the empirical section, we will analyze the effects of this type of dynamic misspecification further.

Finally, it is clear from equation (4) that if volatility is nonstationary, for example if $\sigma_t \to \infty$ as in Jensen and Rahbek (2004a, 2004b), then the risk-return relationship becomes unbalanced in the standard GARCH-M model due to Term 1. However, if we restrict $\mu = 0$, then as long as the ratio $\frac{\sigma_t}{\sigma_{t-1}}$ is stationary, the risk-return relationship maintains a balance. In what follows, this scenario where σ_t may be nonstationary will serve as one of the main focus points of the theoretical and empirical analysis. We name the model that allows for Terms 2 and 3 in (4), the GARCH-AR(**1**) model.

3 Strict stationarity and weak-GARCH(1,1) processes

Following Drost and Nijman (1993), GARCH models can be classified into three categories: strong GARCH (where the innovation process ϵ_t is i.i.d. with zero mean and unit variance); the semi-strong form (where the innovation process is a martingale difference sequence), and finally the weak form, where also the martingale difference sequence assumption is relaxed.

Asymptotic normality of the QMLE has been established under a wide range of alternative assumptions for strong-GARCH models (see e.g. Lumsdaine (1996) and Berkes and Horváth (2004)). One of the most recent contribution is by Jensen and Rahbek (2004a, 2004b), who show that the QMLE is asymptotically normal both in the stationary and the nonstationary region of the parameter space. Lee and Hansen (1994) show asymptotic normality of the QMLE for the semi-strong form. In relation to weak-form GARCH models, the asymptotic normality of the QML estimator has not yet been established, see, e.g., Linton and Mammen (2005, see page 787). Indeed Franqc and Zakoïan (2000) consider a weak-GARCH type model where the squared of the errors are modelled explicitly with ARMA (Autoregressive Moving Average) processes, but they show the properties of estimating such a model with two-stage least squares instead of with QML.

GARCH processes where the innovation is not i.i.d. have mainly been treated semiparametrically in the literature, see, e.g., Meddahi and Renault (2004), Linton and Mammen (2005) and Dahl and Levine (2006). However, common for the semiparametric approaches, they do not permit nonstationary and explosive GARCH processes. The *first* main contribution of this paper is to analyze a fully parametrically specified GARCH process, where the standardized disturbances are neither i.i.d. nor a martingale difference sequence and by allowing for explosive and nonstationary behavior of the GARCH process. Papers such as Francq and Zakoïan (2000), Meddahi and Renault (2004), Linton and Mammen (2005) and Dahl and Levine (2006) provide evidence that allowing for a structure in ϵ_t that is non-i.i.d. is relevant in Economics and Finance. We are hereby extending the work of Jensen and Rahbek (2004a, 2004b) from strong-GARCH processes to GARCH with serially dependent standardized disturbances. Linton and Mammen (2005) argue that one of the important challenges in relation to modelling weak-form GARCH processes parametrically is to avoid higher order moment restrictions on the dependent variable. That is the reason of why all GARCH processes with non i.i.d.-noise have been treated so far only semiparametrically or nonparametrically. Franqc and Zakoïan (2000) are an exception, but they use two-stage least squares as the estimation method instead of QML. Importantly, we show the moment restrictions needed for asymptotic normality of the QMLE are not stronger for the GARCH process with serially dependent standardized disturbances we consider relative to the strong-GARCH specification in Jensen and Rahbek (2004a, 2004b).

It should be emphasized that the GARCH representation with non-i.i.d. noise we consider in this paper is not new in the literature. It has been used previously, for example, by Gonçalves and White (2004, page 208). Kristensen and Rahbek (2005) also note the relevance of this model. However, the asymptotic properties of this model and its GARCH-M type representations has not previously been explored under the best of our knowledge.

One of the key elements in the derivation of the asymptotic results for GARCH processes, including the new GARCH-M type representation, is strict stationarity of y_t. To keep a high level of generality, we therefore consider the following representation

$$y_t = \sigma_t(\boldsymbol{\theta}^*)\,\epsilon_t, \tag{5}$$

$$\sigma_t^2(\boldsymbol{\theta}^*) = w + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}^*), \tag{6}$$

for $t = 1, 2, ..., T$, where we allow ϵ_t to be a dependent process. Following Jensen and Rahbek (2004b) as to the initial values, the analysis is conditional on the observed value y_0 and if expressions involved in (5)-(6) are well defined, we assume that the initial value of ϵ_t is drawn from the strictly stationary distribution and the unobserved variance is parameterized by $\gamma = \sigma_0^2$. The vector of parameters of interest is then given as $\boldsymbol{\theta}^* = (w, \alpha, \beta, \gamma)' \in \Theta^*$, where the true parameter vector is defined as $\boldsymbol{\theta}_0$. In establishing strict stationarity, we will work under the following set of general assumptions which are maintained throughout the paper

Assumption A

A1 $\Theta^* \subset \mathbb{R}_+^4$ is a compact set.

A2 ϵ_t is an ergodic and strictly stationary process.

For simplicity of notation we write $\sigma_t^2 = \sigma_t^2(\boldsymbol{\theta}_0)$. We can now establish the following result regarding strict stationarity of y_t, which is an extension of the results of Bougerol and Picard (1992a, 1992b).

Lemma 1 Let Assumption A hold. A necessary and sufficient condition for strict stationarity of y_t generated by (5)-(6) is given by

$$\mathrm{E}\log\left(\alpha_0\epsilon_t^2 + \beta_0\right) < 0.$$

Proof of Lemma 1 Given in Appendix 1.

An additional and important new result (extending the results in Nelson (1990)), under the broad definition of ϵ_t in Assumption A,can now be formulated as follows

Lemma 2 Let Assumption A hold. If $\mathrm{E}\log\left(\alpha_0\epsilon_t^2 + \beta_0\right) \geq 0$ as $t \to \infty$, then $\sigma_t^2 \overset{a.s.}{\to} \infty$.

Proof of Lemma 2 Given in Appendix 1.

Lemmas 1 and 2 generalize the well known conditions for strict stationarity of strong-GARCH processes, to include weak-GARCH processes. The results presented in Lemmas 1 and 2 on strict stationarity also apply directly to the new GARCH-M type model, i.e., the GARCH-AR specification.

4 The GARCH(1,1)-AR(1) model and its properties

To obtain quasi-likelihood based estimators of the unknown parameters in $\boldsymbol{\theta}^*$, we proceed by parameterizing the process generating ϵ_t. We will begin under the simplifying assumption that ϵ_t follows a strictly stationary and ergodic AR(1) process. Later we will generalize all the results to the AR(p) case. In order to be precise about the dynamic order of the new GARCH-M type model we will, for the remainder of the paper, mostly refer to it as the GARCH(1,1)-AR(p) representation. To keep matters simple, however, we begin considering the GARCH(1,1)-AR(1) representation given as

$$
\begin{aligned}
y_t &= \sigma_t(\boldsymbol{\theta})\,\epsilon_t(\boldsymbol{\theta}), & (7)\\
\epsilon_t(\boldsymbol{\theta}) &= \rho\epsilon_{t-1}(\boldsymbol{\theta}) + v_t, & (8)\\
\sigma_t^2(\boldsymbol{\theta}) &= w + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}), & (9)
\end{aligned}
$$

where the parameter vector is redefined as $\boldsymbol{\theta}^* = (w, \alpha, \beta, \gamma, \rho)' \in \Theta$ with true values $\boldsymbol{\theta}_0 = (w_0, \alpha_0, \beta_0, \gamma_0, \rho_0)'$ and where v_t is a white noise process defined by Assumption B below. We write $\left(\epsilon_t, \sigma_t^2\right) = \left(\epsilon_t(\boldsymbol{\theta}_0), \sigma_t^2(\boldsymbol{\theta}_0)\right)$.

To illustrate a potential and important difference, we will first consider the unconditional variance of the process given by (7) - (9) and compare it to the unconditional variance generated from the strong-GARCH process. For illustration only, let us restrict attention to the case where the unconditional variance is bounded and σ_t^2 is a strictly stationary and ergodic process. We immediately have that

$$
\mathrm{E}\left(\sigma_t^2\right) = w_0 + \alpha_0\,\mathrm{E}\left(\sigma_{t-1}^2\epsilon_{t-1}^2\right) + \beta_0\,\mathrm{E}\left(\sigma_{t-1}^2\right),
$$

where

$$
\begin{aligned}
\mathrm{E}\left(\sigma_{t-1}^2\epsilon_{t-1}^2\right) &= \mathrm{E}\left(\sigma_{t-1}^2\left(\rho_0\epsilon_{t-2} + v_{t-1}\right)^2\right),\\
&= \rho_0^2\,\mathrm{E}\left(\sigma_{t-1}^2\epsilon_{t-2}^2\right) + \mathrm{E}\left(\sigma_{t-1}^2\right),
\end{aligned}
$$

such that,

$$\mathrm{E}\left(\sigma_t^2\right) = \frac{w_0 + \alpha_0 \rho_0^2 \,\mathrm{E}\left(\sigma_{t-1}^2 \epsilon_{t-2}^2\right)}{(1 - \alpha_0 - \beta_0)}.$$

Consequently, the difference between the unconditional variance of the standard GARCH(1,1) process and the GARCH(1,1)-AR(1) generated from (7) - (9) is equal to $\alpha_0 \rho_0^2 \,\mathrm{E}\left(\sigma_{t-1}^2 \epsilon_{t-2}^2\right) / (1 - \alpha_0 - \beta_0)$ and the GARCH(1,1)-AR(1) model will be able to generate relative larger (smaller) unconditional variance depending on the magnitude of this term. Again we would like to emphasize that stationarity of σ_t^2 is not required in the estimation framework presented below, so the unconditional variance may not exist in the GARCH(1,1)-AR(1) setting. This feature introduces additional flexibility in the conditional volatility process modelling relative to the traditional GARCH(1,1) model (since when $\rho_0 = 0$ we are back in the traditional GARCH(1,1)).

A second important issue is how to interpret the conditional variance function $\sigma_t^2(\boldsymbol{\theta}^*)$ in the representation given by (7) - (9). In what follows, let $\eta_t = \sigma_t(\boldsymbol{\theta}^*)\, v_t$ and define the information set at time $t-1$ as $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$. To help the understanding, let us compare to the traditional AR(1)-ARCH(1) model, see e.g., Ling and McAleer (2003, page 283), where

$$\begin{aligned} y_t - \phi y_{t-1} &= \eta_t, & (10)\\ \sigma_t^2(\boldsymbol{\theta}^*) &= w + \alpha \eta_{t-1}^2. & (11) \end{aligned}$$

Here the interpretation is that $\sigma_t^2(\boldsymbol{\theta}^*)$ is the conditional variance (conditional on information up to and including $t-1$) since $\sigma_t^2(\boldsymbol{\theta}^*) = \mathrm{E}\left((y_t - \phi y_{t-1})^2 \,|I_{t-1}\right) = \mathrm{var}_t(\eta_t)$, where $\mathrm{var}_t(\eta_t)$ is the conditional variance of η_t. Note, that in (10)- (11), $\sigma_t^2(\boldsymbol{\theta}^*)$ is a function of η_{t-1}^2 which is the original idea of Engle (1982). Alternatively, consider the double autoregressive model of Ling (2004), where

$$\begin{aligned} y_t - \phi y_{t-1} &= \eta_t,\\ \sigma_t^2(\boldsymbol{\theta}^*) &= w + \alpha y_{t-1}^2. \end{aligned}$$

Again the interpretation is that $\sigma_t^2(\boldsymbol{\theta}^*) = \mathrm{E}\left((y_t - \phi y_{t-1})^2 \,|I_{t-1}\right) = \mathrm{var}_t(\eta_t)$. It does not change this interpretation that $\sigma_t^2(\boldsymbol{\theta}^*)$ is a function of y_{t-1}^2 which is different from η_{t-1}^2 in (11) whenever $\phi \neq 0$. Regarding the interpretation of $\sigma_t^2(\boldsymbol{\theta}^*)$ in the model given by (7) - (9), note that it can be re-represented as

$$\begin{aligned} y_t - \rho \sigma_t(\boldsymbol{\theta}^*)\, \sigma_{t-1}^{-1}(\boldsymbol{\theta}^*)\, y_{t-1} &= \eta_t, & (12)\\ \sigma_t^2(\boldsymbol{\theta}^*) &= w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\boldsymbol{\theta}^*), & (13) \end{aligned}$$

since the model allows the standardized disturbances as defined in the traditional GARCH model to be partially predicted through an autoregressive process. Note that with (12), we have returned to traditional context of having an i.i.d. innovation term η_t (therefore $\text{var}_t(\eta_t) = \text{var}(\eta_t)$). Also, this emphasizes the differences between a traditional AR-GARCH model with a constant AR coefficient given by ϕ in the mean equation versus the GARCH-AR having variable AR coefficients in the mean equation given by $\rho\sigma_t(\boldsymbol{\theta}^*)\sigma_{t-1}^{-1}(\boldsymbol{\theta}^*)$. However, because of the properties of η_t we again have that

$$\begin{aligned} \mathrm{E}\left(\left(y_t - \rho\sigma_t(\boldsymbol{\theta}^*)\sigma_{t-1}^{-1}(\boldsymbol{\theta}^*)y_{t-1}\right)^2 | I_{t-1}\right) &= \sigma_t^2(\boldsymbol{\theta}^*), \\ &= \text{var}_t(\eta_t). \end{aligned}$$

In conclusion, the interpretation of $\sigma_t^2(\boldsymbol{\theta}^*)$ as the conditional variance still holds in the GARCH(1,1)-AR(1) model and this result easily generalizes to any GARCH(1,1)-AR(p) model. This result is not surprising as the GARCH(1,1) with serially dependent standardized disturbances given by (7) - (9) is a generalization of a Ling's (2004) double autoregressive model with a function of the volatility in the mean equation (i.e. allowing for a risk premium effect).

From (12), it also becomes clear how by specifying an AR(1) process for the standardized disturbance, a measure of the volatility is introduced into the mean equation of the process y_t. Interestingly, this gives an easy test for the presence of risk effects in the mean equation: if the hypothesis that $\rho = 0$ cannot be rejected, then data supports the traditional GARCH(1,1) process without risk effects in the mean . Furthermore, as pointed out by Dufour, Khalaf, Bernard and Genest (2004), it is difficult to test for GARCH effects within the traditional GARCH-M framework due to the presence of identification problems. These problems are not present in the GARCH-AR specification. To see this let us define $\widetilde{\rho}_t = \rho\sigma_t(\boldsymbol{\theta}^*)\sigma_{t-1}^{-1}(\boldsymbol{\theta}^*)$. A test for no GARCH effects would reduce to a test of the null hypothesis $\widetilde{\rho}_t = \rho$ for all t, i.e., a test whether $\widetilde{\rho}_t$ is a constant versus a time varying parameter. This test is particular simple since under the null hypothesis as the standardized disturbance term will be i.i.d..

4.1 Estimation and the Asymptotics

In order to derive the asymptotic properties of the QMLE, in addition, we need to modify the set of assumptions we will be work under.

Assumption B

B1 $\Theta \subset \mathbb{R}_+^4 \times \mathbb{R}$ is a compact set, $\boldsymbol{\theta}_0 \in interior(\Theta)$ and $|\rho_0| < 1$.

B2 $\mathrm{E}\log\left(\alpha_0\epsilon_t^2 + \beta_0\right) \geq 0$.

B3 $v_t \sim i.i.d.(0,1)$ with $\mathrm{E}\left(v_t^4 - 1\right) = \zeta < \infty$.

Note that we impose Assumption B2 to explicitly focus first on the nonstationarity region for the GARCH process given by (5).

We now present the result of the asymptotic behavior of the QMLE in this setting. The QML estimators will maximize the quasi log-likelihood (with $\boldsymbol{\theta} = (w_0, \alpha, \beta, \gamma_0, \rho)'$) given as

$$l_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left(w_0 + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta})\right) + \frac{\left(y_t - \rho\sigma_t(\boldsymbol{\theta})\,\sigma_{t-1}^{-1}(\boldsymbol{\theta})\,y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta})}\right). \quad (14)$$

We then have,

Theorem 1 Let Assumption B hold with (w, γ) fixed at their true values, $(w_0, \gamma_0)'$, and assume that y_t is generated from (7) - (9). Then there exists a fixed open neighborhood $U = U(\alpha_0, \beta_0, \rho_0)$ of $(\alpha_0, \beta_0, \rho_0)'$ such that with probability tending to one as $T \longrightarrow \infty$, $l_T(\boldsymbol{\theta})$ given by (14) has a unique maximum point $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}\right)'$ in U. In addition, the QML estimator $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}\right)'$ is consistent and asymptotically normal, i.e.,

$$\sqrt{T}\left[\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}\right) - (\alpha_0, \beta_0, \rho_0)\right]' \overset{d}{\longrightarrow} \mathrm{N}\left(0, \boldsymbol{\Omega}^{-1}\right),$$

with $\mu_i = \mathrm{E}\left(\beta_0 / \left(\alpha_0\epsilon_t^2 + \beta_0\right)\right)^i, i = 1, 2$ and where

$$\boldsymbol{\Omega} = \begin{pmatrix} \frac{1}{\zeta\alpha_0^2} & \frac{\mu_1}{\zeta\alpha_0\beta_0(1-\mu_1)} & 0 \\ \frac{\mu_1}{\zeta\alpha_0\beta_0(1-\mu_1)} & \frac{(1+\mu_1)\mu_2}{\zeta\beta_0^2(1-\mu_1)(1-\mu_2)} & 0 \\ 0 & 0 & \frac{1}{(1-\rho_0^2)} \end{pmatrix}.$$

Proof of Theorem 1 Given in Appendix 1.

Three remarks in relation to Theorem 1:

Remark 1: Note that Theorem 1 relies on the nonstationary case and later Theorem 5 will focus on the stationary case. Since in this paper our proofs consider both the stationary and the nonstationary cases, that means that we do not need to estimate GARCH-AR models by QML constrained to the restriction of stationarity.

Remark 2: Note that Theorem 1 involves to deal with $\mu_i = \mathrm{E}\left(\beta_0/\left(\alpha_0\epsilon_t^2+\beta_0\right)\right)^i, i = 1, 2$ where we use ϵ_t instead of the i.i.d. innovation process in Jensen and Rahbek (2004b). Also $\mathrm{E}\left(v_t^4 - 1\right) = \zeta$.

Remark 3: When $\rho = 0$, The results in Theorem 1 match with the results in Jensen and Rahbek (2004b).
Furthermore,

Theorem 2 If $\mathrm{E}\log\left(\alpha_0\epsilon_t^2+\beta_0\right) > 0$, under the conditions given in Theorem 1, then the results in Theorem 1 hold for any arbitrary value of $w > 0$ and $\gamma > 0$.

Proof of Theorem 2 Theorem 2 is proved by a direct application of Theorem 2 of Jensen and Rahbek (2004b).

In order to make the proofs easier to follow, we have structured the derivatives in the Appendix such that comparisons to the results of Jensen and Rahbek (2004a, 2004b) are straightforward to make.

5 Asymptotics of the QMLE for the GARCH(1,1)-AR(p) process

We now generalize the results in Theorems 1 and 2 to the AR(p) case. Consider now the GARCH(1,1)-AR(p) model given as

$$y_t = \sigma_t(\boldsymbol{\theta})\,\epsilon_t(\boldsymbol{\theta}), \tag{15}$$

$$\rho(L)\,\epsilon_t(\boldsymbol{\theta}) = v_t, \tag{16}$$

$$\sigma_t^2(\boldsymbol{\theta}) = w + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}), \tag{17}$$

where $\rho(L) = 1 - \sum_{i=1}^p \rho_i L^i$ and $\boldsymbol{\theta} = (w_0, \alpha, \beta, \gamma_0, \rho_1, ..., \rho_p)' \in \Theta$ with its true value at $\boldsymbol{\theta}_0$. Again we write $\{\epsilon_t, \sigma_t^2\} = \{\epsilon_t(\boldsymbol{\theta}_0), \sigma_t^2(\boldsymbol{\theta}_0)\}$. As in the previous section it can be shown that the unconditional volatility of this process relative to the unconditional volatility (provided that they exists) will depend on $\boldsymbol{\theta}_0$, in particular, on the term

$$\frac{\alpha_0}{(1-\alpha_0-\beta_0)}\left(\sum_i \rho_{i0}^2\,\mathrm{E}\left(\sigma_{t-1}^2\epsilon_{t-i-1}^2\right) + 2\sum_i \rho_{i-10}\rho_{i0}\,\mathrm{E}\left(\sigma_{t-1}^2\epsilon_{t-i}\epsilon_{t-i-1}\right)\right).$$

It is again clear that there are some new extra-terms in the unconditional volatility that would not be present given that the data was generated from the traditional GARCH(1,1) process. Hence, more flexibility is introduced in the conditional variance of the GARCH(1,1)-AR(p) versus the GARCH(1,1).

For estimation purpose we next impose the following set of assumptions

Assumption C

C1 $\Theta \subset \mathbb{R}_+^4 \times \mathbb{R}^p$ is a compact set, $\boldsymbol{\theta}_0 \in interior\,(\Theta)$ and the roots of $\rho\,(z)$ are all outside the unit circle.

C2 $\mathrm{E} \log \left(\alpha_0 \epsilon_t^2 + \beta_0\right) \geq 0.$

C3 $v_t \sim i.i.d.\,(0,1)$ with $\mathrm{E}\left(v_t^4 - 1\right) = \zeta < \infty$

As Assumption C is more restrictive than Assumption A, Lemmas 1 and 2 still hold when data under this assumption is generated from (15)-(17). In order to deriving the limiting properties of the QML estimators, we represent (15)-(17) as

$$\begin{aligned} y_t - \sum_{i=1}^{p} \rho_i \sigma_t(\boldsymbol{\theta})\, \sigma_{t-i}^{-1}(\boldsymbol{\theta})\, y_{t-i} &= \sigma_t(\boldsymbol{\theta})\, v_t, \\ \sigma_t^2(\boldsymbol{\theta}) &= w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\boldsymbol{\theta}). \end{aligned}$$

As in the previous section, we note that the AR(p) structure in the innovation process generates a measure of risk premium in the mean equation. With (w, γ) fixed at their true values, (w_0, γ_0) ,the quasi log-likelihood function associated with (15)-(17) then writes

$$l_T(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{T} \left(\log\left(w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\boldsymbol{\theta})\right) + \frac{\left(y_t - \sum_{i=1}^{p} \rho_i \sigma_t(\boldsymbol{\theta})\, \sigma_{t-i}^{-1}(\boldsymbol{\theta})\, y_{t-i}\right)^2}{\sigma_t^2(\boldsymbol{\theta})} \right). \tag{18}$$

We can now state the most important results of the paper.

Theorem 3 Let Assumption C hold and let data be generated according to the GARCH(1,1)-AR(p) process given by (15)-(17) Then the results of Theorem 1 hold. The variance-covariance matrix $\boldsymbol{\Omega}$ will be block diagonal. The upper block will be unchanged while the lower block will consist of the information matrix of the AR(p) process as given in Hamilton (1994, page 125) using the Galbraith and Galbraith (1974) equation.

Proof of Theorem 3 Given in Appendix 2.

As in the GARCH(1,1)-AR(1) case we have the following companion result to Theorem 3.

Theorem 4 If $\mathrm{E}\log\left(\alpha_0\epsilon_t^2+\beta_0\right)>0$, then the results in Theorem 3 hold for any arbitrary value of $w>0$ and $\gamma>0$.

Proof of Theorem 4 Theorem 4 is proved by a direct application of Theorem 2 of Jensen and Rahbek (2004b).

Very importantly, it is straightforward to show that, using similar techniques as in Jensen and Rahbek (2004b), the results in Theorem 1 and 3 carry over to the stationary case by the ergodic theorem. This important result can be summarized as follows

Theorem 5 Let Assumptions C1 and C3 hold and assume that $\mathrm{E}\log\left(\alpha_0\epsilon_t^2+\beta_0\right)<0$. Furthermore, let data be generated according to the GARCH(1,1)-AR(p) process given by (15)-(17). Then the results of Theorem 3 hold.

Proof of Theorem 5 Given in Appendix 3.

By Theorems 3, 4 and 5, we have generalized the main results of Lumsdaine (1996) and Jensen and Rahbek (2004a, 2004b) to a much richer class of volatility models. In the next section we will turn to the subject of the empirical relevance of the new approach.

6 Illustrations

In this section we illustrate the importance and need of carefully considering whether to specify the dynamics of the GARCH process in the conditional mean function (i.e., by choosing the AR-GARCH specification), a GARCH-M model or to specify the dynamics in the standardized disturbance by working with the GARCH-AR specification. We begin by analyzing the effects on the estimated parameters in the conditional mean function if it is misspecified, i.e., if the researcher estimates an AR-GARCH model when the true data generating process is GARCH-AR. We derive the relative inconsistency theoretically and we quantify it under various distributional assumptions by conducting simulations. Second, we present an LM test for neglected serially dependent standardized disturbances in GARCH models and we analyze its size properties in finite samples. Third, we provide

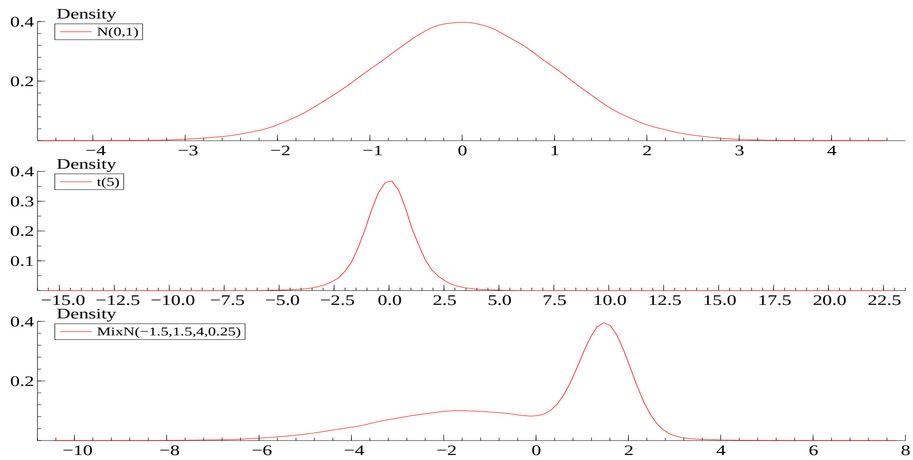

Figure 1: Alternative densities for v_t.

an empirical illustration showing that the GARCH-AR model, in some cases, may be a better alternative relative to the AR-GARCH model when fitting leading stock return series.

6.1 Effects of misspecification: A simple illustration

Consider the time series of interest y_t being generated according to the following GARCH(1,1)-AR(1) process parameterized as

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t, && (19)\\
\epsilon_t &= \phi_0 \epsilon_{t-1} + v_t, && (20)
\end{aligned}
$$

where v_t is white noise and $\sigma_t^2 = 1 + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$. Assume, for simplicity that a) the researchers primary interest is in estimating ϕ_0 and b) that she can observe the conditional variance function perfectly. However, the researcher "wrongly" assumes that the model is given by an AR(1)-GARCH(1,1), hence uses the representation in the mean given as

$$y_t = \phi y_{t-1} + \sigma_t v_t, \tag{21}$$

for estimation of the parameter ϕ_0. The main interest is to analyze the consequences of using the wrong model and to establish the asymptotic properties of the estimator of ϕ based on (21) given that σ_t is known and data is generated from (19). When σ_t is known,

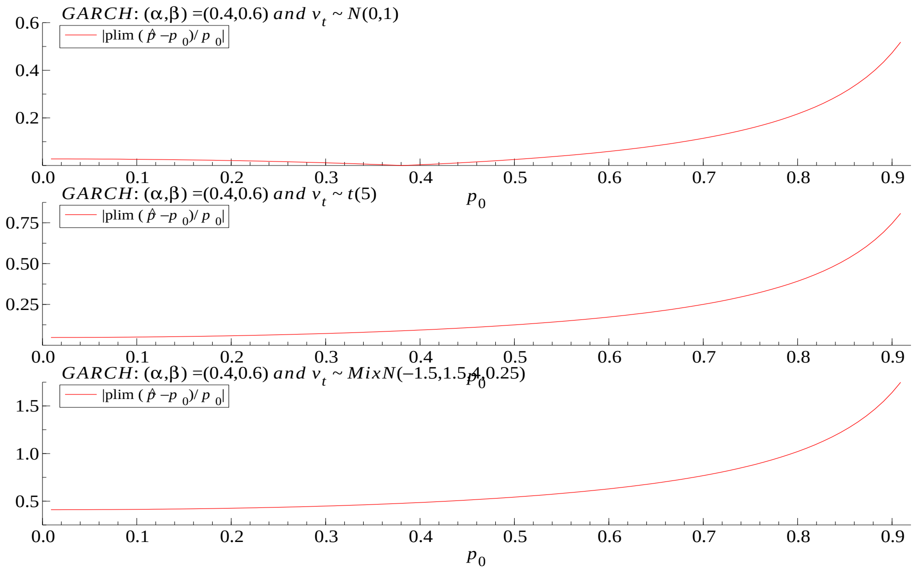

Figure 2: Measures of absolute relative inconsistency for alternative values of $|\phi_0|$ and alternative distributional assumptions.

a proper estimator of ϕ_0 is the WLS (Weighted Least Squares) estimator $\widehat{\phi}$ given as

$$\begin{aligned} \widehat{\phi} &= \frac{\frac{1}{T}\sum \sigma_t^{-2} y_t y_{t-1}}{\frac{1}{T}\sum \sigma_t^{-2} y_{t-1}^2}, \qquad (22) \\ &= \phi_0 \frac{\frac{1}{T}\sum \left(\sigma_t^2/\sigma_{t-1}^2\right)^{1/2} \left(\sigma_t^{-2} y_{t-1}^2\right)}{\frac{1}{T}\sum \sigma_t^{-2} y_{t-1}^2} + o_p(1), \end{aligned}$$

For simplicity assume that $y_t^2 \overset{a.s.}{\to} \infty$. Then from straightforward calculations we have $\sigma_t^{-2} y_{t-1}^2 \to 1/\alpha$, and $\sigma_t^2/\sigma_{t-1}^2 = 1/\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \beta$. By inserting in (22) it follows that

$$\left| \frac{\text{plim}\left(\widehat{\phi} - \phi_0\right)}{\phi_0} \right| = \left| \text{plim}\left(\frac{1}{T} \sum \left(\alpha\epsilon_{t-1}^2 + \beta\right)^{1/2} \right) - 1 \right|. \qquad (23)$$

We will denote $\left|\text{plim}\left(\widehat{\phi} - \phi_0\right)/\phi_0\right|$ as the measure of absolute relative inconsistency. From (23) it can be seen that estimating ϕ_0 based on (21) when (19) is the actual data generating process generally leads to relative inconsistency except in the trivial case when $\phi_0 = 0$. In general plim $\frac{1}{T}\sum\left(\alpha\epsilon_{t-1}^2 + \beta\right)^{1/2}$ can be far away from unity in particular for larger values of $|\phi_0|$ that will generate a large variance of ϵ_{t-1}^2.

To quantify the measure of absolute relative inconsistency under different distributional assumptions on v_t and for alternative values of $|\phi_0|$, we next conduct a small simulation study. We consider v_t being generated from three alternative densities depicted in Figure 1. As we expect the relative inconsistency to grow with the variance of v_t, we include the standard Gaussian density as well as a student$-t$ density with 5 degrees of freedom and a Gaussian mixture density.

In Figure 2 we have depicted the measure of absolute relative asymptotic inconsistency for alternative values of $|\phi_0|$ when $(\alpha, \beta) = (0.4, 0.6)$. Not surprisingly, the inconsistencies become more severe as the variance of ϵ_{t-1} increases. Note that $var(\epsilon_{t-1})$ increases with ϕ_0 and when the density of v_t goes from the normal to densities with fatter tails such as the student$-t$ and Gaussian mixture. Such fat-tailed distributions are common in financial time series. In general, the inconsistencies seem significant even for very small values of $|\phi_0|$ ranging from 5% in the Gaussian case to about 45% for the Gaussian mixture. This simple example stresses that careful model evaluation is needed to determine whether to model the dynamics in the standardized disturbance or in the conditional mean function, particularly when the density of the data exhibit fat-tail behavior. This evidence emphasizes again that the choice between the traditional AR-GARCH model with constant AR coefficients in the mean equation versus the GARCH-AR with variable AR coefficients potentially is of great importance.

6.2 Empirical illustration

In this section we provide an empirical motivation for the GARCH-AR specification. First, we suggest a simple testing procedure for the presence of the new GARCH-M type risk-return trade-off in AR-GARCH models. Secondly we illustrate the advantages of the GARCH-AR specification versus other traditional GARCH-type models in terms of model fitting leading US stock returns.

6.2.1 Testing

In order to test for the presence of serially correlated standardized disturbances in GARCH models, we propose an extension of the LM test of Lundbergh and Teräsvirta (2002). In the model given by equations (15)-(17), the null hypothesis corresponds to the case where $H_0 : \rho_i = 0, \forall i = 1, ...p$. The LM test can therefore be computed according to the following 3-step procedure:

Step 1: (Estimating the model under the null) Estimate the parameters of the conditional variance model using a traditional GARCH model under $H_0 : \rho_i = 0, \forall i = 1, ...p$.

Compute the standardized residuals $\widehat{\epsilon}_t$ and the residual sum of squares under the null hypothesis $SSR_0 = \sum_{t=1}^{T} \widehat{\epsilon}_t^2$.

Step 2: (Estimating the auxiliary regression) Regress $\widehat{\epsilon}_t$ from Step 1 on p lags of $\widehat{\epsilon}_t$, i.e. $(\widehat{\epsilon}_{t-1}, ..., \widehat{\epsilon}_{t-p})'$, and compute the sum of squared residuals SSR_1 associated with this regression.

Step 3: (Form the LM test statistic) Compute the χ^2 version of the test statistic as

$$LM_{\chi^2} = T\left(SSR_0 - SSR_1\right)/SSR_0 \sim \chi^2\left(p\right),$$

or the $F-$version as

$$LM_F = \frac{\left(SSR_0 - SSR_1\right)/p}{SSR_1/\left(T-3-p\right)} \sim F\left(p, T-3-p\right).$$

In Table 1, the extent to which the asymptotic distribution of the LM statistic approximates its distribution in finite samples is illustrated. We use two models for simulations under the null; the GARCH(1,1) and an AR(5)-GARCH(1,1) with population parameters fixed at the QML estimated parameters using IBM stock return series (to secure empirical relevance of the simulations). From inspection of Table 1 it is apparent that the nominal size is quite well matched by the simulated small sample distribution of the LM_{χ^2} distribution. The test seems to be only slightly oversized, even when the AR(5) mean function has to be estimated.

[Table 1 about here]

6.2.2 Estimation

Based on 11 leading US Stock returns series[1], we have reported results from estimation of the simple GARCH(1,1) model in Table 2.[2] A few stylized facts are confirmed: First, the estimated sum of α and β is close to unity, and secondly, dynamic misspecification in the conditional mean is often observed across the columns of the table. Specifically,

[1]Data corresponds to the following 11 companies: IBM, Intel, Microsoft, General Motors (GM), Ford, General Electric (GE), Boing, Walmart, McDonald's, TimeWarner and Sony.

[2]Data on the individual stock are collected from Yahoo Finance and are closing prices. Returns are computed as daily log returns. The last observation for all stock are 9.10.2007. Information on the dates for the first observation are given in Tables 2-4.

the GARCH(1,1) is rejected for GM, Ford, GE, Walmart, McDonald's, Time Warner and Sony at 5% significance level based on the LM test described in the previous section.[3]

[Table 2 about here]

In Table 3, we have augmented the GARCH(1,1) model by including dynamics in the conditional mean function. Indeed, this is the common practice when the GARCH(1,1) model appears dynamically misspecified. The effect of including dynamics on the estimated values of α and β is negligible. In all the models where the GARCH(1,1) model was dynamically misspecified according to the LM test in Table 2, one or more of the lagged dependent variables are significant in Table 3 (except for Intel, Microsoft, and Boing). However, only in the model for Ford and Sony, the inclusion of lagged dependent variables matters in terms of dynamic misspecification according to the LM test at the 5% significance level. In summary, the LM test indicates that the AR-GARCH models for GM, GE, Walmart, McDonald's and Time Warner still appear misspecified after the inclusion of lagged dependent variables in the conditional mean function.[4]

[Table 3 about here]

In Table 4, the results based on the GARCH(1,1)-AR(1) model are presented. For all the models where the GARCH(1,1) specification is misspecified, the dynamic term, represented by ϵ_{t-1}, becomes significant. Secondly, simply by including only a single additional term/coefficient in the model, the null hypothesis of a dynamically well specified model cannot be rejected for any of the return series considered, which is remarkable. Finally, it should be noticed that the estimates of ρ take positive as well as negative values depending on the return series considered.[5] Moreover, it is exactly for the series of GM, GE, Walmart, McDonald's and Time Warner (the ones where the AR-ARCH is not dynamically correctly specified in Table 3) where the estimate of ρ is statistically significant even at the 1% in Table 4.

[3] The LM tests in Tables 2-4 are all testing for serial correlation of order 1. The results, however, do not change when testing for serial correlation of higher order.

[4] Although not reported here, we estimated and tested a wide range of alternative dynamic specifications including up to 20 lags of the dependent variable. However, at large, this did not matter for the outcome of the LM test. Typically, higher order lagged terms were highly insignificant.

[5] We also estimated various versions of the classical GARCH-M model. The M-terms never turned up significant in any of the US stock return series.

[Table 4 about here]

7 Conclusion

In this paper we introduce a new parametric volatility model. One of the key features of the new model is that it can generate a volatility-return trade-off even when volatility is nonstationary and returns have short memory. A second important feature is that the model can be specified sufficiently simple facilitating a complete characterization of its stationarity properties and the asymptotics of the estimated parameters. Finally, we provide evidence of the usefulness of the new representation in a Monte Carlo experiment and in practical applications. We first show the consequences of dynamically misspecifying the GARCH model when the actual data generating mechanism is an GARCH-AR specification. We provide evidence of the large inconsistencies that this type of misspecification can generate. Next, we propose an LM test for neglected serially dependent standardized disturbances in AR-GARCH models. And finally, we provide an application based on a broad range of leading US stock return series where we show the empirical relevance of our new model.

Appendix 1

Appendix 1 provides the proof of Theorem 1. We need for that the use of Lemmas 1-5. The proof of Lemmas 1-5 are in the technical appendix which is available upon request from the authors.

Proof of Theorem 1 In order to allow for nonstationarity in the GARCH along the lines of Jensen and Rahbek (2004a, 2004b), we first find the expressions for the first, second and third order derivatives that are given in Results 1, 2 and 3. Lemmas 1 and 2 are used in our results. Later, Lemmas 3, 4 and 5 establish the Cramér type conditions. As in Jensen and Rahbek (2004a, 2004b) we also use the central limit theorem in Brown (1971). In order to make our results clear, we order the terms of the derivatives to find a similar structure as in Jensen and Rahbek (2004a, 2004b), in all those cases where this is possible. We also use Lemma 1 of Jensen and Rahbek (2004b) to prove uniqueness and the existence of the consistent and asymptotically Gaussian estimator. ■

Result 1: First order derivatives The first order derivatives are given by

$$\frac{\partial}{\partial z} l_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{\left(y_t-\rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta})}\right)\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})}+\frac{\frac{\partial\left(y_t-\rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)^2}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})}\right);\forall z=\alpha,\beta,$$

with

$$\begin{aligned}
\frac{\partial}{\partial\alpha} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{1t}(\boldsymbol{\theta}),\\
\frac{\partial}{\partial\beta} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{2t}(\boldsymbol{\theta}),\\
\frac{\partial}{\partial\rho} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{3t}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\frac{\left(y_t-\rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}}{\sigma_t(\boldsymbol{\theta})},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\alpha}}{\sigma_t^2(\boldsymbol{\theta})} &= \frac{\sum_{j=1}^{t}\beta^{j-1}y_{t-j}^2}{\sigma_t^2(\boldsymbol{\theta})},\\
\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\beta}}{\sigma_t^2(\boldsymbol{\theta})} &= \sum_{j=1}^{t}\beta^{j-1}\frac{\sigma_{t-j}^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})},\\
\frac{\partial\left(y_t-\rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)^2}{\partial z} &= \frac{-\left(y_t-\rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)\rho y_{t-1}}{\left(w+\alpha y_{t-2}^2+\beta\sigma_{t-2}^2(\boldsymbol{\theta})\right)^{3/2}\left(w+\alpha y_{t-1}^2+\beta\sigma_{t-1}^2(\boldsymbol{\theta})\right)^{1/2}}\\
&\quad\times\left(\left(w+\alpha y_{t-2}^2+\beta\sigma_{t-2}^2(\boldsymbol{\theta})\right)\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial z}-\left(w+\alpha y_{t-1}^2+\beta\sigma_{t-1}^2(\boldsymbol{\theta})\right)\frac{\partial\sigma_{t-1}^2(\boldsymbol{\theta})}{\partial z}\right)
\end{aligned}$$

for $\forall z=\alpha,\beta$. Then

Lemma 3 Let Assumption B hold. Define $s_{it}=s_{it}(\boldsymbol{\theta}_0)$, $\forall i=1,2,3$ where $s_{it}(\boldsymbol{\theta}_0)$ is defined in Result 1. Then

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{1t} &\xrightarrow{d} \mathrm{N}\left(0,\frac{\zeta}{4\alpha_0^2}\right),\\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{2t} &\xrightarrow{d} \mathrm{N}\left(0,\frac{\zeta(1+\mu_1)\mu_2}{4\beta_0^2(1-\mu_1)(1-\mu_2)}\right),\\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{3t} &\xrightarrow{d} \mathrm{N}\left(0,\frac{1}{(1-\rho_0^2)}\right),
\end{aligned}$$

with $\mu_i=\mathrm{E}\left(\beta_0/\left(\alpha_0\epsilon_t^2+\beta_0\right)\right)^i, i=1,2$ as $T\longrightarrow\infty$.

Result 2: Second order derivatives We have,

$$\begin{aligned}
\frac{\partial^2}{\partial z_1\partial z_2}l_T(\boldsymbol{\theta}) &= \frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta})}\right)\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4(\boldsymbol{\theta})}+\left(\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta})}-1\right)\frac{\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^2(\boldsymbol{\theta})}\right)\\
&\quad+\frac{1}{2}\sum_{t=1}^{T}\left(-\frac{\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2}}{\sigma_t^2}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\right)}{\sigma_t^4}\right),\\
\frac{\partial^2}{\partial\rho^2}l_T(\boldsymbol{\theta}) &= -\sum_{t=1}^{T}\sigma_{t-1}^{-2}y_{t-1}^2,\\
\frac{\partial^2}{\partial z\partial\rho}l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\left(\frac{\frac{\partial\left[\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z}}{\sigma_t^2}-\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_{t-1}^{-1}y_{t-1}\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^3}\right),
\end{aligned}$$

where

$$
\begin{aligned}
\frac{\frac{\partial^2 \sigma_t^2}{\partial \alpha^2}}{\sigma_t^2} &= 2\frac{\sum_{j=1}^{t}(j-1)\beta^{j-2}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^2 \sigma_t^2}{\partial \beta^2}}{\sigma_t^2} &= 2\sum_{j=1}^{t}(j-1)\beta^{j-2}\frac{\sigma_{t-j}^2}{\sigma_t^2}, \\
\frac{\partial\left[\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z} &= \frac{y_t y_{t-1}}{2\sigma_t\sigma_{t-1}}\left(\frac{\partial\sigma_t^2}{\partial z}-\frac{\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right)-\rho y_{t-1}^2\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_{t-1}^2}-\frac{\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^4}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2} &= \left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)(-\rho y_{t-1})\left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right] \\
&\quad+\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)(\rho y_{t-1})\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t+\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^5\sigma_t^2}\right] \\
&\quad+\rho^2y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2},
\end{aligned}
$$

for $\forall z, z_1, z_2 = \alpha, \beta$. Then

Lemma 4 Under Assumption B, with the expressions of the second order derivatives from Result 2 evaluated at $\boldsymbol{\theta}_0$

(a) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha^2}l_T(\boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\xrightarrow{p}\frac{1}{2\alpha_0^2}>0$,

(b) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta^2}l_T(\boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\xrightarrow{p}\frac{(1+\mu_1)\mu_2}{2\beta_0^2(1-\mu_1)(1-\mu_2)}>0$,

(c) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\beta}l_T(\boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\xrightarrow{p}\frac{\mu_1}{2\alpha_0\beta_0(1-\mu_1)}$,

(d) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\rho}l_T(\boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\xrightarrow{p}0$,

(e) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\rho}l_T(\boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\xrightarrow{p}0$,

(f) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\rho^2}l_T(\boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\xrightarrow{p}\frac{1}{\left(1-\rho_0^2\right)}>0$,

as $T\longrightarrow\infty$.

Result 3: Third order derivatives We have,

$$\begin{aligned}
\frac{\partial^3}{\partial z_1^2 \partial z_2} l_T(\boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^3\sigma_t^2}{\partial z_1^2\partial z_2}}{\sigma_t^2} \\
&- \sum_{t=1}^{T}\left(1-\frac{3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6} \\
&+ \sum_{t=1}^{T}\frac{\frac{1}{2}\left(\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}+\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2}\right)}{\sigma_t^4} \\
&+ \sum_{t=1}^{T}\frac{\left(\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\right)}{\sigma_t^4} \\
&- \sum_{t=1}^{T}\left(2\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}-1\right)\frac{\left(\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\frac{1}{2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^4} - \frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1^2\partial z_2}}{\sigma_t^2} \\
&- \sum_{t=1}^{T}\frac{\left(\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2}\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2+2\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^6}, \\
\frac{\partial^3}{\partial\rho^3} l_T(\boldsymbol{\theta}) &= 0, \\
\frac{\partial^3}{\partial z_1\partial z_2\partial\rho} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6}+\sum_{t=1}^{T}\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^4} \\
&+\frac{1}{2}\sum_{t=1}^{T}\left(-\frac{\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2\partial\rho}}{\sigma_t^2}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2\partial\rho}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial\rho}\right)}{\sigma_t^4}\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^3}{\partial z\partial\rho^2} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{y_{t-1}^2\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^4}, \\
\frac{\frac{\partial^3\sigma_t^2}{\partial\alpha^3}}{\sigma_t^2} &= 3\frac{\sum_{j=1}^{T}(j-1)(j-2)\beta^{j-3}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^3\sigma_t^2}{\partial\beta^3}}{\sigma_t^2} &= 3\sum_{j=1}^{t}(j-1)(j-2)\beta^{j-3}\frac{\sigma_{t-j}^2}{\sigma_t^2},
\end{aligned}$$

for $\forall z, z_1, z_2 = \alpha, \beta$ where

$$
\begin{aligned}
\frac{\partial^3 \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1^2 \partial z_2} \quad = \quad & -\rho y_{t-1}\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)[\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2} + \frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\sigma_{t-1}^2 - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2} - \sigma_t^2\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t} \\
& -\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(3\sigma_{t-1}\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_t^{-1}\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-1}^6\sigma_t^2} \\
& -\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} + \frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t} \\
& +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(2\sigma_{t-1}^2\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \frac{1}{2}\sigma_t^{-1}\sigma_{t-1}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^8\sigma_t^2} - \frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^2\sigma_t^3} + \frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^{-1}} \\
& +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_t\sigma_{t-1}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t^6} \\
& -\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2\left(3\sigma_{t-1}^4\sigma_t^{-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2} - \sigma_t^{-3}\sigma_{t-1}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^{-2}}] \\
& -\frac{\rho^2 y_{t-1}^2}{2}[\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^4\sigma_t} + \frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-1}^2\sigma_t^3} - \frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2}{\sigma_{t-1}^6\sigma_t^{-1}}\right] \\
& -\frac{2\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2} \\
& +\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\sigma_{t-1}^4\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^4} \\
& -\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}],
\end{aligned}
$$

$$\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z\partial\rho} = \left[\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2\right]\left[\frac{\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^3}-\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t\sigma_{t-1}}\right],$$

$$\begin{aligned}\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2\partial\rho} &= \left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right] \\ &\times\left[\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2-\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}\right] \\ &+\left[\frac{\left(\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t+\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^6\sigma_t^2}\right] \\ &\times\left[\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2-\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}\right] \\ &+2\rho y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2},\end{aligned}$$

again for $\forall z, z_1, z_2 = \alpha, \beta$.

Definition 1 Following Jensen and Rahbek (2004b), we introduce the following lower and upper bounds on each parameter in $\boldsymbol{\theta}_0$ as

$$w_L < w_0 < w_U; \quad \alpha_L < \alpha_0 < \alpha_U,$$

$$\beta_L < \beta_0 < \beta_U; \quad \gamma_L < \gamma_0 < \gamma_U; \quad \rho_L < \rho_0 < \rho_U,$$

and we define the neighborhood $N(\boldsymbol{\theta}_0)$ around $\boldsymbol{\theta}_0$ as

$$N(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta}\backslash w_L \leq w \leq w_U, \alpha_L \leq \alpha \leq \alpha_U, \beta_L \leq \beta \leq \beta_U, \gamma_L < \gamma < \gamma_U \text{ and } \rho_L \leq \rho \leq \rho_U\}. \tag{24}$$

Lemma 5 Under Assumption B, there exists a neighborhood $N(\boldsymbol{\theta}_0)$ given by (24) in Definition 1 for which

(a) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{\partial^3}{\partial\alpha^3}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{1t}$, (b) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{\partial^3}{\partial\beta^3}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{2t}$,

(c) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\rho^3}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{3t}$, (d) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\beta}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{4t}$,

(e) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\rho}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{5t}$, (f) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\beta^2\partial\rho}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{6t}$,

(g) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta^2}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{7t}$, (h) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\rho^2}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{8t}$,

(i) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\beta\partial\rho^2}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{9t}$, (j) $\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta\partial\rho}l_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^T w_{10t}$,

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

Appendix 2

Proof of Theorem 3 We begin again with the first order derivatives (Lemma 6). Later, we move to the second and third order derivatives (Lemmas 7 and 8). Brown (1971) provides the type of central limit theorem we need. Lemmas 1 and 2 also hold for the AR(p), and they are used in our proofs. The first, second and third order derivatives are given in Results 4-6. The proof of Lemmas 6-8 are in the technical appendix that is available upon request from the authors. ■

Result 4: First order derivatives The first order derivatives associated with (18) are given by ($\forall z = \alpha, \beta$)

$$\frac{\partial}{\partial z} l_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1 - \frac{(\sigma_t(\boldsymbol{\theta}) v_t)^2}{\sigma_t^2(\boldsymbol{\theta})}\right)\frac{\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})} + \frac{\frac{\partial\left(y_t - \sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})}\right),$$

with

$$\begin{aligned}
\frac{\partial}{\partial \alpha} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{1t}(\boldsymbol{\theta}), \\
\frac{\partial}{\partial \beta} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{2t}(\boldsymbol{\theta}),
\end{aligned}$$

and

$$\frac{\partial}{\partial \rho_i} l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} s_{(2+i)t}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \frac{(\sigma_t(\boldsymbol{\theta}) v_t)\,\sigma_{t-i}^{-1}(\boldsymbol{\theta})\, y_{t-i}}{\sigma_t(\boldsymbol{\theta})}; \forall i = 1, ..., p,$$

where

$$\begin{aligned}
\frac{\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \alpha}}{\sigma_t^2(\boldsymbol{\theta})} &= \frac{\sum_{j=1}^{t} \beta^{j-1} y_{t-j}^2}{\sigma_t^2(\boldsymbol{\theta})}, \\
\frac{\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \beta}}{\sigma_t^2(\boldsymbol{\theta})} &= \sum_{j=1}^{t} \beta^{j-1} \frac{\sigma_{t-j}^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})},
\end{aligned}$$

$$\frac{\partial\left(y_t - \sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\,\sigma_{t-i}^{-1}(\boldsymbol{\theta})\, y_{t-i}\right)^2}{\partial z} = -\left(\sigma_t(\boldsymbol{\theta}) v_t\right)\left[\sum_{i=1}^{p} \frac{\left(\sigma_{t-i}^2(\boldsymbol{\theta})\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial z} - \sigma_t^2(\boldsymbol{\theta})\frac{\partial \sigma_{t-i}^2(\boldsymbol{\theta})}{\partial z}\right)\rho_i y_{t-i}}{\sigma_{t-i}^3(\boldsymbol{\theta})\,\sigma_t(\boldsymbol{\theta})}\right],$$

Lemma 6 Let Assumption C hold and write the scores associated with (18), derived in Result 4, as $s_{lt} = s_{lt}(\boldsymbol{\theta}_0)$. Then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{1t} \overset{d}{\longrightarrow} \mathrm{N}\left(0, \frac{\zeta}{4\alpha_0^2}\right),$$
$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{2t} \overset{d}{\longrightarrow} \mathrm{N}\left(0, \frac{\zeta(1+\mu_1)\mu_2}{4\beta_0^2(1-\mu_1)(1-\mu_2)}\right),$$
$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{(2+i)t} \overset{d}{\longrightarrow} \mathrm{N}\left(0, \frac{1}{(1-\rho_{i0}^2)}\right),$$

for $\forall i = 1, ..., p$, with $\mu_j = \mathrm{E}\left(\beta_0/\left(\alpha_0\epsilon_t^2+\beta_0\right)\right)^j, \quad j = 1, 2$ as $T \longrightarrow \infty$.

Result 4: Second order derivatives The second order derivatives associated with (18) are given by

$$\begin{aligned}
\frac{\partial^2}{\partial z_1 \partial z_2} l_T(\boldsymbol{\theta}) &= \frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{2(\sigma_t v_t)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4} + \left(\frac{(\sigma_t v_t)^2}{\sigma_t^2}-1\right)\frac{\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^2} - \frac{\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1\partial z_2}}{\sigma_t^2}\right) \\
&\quad + \frac{1}{2}\sum_{t=1}^{T}\left(\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_2} + \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1}}{\sigma_t^4}\right), \\
\frac{\partial^2}{\partial \rho_i \partial \rho_j} l_T(\boldsymbol{\theta}) &= -\sum_{t=1}^{T}\sigma_{t-i}^{-1}y_{t-i}\sigma_{t-j}^{-1}y_{t-j}, \\
\frac{\partial^2}{\partial z \partial \rho_i} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{\frac{\partial\left[\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)\sigma_t\sigma_{t-i}^{-1}y_{t-i}\right]}{\partial z}}{\sigma_t^2} - \sum_{t=1}^{T}\frac{(\sigma_t v_t)\sigma_{t-i}^{-1}y_{t-i}\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^3},
\end{aligned}$$

for $\forall i,j=1,...,p$,and $\forall z,z_1,z_2=\alpha,\beta$, where

$$\begin{aligned}
\frac{\frac{\partial^2\sigma_t^2}{\partial\alpha^2}}{\sigma_t^2} &= 2\frac{\sum_{j=1}^{t}(j-1)\beta^{j-2}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^2\sigma_t^2}{\partial\beta^2}}{\sigma_t^2} &= 2\sum_{j=1}^{t}(j-1)\beta^{j-2}\frac{\sigma_{t-j}^2}{\sigma_t^2}, \\
\frac{\partial^2(\sigma_t v_t)^2}{\partial z_1\partial z_2} &= -(\sigma_t v_t)\left[(\rho_1 y_{t-1})\left(\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right)\right] \\
&\quad -...-(\sigma_t v_t)\left[(\rho_p y_{t-p})\left(\frac{\frac{\partial\sigma_{t-p}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-p}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}}{\sigma_{t-p}^3\sigma_t}\right)\right] \\
&\quad +(\sigma_t v_t)\left[(\rho_1 y_{t-1})\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t+\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^5\sigma_t^2}\right] \\
&\quad +...+(\sigma_t v_t)\left[(\rho_p y_{t-p})\frac{\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(3\frac{\partial\sigma_{t-p}^2}{\partial z_2}\sigma_t+\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-p}^5\sigma_t^2}\right] \\
&\quad +\rho_1^2 y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2} \\
&\quad +...+\rho_p^2 y_{t-p}^2\frac{\left[\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right]\left[\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right]}{2\sigma_{t-p}^6\sigma_t^2},
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial\left[\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)\sigma_t\sigma_{t-i}^{-1}y_{t-i}\right]}{\partial z} = -\frac{y_t y_{t-i}}{2}\left[\sigma_t\sigma_{t-i}^{-3}\frac{\partial\sigma_{t-i}^2}{\partial z}-\sigma_t^{-1}\sigma_{t-i}^{-1}\frac{\partial\sigma_t^2}{\partial z}\right] \\
&\quad -\rho_1 y_{t-1}y_{t-i}\left[\sigma_{t-1}^{-1}\sigma_{t-i}^{-1}\frac{\partial\sigma_t^2}{\partial z}-\frac{\sigma_t^2}{2}\left[\sigma_{t-1}^3\sigma_{t-i}^{-1}\frac{\partial\sigma_{t-1}^2}{\partial z}+\sigma_{t-1}^{-1}\sigma_{t-i}^{-3}\frac{\partial\sigma_{t-i}^2}{\partial z}\right]\right] \\
&\quad -...-\rho_p y_{t-p}y_{t-i}\left[\sigma_{t-p}^{-1}\sigma_{t-i}^{-1}\frac{\partial\sigma_t^2}{\partial z}-\frac{\sigma_t^2}{2}\left[\sigma_{t-p}^3\sigma_{t-i}^{-1}\frac{\partial\sigma_{t-p}^2}{\partial z}+\sigma_{t-p}^{-1}\sigma_{t-i}^{-3}\frac{\partial\sigma_{t-i}^2}{\partial z}\right]\right].
\end{aligned}$$

Lemma 7 Under Assumption C, with the expressions of the second order derivatives given by Result 5 and evaluated at $\boldsymbol{\theta}_0$

(a) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha^2}l_T\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \xrightarrow{p} \frac{1}{2\alpha_0^2} > 0,$

(b) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta^2}l_T\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \xrightarrow{p} \frac{(1+\mu_1)\mu_2}{2\beta_0^2(1-\mu_1)(1-\mu_2)} > 0,$

(c) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\beta}l_T\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \xrightarrow{p} \frac{\mu_1}{2\alpha_0\beta_0(1-\mu_1)},$

(d) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\rho_i}l_T\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \xrightarrow{p} 0, \forall i = 1, ..., p,$

(e) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\rho_i}l_T\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \xrightarrow{p} 0, \forall i = 1, ..., p,$

(f) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\rho_i^2}l_T\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \xrightarrow{p} \frac{1}{\left(1-\rho_{i0}^2\right)} > 0, \forall i = 1, ..., p,$

with $\mu_j = \mathrm{E}\left(\beta_0/\left(\alpha_0\epsilon_t^2+\beta_0\right)\right)^j, \quad j = 1, 2$ as $T \longrightarrow \infty$.

Result 6: Third order derivatives The third order derivatives associated with (18) are given by

$$\begin{aligned}
\frac{\partial^3}{\partial z_1^2 \partial z_2} l_T(\boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{(\sigma_t v_t)^2}{\sigma_t^2}\right)\frac{\frac{\partial^3 \sigma_t^2}{\partial z_1^2 \partial z_2}}{\sigma_t^2} - \sum_{t=1}^{T}\left(2\frac{(\sigma_t v_t)^2}{\sigma_t^2}-1\right)\frac{\left(\frac{\partial^2\sigma_t^2}{\partial z_1 \partial z_2}\frac{\partial \sigma_t^2}{\partial z_1}+\frac{1}{2}\frac{\partial^2 \sigma_t^2}{\partial z_1^2}\frac{\partial \sigma_t^2}{\partial z_2}\right)}{\sigma_t^4} \\
&- \sum_{t=1}^{T}\left(1-\frac{3(\sigma_t v_t)^2}{\sigma_t^2}\right)\frac{\left(\frac{\partial \sigma_t^2}{\partial z_1}\right)^2\frac{\partial \sigma_t^2}{\partial z_2}}{\sigma_t^6} - \frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_1^2 \partial z_2}}{\sigma_t^2} \\
&+ \sum_{t=1}^{T}\left(\frac{\left(\frac{1}{2}\frac{\partial^2\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_1^2}\frac{\partial \sigma_t^2}{\partial z_2} + \frac{\partial^2\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_1 \partial z_2}\frac{\partial \sigma_t^2}{\partial z_1}\right)}{\sigma_t^4}\right) \\
&+ \sum_{t=1}^{T}\left(\frac{\left(\frac{1}{2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_2} + \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2}\frac{\partial\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_1}\right)}{\sigma_t^4}\right) \\
&- \sum_{t=1}^{T}\frac{\left(\frac{\partial\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_2}\left(\frac{\partial \sigma_t^2}{\partial z_1}\right)^2 + 2\frac{\partial\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_2}\right)}{\sigma_t^6}, \\
\frac{\partial^3}{\partial z_1 \partial z_2 \partial \rho_i} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{2(\sigma_t v_t)\sigma_t y_{t-i}\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_2}}{\sigma_{t-i}\sigma_t^6} + \sum_{t=1}^{T}\frac{(\sigma_t v_t)\sigma_t y_{t-i}\frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2}}{\sigma_{t-i}\sigma_t^4} - \sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_1 \partial z_2 \partial \rho_i}}{2\sigma_t^2} \\
&+ \frac{1}{2}\sum_{t=1}^{T}\left(\frac{\left(\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial^2\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_2 \partial \rho_i} + \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial^2\left(y_t - \sum_{i=1}^p \rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z_1 \partial \rho_i}\right)}{\sigma_t^4}\right), \\
\frac{\partial^3}{\partial z \partial \rho_i \partial \rho_j} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{y_{t-i} y_{t-j}\left(\sigma_{t-i}\frac{\partial \sigma_{t-j}}{\partial z} + \sigma_{t-j}\frac{\partial \sigma_{t-i}}{\partial z}\right)}{\sigma_{t-i}^2 \sigma_{t-j}^2}, \\
\frac{\partial^3}{\partial \rho_i \partial \rho_j \partial \rho_k} l_T(\boldsymbol{\theta}) &= 0,
\end{aligned}$$

for $\forall z_1, z_2 = \alpha, \beta$ and $\forall i, j, k = 1, ..., p$ where

$$\begin{aligned}
\frac{\frac{\partial^3 \sigma_t^2}{\partial \alpha^3}}{\sigma_t^2} &= 3\frac{\sum_{j=1}^{t}(j-1)(j-2)\beta^{j-3} y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^3 \sigma_t^2}{\partial \beta^3}}{\sigma_t^2} &= 3\sum_{j=1}^{t}(j-1)(j-2)\beta^{j-3}\frac{\sigma_{t-j}^2}{\sigma_t^2},
\end{aligned}$$

and

$$\frac{\partial^3\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t\left(\boldsymbol{\theta}\right)\sigma_{t-i}^{-1}\left(\boldsymbol{\theta}\right)y_{t-i}\right)^2}{\partial z_1^2\partial z_2}=-\rho_1 y_{t-1}\left(\sigma_t v_t\right)\left[\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}+\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\sigma_{t-1}^2-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t}\right.$$

$$-\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(3\sigma_{t-1}\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_t^{-1}\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-1}^6\sigma_t^2}-\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t}$$

$$+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(2\sigma_{t-1}^2\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\frac{1}{2}\sigma_t^{-1}\sigma_{t-1}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^8\sigma_t^2}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^2\sigma_t^3}+\frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^{-1}}$$

$$\left.+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_t\sigma_{t-1}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t^6}-\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2\left(3\sigma_{t-1}^4\sigma_t^{-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}-\sigma_t^{-3}\sigma_{t-1}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^{-2}}\right]$$

$$-\ldots-\rho_p y_{t-p}\left(\sigma_t v_t\right)\left[\frac{\left(\frac{\partial\sigma_{t-p}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}+\frac{\partial^3\sigma_{t-p}^2}{\partial z_1^2\partial z_2}\sigma_{t-p}^2-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^3\sigma_{t-p}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-p}^3\sigma_t}\right.$$

$$-\frac{\left(\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}\right)\left(3\sigma_{t-p}\sigma_t\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_t^{-1}\sigma_{t-p}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-p}^6\sigma_t^2}-\frac{\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^4\sigma_t}$$

$$+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(2\sigma_{t-p}^2\sigma_t\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\frac{1}{2}\sigma_t^{-1}\sigma_{t-p}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^8\sigma_t^2}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^2\sigma_t^3}+\frac{3\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^6\sigma_t^{-1}}$$

$$\left.+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_t\sigma_{t-p}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^4\sigma_t^6}-\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)^2\left(3\sigma_{t-p}^4\sigma_t^{-1}\frac{\partial\sigma_{t-p}^2}{\partial z_2}-\sigma_t^{-3}\sigma_{t-p}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^{12}\sigma_t^{-2}}\right]$$

$$-\frac{\rho_1^2y_{t-1}^2}{2}\left[\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^4\sigma_t}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-1}^2\sigma_t^3}-\frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2}{\sigma_{t-1}^6\sigma_t^{-1}}\right]\right.$$

$$-\frac{2\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}$$

$$\left.+\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\sigma_{t-1}^4\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^4}-\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}\right]$$

$$-\ldots-\frac{\rho_p^2y_{t-p}^2}{2}\left[\left[\frac{\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right)}{\sigma_{t-p}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)}{\sigma_{t-p}^4\sigma_t}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-p}^2\sigma_t^3}-\frac{3\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)^2}{\sigma_{t-p}^6\sigma_t^{-1}}\right]\right.$$

$$-\frac{2\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-p}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}$$

$$\left.+\frac{\left(\sigma_{t-p1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(3\sigma_{t-p}^4\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_{t-p}^{33}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^{12}\sigma_t^4}+\frac{\left(\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}\right)\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right)}{\sigma_{t-p}^6\sigma_t^2}\right],$$

for $\forall z_1, z_2 = \alpha, \beta$. Furthermore,

$$\frac{\partial^2\left(y_t - \sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\,\sigma_{t-i}^{-1}(\boldsymbol{\theta})\,y_{t-i}\right)^2}{\partial z\partial\rho} = \left[(\sigma_t v_t)\,y_{t-1} - \rho_1\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2\right]\left[\frac{\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^3} - \frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t\sigma_{t-1}}\right]$$
$$+ ... + \left[(\sigma_t v_t)\,y_{t-p} - \rho_p\sigma_t\sigma_{t-p}^{-1}y_{t-p}^2\right]\left[\frac{\sigma_t\frac{\partial\sigma_{t-p}^2}{\partial z}}{\sigma_{t-p}^3} - \frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t\sigma_{t-p}}\right],$$

$$\frac{\partial^3(\sigma_t v_t)^2}{\partial z_1\partial z_2\partial\rho} = \left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1} + \sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right]$$
$$\times\left[\rho_1\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2 - (\sigma_t v_t)\,y_{t-1}\right]$$
$$+ ... + \left[\frac{\frac{\partial\sigma_{t-p}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1} + \sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-p}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}}{\sigma_{t-p}^3\sigma_t}\right]$$
$$\times\left[\rho_p\sigma_t\sigma_{t-p}^{-1}y_{t-p}^2 - (\sigma_t v_t)\,y_{t-p}\right]$$
$$+\left[\frac{\left(\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1} - \sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t + \sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^6\sigma_t^2}\right]$$
$$\times\left[\rho_1\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2 - (\sigma_t v_t)\,y_{t-1}\right]$$
$$+ ... + \left[\frac{\left(\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1} - \sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-p}\frac{\partial\sigma_{t-p}^2}{\partial z_2}\sigma_t + \sigma_{t-p}^3\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-p}^6\sigma_t^2}\right]$$
$$\times\left[\rho_p\sigma_t\sigma_{t-p}^{-1}y_{t-p}^2 - (\sigma_t v_t)\,y_{t-p}\right]$$
$$+2\rho_1 y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2}$$
$$+ ... + 2\rho_p y_{t-p}^2\frac{\left[\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right]\left[\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right]}{2\sigma_{t-p}^6\sigma_t^2},$$

for $\forall z, z_1, z_2 = \alpha, \beta$.

Definition 2 Similar to Definition 1 we introduce lower and upper values for each parameter in $\boldsymbol{\theta}_0$ as

$$w_L < w_0 < w_U; \quad \alpha_L < \alpha_0 < \alpha_U; \quad \beta_L < \beta_0 < \beta_U,$$

$$\gamma_L < \gamma_0 < \gamma_U; \quad \rho_L < \rho_{10} < \rho_U; ...; \rho_L < \rho_{p0} < \rho_U,$$

and we define the neighborhood $N(\boldsymbol{\theta}_0)$ around $\boldsymbol{\theta}_0$ as

$$N(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \backslash w_L \le w \le w_U, \alpha_L \le \alpha \le \alpha_U, \beta_L \le \beta \le \beta_U, \gamma_L < \gamma < \gamma_U, \rho_L < \rho_1 < \rho_U, \rho_L \le \rho_p \le \rho_U\}, \tag{25}$$

Lemma 8 Under Assumption C, there exists a neighborhood $N(\boldsymbol{\theta}_0)$ given (25) in Definition 2 for which

(a)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{\partial^3}{\partial\alpha^3} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{1t},$	(b)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{\partial^3}{\partial\beta^3} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{2t},$
(c)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\rho_i\partial\rho_j\partial\rho_k} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{3t},$	(d)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\beta} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{4t},$
(e)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\rho_i} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{5t},$	(f)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\beta^2\partial\rho_i} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{6t},$
(g)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta^2} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{7t},$	(h)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\rho_i\partial\rho_j} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{8t},$
(i)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\beta\partial\partial\rho_i\partial\rho_j} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{9t},$	(j)	$\sup_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} \left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta\partial\rho_i} l_T(\boldsymbol{\theta})\right| \le \frac{1}{T}\sum_{t=1}^T w_{10t},$

with $i, j, k = 1, ...p$, where $w_{1t}, ..., w_{9t}$ and w_{10t} are stationary and have finite moments, $\mathrm{E}\, w_{lt} = M_l < \infty, \forall l = 1, ..., 10$. Furthermore $\frac{1}{T}\sum_{t=1}^T w_{lt} \xrightarrow{a.s.} M_l, \forall l = 1, ..., 10$.

Appendix 3

Proof of Theorem 5 The proof of Theorem 5 uses the same expressions of the first, second and third order derivatives of Theorem 3. Lemmas 6 and 7 apply directly to the stationary region by using the ergodic theorem for the observed information as in Jensen and Rahbek (2004a , page 641 and 2004b, Remark 4, page 1212). Lemma 8 does not use the nonstationary condition and therefore it applies directly both in the stationary and the nonstationary case (see Jensen and Rahbek (2004b, page 1216 and Remark 5 page 1218). Brown (1971) provides the type of central limit theorem we need. We also use Lemma 1 of Jensen and Rahbek (2004b) to prove uniqueness and the existence of the consistent and asymptotically Gaussian estimator.■

REFERENCES

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2003), Modelling and Forecasting Realized Volatility, *Econometrica* 71, 579–625.

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.

Baillie, R. T., T. Bollerslev and H. O. Mikkelsen (1996), Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 74, 3–30.

Baillie, R. T. and R. P. DeGennaro (1990), Stock Returns and Volatility, *Journal of Financial and Quantitative Analysis* 25, 2, 203-214.

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.

Bougerol P. and N. Picard (1992a), Strict Stationarity of Generalized Autoregressive Processes, *The Annals of Probability* 20, 4, 1714-1730.

Bougerol P. and N. Picard (1992b), Stationarity of GARCH Processes and of some nonnegative time series, *Journal of Econometrics* 52, 115-127.

Breidt, F. J., N. Crato and P. de Lima (1998), The Detection and Estimation of Long-Memory in Stochastic Volatility, *Journal of Econometrics* 83, 325–348.

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

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.

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

Christensen, B. J., C. M. Dahl and E. M. Iglesias (2008), A Class of Semiparametric GARCH in Mean Models and its Semiparametric Efficiency Bound, *CREATES Research Papers, number 2008-46.*

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.

Christensen, B. J., M. Ø. Nielsen and J. Zhu (2009), Long Memory in Stock Market Volatility and the Volatility-in-mean Effect: the FIEGARCH-M Model, *CREATES Research Papers, number 2009-1207.*

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.*

Crato, N. and de Lima, P. (1994), Long-Range Dependence in the Conditional Variance of Stock Returns, *Economics Letters* 45, 281–285.

Dahl, C. M., and M. Levine (2006), Nonparametric estimation of volatility models with serially dependent innovations, *Statistics and Probability Letters* 76, 2007-2016.

Ding, Z. and C. W. J. Granger (1996), Modelling Volatility Persistence of Speculative Returns: A new approach, *Journal of Econometrics* 73, 185–215.

Drost, F. C. and T. E. Nijman (1993), Temporal Aggregation of GARCH Processes, *Econometrica* 61, 909-927.

Dufour, J. M., L. Khalaf, J.-T.- Bernard and I. Genest (2004), Simulation-Based Finite-Sample Tests for Heteroscedasticity and ARCH Effects, *Journal of Econometrics*, 122, 2, 317-347.

Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987-1007.

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, 2, 391-407.

Franqc, C. and J.-M. Zakoïan (2000), Estimating Weak-GARCH Representations, *Econometric Theory* 16, 692-728.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics* 19, 3-29.

Galbraith, R. F. and J. I. Galbraith (1974), On the Inverses of Some Patterned Matrices Arising in the Theory of Stationary Time Series, *Journal of Applied Probability* 11, 63-71.

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.

Gonçalves, S. and H. White (2004), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics* 119, 199-219.

Hamilton, J. D. (1994), *Time Series Analysis.* Princeton University Press, Princeton.

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.

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

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

Kristensen, D. and A. Rahbek (2005), Asymptotics of the QMLE for a Class of ARCH(q) Models, *Econometric Theory* 21, 946-961.

Lee, S. and B. Hansen (1994), Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, *Econometric Theory* 10, 29-52.

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).

Ling, S. and McAleer, M. (2003), Asymptotic theory for a vector ARMA-GARCH model, *Econometric Theory* 19, 280-310.

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

Linton, O. and E. Mammen (2005), Estimating Semiparametric ARCH(∞) Models by Kernel Smoothing Methods, *Econometrica* 73, 3, 771-836.

Lumsdaine, R. L. (1996), Consistency and Asymptotic Normality of the Quasi Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, *Econometrica* 64, 3, 575-596.

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.

Lundbergh, S. and T. Teräsvirta (2002), Evaluating GARCH Models, *Journal of Econometrics* 110, 417-435.

Meddahi, N. and E. Renault (2004), Temporal Aggregation of Volatility Models, *Journal of Econometrics* 119, 355-379.

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

Nelson, D. B. (1990), Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory* 6, 318-334.

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

Robinson, P. M. (1991), Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regressions, *Journal of Econometrics* 47, 67–84.

Robinson, P. M. (2001), The Memory of Stochastic Volatility Models, *Journal of Econometrics* 101, 195–218.

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

White, H. (1984), *Asymptotic Theory for Econometricians.* New York: Academic Press.

Table 1: Size results of the LM_{χ^2} test statistic. Rejection frequencies reported

Nominal size:	50%	40%	30%	20%	15%	10%	5%	1%
GARCH(1,1)								
500	0.57090	0.45040	0.33420	0.21670	0.16750	0.11520	0.056800	0.012300
2000	0.56360	0.44360	0.33120	0.21500	0.16050	0.10930	0.055600	0.0092000
10000	0.54730	0.43310	0.31930	0.21580	0.16010	0.10770	0.054600	0.010100
AR(5)-GARCH(1,1)								
500	0.58380	0.46550	0.35270	0.23640	0.17880	0.12320	0.063700	0.012900
2000	0.58010	0.45590	0.33910	0.22680	0.17100	0.11330	0.059000	0.014300
10000	0.55780	0.44380	0.33450	0.22480	0.17430	0.11890	0.061600	0.012000

Simulations based on 10000 Monte Carlo replications.
The parameters values used in the GARCH(1,1) and AR(5)-GARCH(1,1) models are based on the IBM data.

Table 2: Estimation results based on the model $y_t = c + \sigma_t \epsilon_t, \sigma_t^2 = w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \epsilon_t \sim$iid$(0,1)$.

	IBM	Intel	Microsoft	GM	Ford	GE	Boing	Walmart	McDonalds	TimeWarner	Sony
ω	0.026**	0.086*	0.032	0.028**	0.034**	0.008***	0.023**	0.001	0.025***	0.004	0.105*
	(0.011)	(0.052)	(0.022)	(0.011)	(0.013)	(0.003)	(0.010)	(0.002)	(0.006)	(0.004)	(0.057)
α	0.068***	0.055***	0.051***	0.057***	0.048***	0.032***	0.032***	0.020***	0.047***	0.042***	0.073***
	(0.020)	(0.019)	(0.015)	(0.013)	(0.011)	(0.005)	(0.007)	(0.004)	(0.007)	(0.014)	(0.023)
β	0.925***	0.935***	0.945***	0.935***	0.945***	0.966***	0.964***	0.980***	0.946***	0.959***	0.903***
	(0.021)	(0.023)	(0.015)	(0.015)	(0.012)	(0.005)	(0.009)	(0.004)	(0.007)	(0.012)	(0.035)
c	0.022*	0.085***	0.080***	0.018	0.022	0.049***	0.053***	0.055***	0.060***	0.041	0.034
	(0.012)	(0.032)	(0.025)	(0.013)	(0.021)	(0.013)	(0.018)	(0.020)	(0.016)	(0.031)	(0.024)
LMTest	3.674	2.652	1.638	21.225	5.208	47.566	0.643	13.789	13.767	9.808	10.944
pval	0.055	0.103	0.201	0.000	0.022	0.000	0.423	0.000	0.000	0.002	0.001
logL	-0.869	-1.434	-1.285	-0.947	-1.170	-0.949	-1.210	-1.387	-1.024	-1.546	-1.140
BIC	1.741	2.874	2.575	1.898	2.344	1.902	2.423	2.778	2.053	3.100	2.286
First obs.	1.3.62	7.10.86	3.14.86	1.3.62	1.3.62	1.3.62	1.3.62	8.28.70	1.5.70	3.20.92	4.7.83

Standard errors in parenthesis. The last observation for all stocks are dated 9.10.2007. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. **LM test:** denotes the test statistics for neglected serial correlation. **pval:** is the p-value associated with the **LM test:**. **LogL:** denotes the value of the log likelihood function evaluated at the estimated parameters. **BIC:** is the Bayesian Information Criterion.

Table 3: Estimation results based on the model $y_t = c + \sum_{i=1}^{5} \delta_i y_{t-i} + \sigma_t \epsilon_t$, $\sigma_t^2 = w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$, $\epsilon_t \sim$iid$(0, 1)$.

	IBM	Intel	Microsoft	GM	Ford	GE	Boing	Walmart	McDonalds	TimeWarner	Sony
ω	0.026**	0.087*	0.032	0.028**	0.035***	0.008***	0.023**	0.002	0.025***	0.004	0.106*
	(0.011)	(0.053)	(0.022)	(0.011)	(0.013)	(0.003)	(0.010)	(0.002)	(0.006)	(0.004)	(0.055)
α	0.069***	0.056***	0.050***	0.058***	0.047***	0.031***	0.032***	0.019***	0.047***	0.043***	0.074***
	(0.020)	(0.020)	(0.015)	(0.013)	(0.011)	(0.005)	(0.007)	(0.004)	(0.007)	(0.014)	(0.023)
β	0.924***	0.934***	0.946***	0.934***	0.946***	0.967***	0.964***	0.981***	0.945***	0.958***	0.902***
	(0.021)	(0.023)	(0.015)	(0.015)	(0.012)	(0.005)	(0.009)	(0.004)	(0.008)	(0.013)	(0.034)
c	0.021*	0.088***	0.086***	0.018	0.023	0.055***	0.055***	0.064***	0.062***	0.039	0.033
	(0.012)	(0.033)	(0.026)	(0.013)	(0.021)	(0.014)	(0.018)	(0.020)	(0.016)	(0.031)	(0.023)
δ_1	0.023*	0.023	-0.019	0.053***	-0.027**	-0.069***	-0.007	-0.042***	0.043***	0.058***	0.046***
	(0.012)	(0.015)	(0.016)	(0.012)	(0.013)	(0.010)	(0.011)	(0.013)	(0.012)	(0.020)	(0.016)
δ_2	-0.009	-0.011	-0.024	-0.018*	-0.000	-0.037***	-0.015	-0.050***	-0.022**	-0.023	0.006
	(0.011)	(0.015)	(0.015)	(0.010)	(0.012)	(0.011)	(0.011)	(0.013)	(0.011)	(0.018)	(0.014)
δ_3	0.003	-0.016	-0.017	-0.014	-0.005	-0.017*	-0.004	-0.035***	-0.022**	0.007	0.003
	(0.010)	(0.016)	(0.016)	(0.010)	(0.012)	(0.010)	(0.010)	(0.012)	(0.011)	(0.018)	(0.014)
δ_4	-0.001	-0.012	-0.014	0.007	-0.003	-0.017*	-0.015	-0.009	-0.022**	0.022	-0.033**
	(0.011)	(0.015)	(0.015)	(0.010)	(0.012)	(0.010)	(0.010)	(0.012)	(0.011)	(0.018)	(0.014)
δ_5	0.011	-0.015	-0.002	-0.007	-0.012	-0.003	-0.006	-0.024**	-0.006	-0.019	0.006
	(0.010)	(0.015)	(0.015)	(0.010)	(0.012)	(0.010)	(0.010)	(0.012)	(0.011)	(0.018)	(0.014)
LMTest	2.002	0.136	0.372	11.345	0.613	13.984	3.641	6.171	7.475	9.357	0.652
pval	0.157	0.712	0.542	0.001	0.434	0.000	0.056	0.013	0.006	0.002	0.419
logL	-0.869	-1.433	-1.284	-0.946	-1.169	-0.946	-1.210	-1.385	-1.023	-1.544	-1.139
BIC	1.745	2.881	2.582	1.899	2.349	1.900	2.427	2.778	2.054	3.107	2.291
First obs.	1.3.62	7.10.86	3.14.86	1.3.62	1.3.62	1.3.62	1.3.62	8.28.70	1.5.70	3.20.92	4.7.83

Standard errors in parenthesis. The last observation for all stocks are dated 9.10.2007. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. **LM test:** denotes the test statistics for neglected serial correlation. **pval:** is the p-value associated with the **LM test:**. **LogL:** denotes the value of the log likelihood function evaluated at the estimated parameters. **BIC:** is the Bayesian Information Criterion.

Table 4: Estimation results based on the model $y_t = c + \sigma_t \epsilon_t$, $\sigma_t^2 = w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$, $\epsilon_t = \rho \epsilon_{t-1} + \upsilon_t$, υ_t ~iid(0, 1).

	IBM	Intel	Microsoft	GM	Ford	GE	Boing	Walmart	McDonalds	TimeWarner	Sony
ω	0.026**	0.085	0.032	0.028**	0.035***	0.008***	0.023**	0.002	0.025***	0.004	0.106*
	(0.011)	(0.052)	(0.022)	(0.011)	(0.013)	(0.003)	(0.010)	(0.002)	(0.006)	(0.004)	(0.056)
α	0.069***	0.055***	0.051***	0.058***	0.047***	0.031***	0.032***	0.020***	0.048***	0.044***	0.074***
	(0.020)	(0.019)	(0.015)	(0.013)	(0.011)	(0.005)	(0.007)	(0.004)	(0.007)	(0.015)	(0.023)
β	0.924***	0.935***	0.945***	0.934***	0.946***	0.967***	0.964***	0.980***	0.945***	0.957***	0.902***
	(0.021)	(0.023)	(0.015)	(0.015)	(0.012)	(0.005)	(0.009)	(0.004)	(0.007)	(0.013)	(0.034)
ρ	0.018*	0.022	-0.016	0.043***	-0.026**	-0.065***	-0.007	-0.039***	0.038***	0.049***	0.042***
	(0.010)	(0.014)	(0.014)	(0.011)	(0.012)	(0.010)	(0.011)	(0.013)	(0.011)	(0.018)	(0.014)
c	0.022*	0.085**	0.079***	0.018	0.022	0.048***	0.053***	0.055***	0.060***	0.040	0.034
	(0.013)	(0.033)	(0.025)	(0.014)	(0.020)	(0.013)	(0.018)	(0.019)	(0.017)	(0.033)	(0.024)
LMTest	0.087	0.006	0.260	0.027	0.078	0.066	0.002	0.399	0.030	0.777	0.004
pval	0.769	0.939	0.610	0.870	0.780	0.798	0.963	0.528	0.863	0.378	0.950
logL	-0.869	-1.434	-1.282	-0.946	-1.169	-0.947	-1.210	-1.386	-1.024	-1.544	-1.139
BIC	1.742	2.875	2.573	1.897	2.345	1.898	2.424	2.778	2.052	3.099	2.286
First obs.	1.3.62	7.10.86	3.14.86	1.3.62	1.3.62	1.3.62	1.3.62	8.28.70	1.5.70	3.20.92	4.7.83

Standard errors in parenthesis. The last observation for all stocks are dated 9.10.2007. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. **LM test:** denotes the test statistics for neglected serial correlation. **pval:** is the p-value associated with the **LM test:**. **LogL:** denotes the value of the log likelihood function evaluated at the estimated parameters. **BIC:** is the Bayesian Information Criterion.

Supplementary Technical Appendix for "Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary", by Christian M. Dahl and Emma M. Iglesias.

In this Appendix, we provide a detailed proof of Lemmas 1-8. Also we show the expressions of the first (given in Result 1), second (given in Result 2) and third order derivatives (given in Result 3) for the proof of Theorem 1, and of the first (given in Result 4), second (given in Result 5) and third (shown in Result 6) order derivatives that are needed for the proof of Theorem 3 in the paper: "Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary", by Christian M. Dahl and Emma M. Iglesias.

Proof of Lemma 1 Let Assumption **A** hold and write the process

$$\sigma_t^2 = w_0 + \left(\alpha_0 \epsilon_{t-1}^2 + \beta_0\right) \sigma_{t-1}^2 = B_t + A_t \sigma_{t-1}^2,$$

where $A_t = \left(\alpha_0 \epsilon_{t-1}^2 + \beta_0\right)$ and $B_t = w_0$. Then, applying Theorem 1.1 of Bougerol and Picard (1992a, page 1715) with $||$ being an operator norm, we verify the conditions that $\mathrm{E}\left(\log\left(\max\left\{|1, A_0|\right\}\right)\right) < \infty$, $\mathrm{E}\left(\log\left(\max\left\{|1, B_0|\right\}\right)\right) < \infty$ (since by Assumption **A** $w_0, \alpha_0, \beta_0 > 0$), and also σ_t^2 is strictly stationary if the Lyapunov exponent τ

$$\tau = \inf\left\{\mathrm{E}\left(\frac{1}{T+1}\log|A_0 \cdots A_T|\right)\right\} < 0.$$

In the case of one-dimensional recurrence equations

$$\frac{1}{T+1}\mathrm{E}\left(\log|A_0 \cdots A_T|\right) = \frac{1}{T+1}\sum_{i=0}^{T}\mathrm{E}\log|A_i| = \mathrm{E}\log|A_0| < 0.$$

Therefore, σ_t^2 is strictly stationary if

$$\mathrm{E}\log|A_0| = \mathrm{E}\log\left(\alpha_0\epsilon_t^2 + \beta_0\right) < 0.$$

This provides the conditions under which σ_t^2 is strictly stationary. Since the pair $\left(y_t, \sigma_t^2\right)' = \left(\sigma_t\epsilon_t, \sigma_t^2\right)'$ is a fixed function of $\left(\sigma_t^2, \epsilon_t\right)'$ which is ergodic and strictly stationary, then it follows that if $\left(\sigma_t^2, \epsilon_t\right)'$ is strictly stationary, then $\left(\sigma_t\epsilon_t, \sigma_t^2\right)'$ is also strictly stationary.

We note that this is the same sufficient and necessary condition when we have and i.i.d. process in ϵ_t. Since we have shown that the sufficient condition is the same regardless if ϵ_t is i.i.d. or ergodic and strictly stationary, then we also prove that this is not only a sufficient but also a necessary condition since this is also a necessary condition when ϵ_t is i.i.d. as proved in Bougerol and Picard (1992a). Therefore, we prove that the results carry over from the i.i.d. case to the ergodic and strictly stationary framework. This is not a trivial result, since almost all theory in the literature is developed only for strong GARCH processes.■

Proof of Lemma 2 Let Assumption A hold. By recursions,

$$\begin{aligned} \sigma_t^2 &= w_0 + \left(\alpha_0 \epsilon_{t-1}^2 + \beta_0\right) \sigma_{t-1}^2, \\ &= B_t + A_t \sigma_{t-1}^2, \\ &= A_t \cdots A_1 \sigma_0^2 + \sum_{i=0}^{t-1} A_t \cdots A_{t-i+1} B_{t-i}, \\ &= \sigma_{1t}^2 + \sigma_{2t}^2, \end{aligned} \tag{1}$$

where $\sigma_{1t}^2 = A_t \cdots A_1 \sigma_0^2$ and $\sigma_{2t}^2 = \sum_{i=0}^{t-1} A_t \cdots A_{t-i+1} B_{t-i}$. Since σ_{2t}^2 is always positive (and β_0 and w_0 are always positive by Assumption A), it suffices to show that $\frac{\log \sigma_{1t}^2}{t} \xrightarrow{a.s.} \mathrm{E} \log\left(\alpha_0 \epsilon_t^2 + \beta_0\right) \geq 0$. After taking logarithms and dividing by t in the expression for σ_{1t}^2 in (1) we obtain

$$\frac{\log \sigma_{1t}^2}{t} = \frac{\sum_{i=0}^{t-1} \log\left(\alpha_0 \epsilon_{t-i}^2 + \beta_0\right) + \log \sigma_0^2}{t},$$

and by the strong law of large numbers for ergodic and strictly stationary processes, when $\mathrm{E} \log A_t \geq 0$, and as $t \to \infty$, $\frac{\log \sigma_{1t}^2}{t} \xrightarrow{a.s.} \mathrm{E} \log\left(\alpha_0 \epsilon_t^2 + \beta_0\right) \geq 0$, since $\frac{\log \sigma_0^2}{t} \xrightarrow{a.s.} 0$. ■

Proof of Lemma 3 Define $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$.As in Jensen and Rahbek (2004b, Lemma 5 and its extension to the α parameter), using the law of iterated expectations and the properties of v_t, we have $\mathrm{E}\left(s_{1t}|I_{t-1}\right) = \mathrm{E}\left(s_{2t}|I_{t-1}\right) = \mathrm{E}\left(s_{3t}|I_{t-1}\right) = 0$. Also, the proof of Lemma 3 requires that

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

We prove now (2). For the first two scores, we have evaluated at $\boldsymbol{\theta}_0$ (note that the second term of the score $\frac{\partial\left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right)^2}{\partial z}$ is divided by σ_t^2)

$$\begin{aligned} &\frac{1}{2}\left(\left(1 - \frac{\left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right)^2}{\sigma_t^2}\right) \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} + \frac{\frac{\partial\left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right)^2}{\partial z}}{\sigma_t^2}\right) \\ &= \frac{1}{2}\left(1 - v_t^2\right) \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} + \frac{\rho_0 \sigma_t v_t y_{t-1}}{2 \sigma_t^3 \sigma_{t-1}^3}\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z}\right) \\ &= \frac{1}{2}\left(1 - v_t^2\right) \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} + \frac{\rho_0}{2} \frac{v_t y_{t-1}\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z}\right)}{\sigma_t^2 \sigma_{t-1}^3} = \\ &\frac{1}{2}\left(1 - v_t^2\right) \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} + \frac{\rho_0}{2} v_t \epsilon_{t-1}\left[\frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} - \frac{\frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right], \qquad \forall z = \alpha, \beta, \end{aligned}$$

where the first term follows from Lemma 5 in Jensen and Rahbek (2004b), and for the second term, note that

$$\left|\frac{\rho_0}{2} v_t \epsilon_{t-1}\left[\frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} - \frac{\frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right]\right| \leq \left|\frac{\rho_0}{2} v_t \epsilon_{t-1}\right|\left[\frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} - \frac{\frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right].$$

In addition

$$
\begin{aligned}
|v_t\epsilon_{t-1}| &= \left|\sum_{j=0}^{\infty}\rho_0^j v_t v_{t-1-j}\right| && (3)\\
&\leq \sum_{j=0}^{\infty}\left|\rho_0^j\right| |v_t v_{t-1-j}|, && (4)
\end{aligned}
$$

and from Holder's inequality

$$
\begin{aligned}
\mathrm{E}\,|v_t v_{t-1-j}| &\leq \sqrt{\mathrm{E}(v_t^2)}\sqrt{\mathrm{E}(v_{t-1-j}^2)} && (5)\\
&= \mathrm{E}(v_t^2) && (6)\\
&= 1. && (7)
\end{aligned}
$$

Finally, $\mathrm{E}\,|v_t\epsilon_{t-1}| < \infty$ (as $\sum_{j=0}^{\infty}\left|\rho_0^j\right| < \infty$), and hence $\mathrm{E}\,|s_{1t}| < \infty$ and $\mathrm{E}\,|s_{2t}| < \infty$. For the third score, we have

$$s_{3t} = v_t\epsilon_{t-1},$$

and from the previous results for the first and second score, it follows directly that

$$\mathrm{E}\,|s_{3t}| = \mathrm{E}\,|v_t\epsilon_{t-1}| < \infty.$$

Besides

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{1t}^2|I_{t-1}\right) &= \frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)^2\\
&+\frac{1}{T}\sum_{t=1}^{T}\frac{\rho_0^2 y_{t-1}^2\left(\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z}-\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}\right)^2}{4\left(1+\alpha_0 y_{t-2}^2\right)^3\left(1+\alpha_0 y_{t-1}^2\right)^2}\\
&\xrightarrow{p}\frac{\zeta}{4\alpha_0^2},
\end{aligned}
$$

using Lemmas 4-6 in Jensen and Rahbek (2004b) and its extension to the α parameter, since

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial\alpha}}{\sigma_t^2}\right)^2 \xrightarrow{p} \frac{\zeta}{4\alpha_0^2}, \tag{8}$$

and

$$\frac{\rho_0^2}{4T}\sum_{t=1}^{T}\epsilon_{t-1}^2\left[\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right)\right]^2 \xrightarrow{p} 0. \tag{9}$$

We now prove (8) and (9). The proof of (8) requires (see Lemmas 3-6 in Jensen and Rahbek (2004b) to show that

$$E_{k=1}^{j}\left(\frac{\beta}{\alpha\epsilon_{t-k}^2+\beta}\right)^p$$

is exponentially decreasing in j for each $p \geq 1$, and it follows since ϵ_t is stationary and ergodic.

The proof of (9), for example for $z=\alpha$, follows from first using Slutsky's theorem and noting that $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\frac{\partial\sigma_t^2}{\partial\alpha}}{\sigma_t^2}\right)^2 \xrightarrow{p} \frac{1}{\alpha_0^2}$ and also that $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\frac{\partial\sigma_{t-1}^2}{\partial\alpha}}{\sigma_{t-1}^2}\right)^2 \xrightarrow{p} \frac{1}{\alpha_0^2}$. We also use the fact that ϵ_t is stationary and ergodic and therefore

$$\frac{1}{T}\sum_{t=1}^{T}\epsilon_{t-1}^2 \xrightarrow{p} \frac{1}{(1-\rho_0^2)}.$$

For the second score and the outer product

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{2t}^2|I_{t-1}\right) &\xrightarrow{p} \frac{\zeta\left(1+\mu_1\right)\mu_2}{4\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)},\\ \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{1t}s_{2t}|I_{t-1}\right) &\xrightarrow{p} \frac{\zeta\mu_1}{4\alpha_0\beta_0\left(1-\mu_1\right)},\end{aligned}$$

since

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\right)^2 \xrightarrow{p} \frac{\zeta\left(1+\mu_1\right)\mu_2}{4\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)},$$

following Jensen and Rahbek (2004b), Lemmas 3, 4 and 5. For the last score

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{3t}^2|I_{t-1}\right) &= \frac{1}{T}\sum_{t=1}^{T}\frac{y_{t-1}^2}{\left(1+\alpha_0 y_{t-2}^2\right)},\\ &\xrightarrow{p} \frac{1}{(1-\rho_0^2)},\end{aligned}$$

under Assumption B. Finally, we can derive a Lindeberg type condition as in Jensen and Rahbek (2004a), where we have

$$\begin{aligned}&\frac{1}{4}\left(\left(1-v_t^2\right)\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}+\rho_0\frac{v_t y_{t-1}(\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}-\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z})}{\sigma_t^2\sigma_{t-1}^3}\right)^2\\ =\ &\frac{1}{4}[\left(1-v_t^2\right)^2\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)^2+\rho_0^2\frac{v_t^2 y_{t-1}^2(\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}-\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z})^2}{\sigma_t^4\sigma_{t-1}^6}\\ &+2\rho_0 v_t\left(1-v_t^2\right)\frac{\frac{\partial\sigma_t^2}{\partial z}y_{t-1}(\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}-\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z})}{\sigma_t^4\sigma_{t-1}^3}],\\ =\ &\frac{1}{4}\left(1-v_t^2\right)^2\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)^2+\frac{1}{4}\rho_0^2\left(v_t\epsilon_{t-1}\right)^2\left(\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_t^2\sigma_{t-1}^4}-\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^4\sigma_{t-1}^2}\right)\\ &+\frac{1}{2}\rho_0\left(v_t\epsilon_{t-1}\right)\left(1-v_t^2\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)\left(\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}-\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right),\end{aligned}$$

and

$$s_{3t}^2=v_t^2\epsilon_{t-1}^2,$$

with the following bounds for s_{1t}^2 (for s_{2t}^2, it would follow the same argument) and s_{3t}^2

$$\begin{aligned} s_{1t}^2 &\leq \mu_{1t}^2 \equiv \frac{1}{4\alpha_0^2}\left(1-v_t^2\right)^2+\rho_0^2\left(v_t\epsilon_{t-1}\right)^2+|\rho_0|\left|(v_t\epsilon_{t-1})\right|\left|\left(1-v_t^2\right)\right|, &(10)\\ s_{3t}^2 &\leq \mu_{3t}^2 \equiv v_t^2\epsilon_{t-1}^2(1+\gamma_0) \text{ for } 0\leq\gamma_0<\infty\ . &(11)\end{aligned}$$

Since v_t and ϵ_{t-1} are stationary and ergodic, then also any measurable mapping of v_t and ϵ_{t-1} will be stationary and ergodic, see, e.g., White (1984, Th. 3.35). Consequently, μ_{1t}^2 and μ_{3t}^2 are stationary and ergodic and it follows that

$$\begin{aligned} \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{it}^2 I\left(|s_{it}|>\sqrt{T}\partial\right)\right) &\leq \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(\mu_{it}^2 I\left(|\mu_{it}|>\sqrt{T}\partial\right)\right)\\ &= \lim_{T\to\infty}\mathrm{E}\left(\mu_{i1}^2 I\left(|\mu_{i1}|>\sqrt{T}\partial\right)\right) \qquad (12)\\ &\to 0. \end{aligned}$$

for $i=1,2,3$. This establishes the Lindeberg type condition as in Jensen and Rahbek (2004a, 2004b). ■

Proof of Lemma 4 For expressions (a), (b) and (c)

$$\begin{aligned} -\frac{1}{2T}\sum_{t=1}^{T}&\left(\left(1-\frac{2\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4}+\left(\frac{\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}-1\right)\frac{\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^2}\right)\\ &\xrightarrow{p}\frac{1}{2\alpha_0^2},\text{ with } z_1=z_2=\alpha,\\ &\xrightarrow{p}\frac{(1+\mu_1)\,\mu_2}{2\beta_0^2\,(1-\mu_1)\,(1-\mu_2)},\text{ with } z_1=z_2=\ \beta,\\ &\xrightarrow{p}\frac{\mu_1}{2\alpha_0\beta_0\,(1-\mu_1)},\text{ with } z_1=\alpha \text{ and } z_2=\beta, \end{aligned}$$

because of Lemma 6 in Jensen and Rahbek (2004b), and its extension to α and the cross products of α and β. Also

$$-\frac{1}{2T}\sum_{t=1}^{T}\left(-\frac{\frac{\partial^2\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2}}{\sigma_t^2}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\right)}{\sigma_t^4}\right)\xrightarrow{p}0;\quad\forall z_1,z_2=\alpha,\beta,$$

by using again Lemmas 3 and 4 in Jensen and Rahbek (2004b) and the same results as in Lemma 3 for our score. An expression of the most complicated term in the previous expression is

$$\frac{\rho_0^2}{4T}\sum_{t=1}^{T}\frac{y_{t-1}^2}{\sigma_{t-1}^2}\left[\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^2\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}}{\sigma_{t-1}^4}\right]\xrightarrow{p}0;\quad\forall z_1,z_2=\alpha,\beta.$$

For expressions (d) and (e)

$$-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\frac{\partial\left[\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z}}{\sigma_t^2}-\frac{\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_{t-1}^{-1}y_{t-1}\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^3}\right)\xrightarrow{p}0;\quad\forall z=\alpha,\beta,$$

since

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\rho_0 y_{t-1}^2}{\sigma_{t-1}^2}\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right)-\frac{y_t y_{t-1}}{2\sigma_t\sigma_{t-1}}\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right)\xrightarrow{p}0;\quad \forall z=\alpha,\beta,$$

and by a simple application again of Lemmas 3 and 4 in Jensen and Rahbek (2004b). Finally, expression (f)

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\sigma_{t-1}^{-2}y_{t-1}^2 &= \frac{1}{T}\sum_{t=1}^{T}\epsilon_{t-1}^2\\ &\xrightarrow{p}\frac{1}{\left(1-\rho_0^2\right)},\end{aligned}$$

by Assumption B.■

Proof of Lemma 5 In this proof we change the notation slightly. We first define σ_t^2 when the conditional variance is evaluated at $\boldsymbol{\theta}$, and $\sigma_t^2(\boldsymbol{\theta}_0)$ when it is evaluated at the true parameter. By definition

$$\frac{\left(y_t-\rho\sigma_t(\boldsymbol{\theta}_0)\,\sigma_{t-1}^{-1}(\boldsymbol{\theta}_0)\,y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta}_0)}=v_t^2.$$

For expressions (a), (b), (d) and (g), then $w_{1t}(\boldsymbol{\theta}), w_{2t}(\boldsymbol{\theta}), w_{4t}(\boldsymbol{\theta})$ and $w_{7t}(\boldsymbol{\theta})$ are given by

$$\begin{aligned}&-\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^3\sigma_t^2}{\partial z_1^2\partial z_2}}{\sigma_t^2}-\sum_{t=1}^{T}\left(1-\frac{3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6}\\&-\sum_{t=1}^{T}\left(2\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}-1\right)\frac{\left(\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\frac{1}{2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^4},\end{aligned}$$

with z_1 and z_2 being the corresponding α and β that are needed. If in all the three previous ratios we replace $\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}$ by

$$\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2}\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t-\rho\sigma_t(\boldsymbol{\theta}_0)\,\sigma_{t-1}^{-1}(\boldsymbol{\theta}_0)\,y_{t-1}\right)^2}v_t^2,$$

then, using Lemmas 3 and 9 of Jensen and Rahbek (2004b) as in their Lemma 10, and knowing that

$$\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2}\le R_1<\infty;\quad \frac{\sigma_t^2}{\sigma_t^2(\boldsymbol{\theta}_0)}\le R_2<\infty\ ,$$

therefore

$$\begin{aligned}\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t-\rho_0\sigma_t(\boldsymbol{\theta}_0)\,\sigma_{t-1}^{-1}(\boldsymbol{\theta}_0)\,y_{t-1}\right)^2}v_t^2 &= \frac{\left(y_t^2+\rho^2\sigma_t^2\sigma_{t-1}^{-2}\sigma_{t-1}^2(\boldsymbol{\theta}_0)\,\epsilon_{t-1}^2-2\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}y_t\right)}{\sigma_t^2(\boldsymbol{\theta}_0)},\\ &\le \epsilon_t^2+\rho^2\frac{\sigma_t^2}{\sigma_t^2(\boldsymbol{\theta}_0)}\frac{\sigma_{t-1}^2(\boldsymbol{\theta}_0)}{\sigma_{t-1}^2}\epsilon_{t-1}^2+\frac{\sigma_t^2}{\sigma_t^2(\boldsymbol{\theta}_0)}\sqrt{\frac{\sigma_{t-1}^2(\boldsymbol{\theta}_0)}{\sigma_{t-1}^2}}2\left|\rho\right|\left|\epsilon_{t-1}\epsilon_t\right|,\\ &\le \epsilon_t^2+\rho^2R_1R_2\epsilon_{t-1}^2+2\left|\rho\right|R_2\sqrt{R_1}\left|\epsilon_{t-1}\epsilon_t\right|,\end{aligned}$$

and, for example, for the second of the previous ratios

$$\begin{aligned}
D_1 &\equiv \left(3\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2}\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t-\rho\sigma_t(\boldsymbol{\theta}_0)\sigma_{t-1}^{-1}(\boldsymbol{\theta}_0)y_{t-1}\right)^2}v_t^2-1\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6}, \\
&\leq \left(3\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\sigma_t^2}\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t-\rho\sigma_t(\boldsymbol{\theta}_0)\sigma_{t-1}^{-1}(\boldsymbol{\theta}_0)y_{t-1}\right)^2}v_t^2+1\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6}, \\
&\leq 3A\left(R_1\epsilon_t^2+\max\left(\rho_L^2,\rho_U^2\right)R_1^2R_2\epsilon_{t-1}^2+2\max\left(|\rho_L|,|\rho_U|\right)R_2R_1|\epsilon_{t-1}\epsilon_t|+1/3\right), \\
&\equiv w_{1t}^*,
\end{aligned}$$

where A is the lower bound that is obtained in Lemma 9 of Jensen and Rahbek (2004b). Finally $\mathrm{E}(w_{it}^*)<\infty$, $\forall i=1,2,4,7$ as $\mathrm{E}(\epsilon_t^2)<\infty$, $\mathrm{E}(\epsilon_{t-1}^2)<\infty$ and $\mathrm{E}(|\epsilon_{t-1}\epsilon_t|)<\infty$ provided that $|\rho_0|<1$ and $\mathrm{E}(v_t^2)<\infty$. For the remaining ratios (and using the results of Lemma 4 with the same type of ratios that have been already analyzed), it is enough to deal with the extra expression

$$\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1^2\partial z_2}.$$

The most complicated terms of the previous expression are the terms

$$\begin{aligned}
&-\frac{\rho^2y_{t-1}^2}{2\sigma_{t-1}^2}\Bigg[\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}\sigma_t^3}\right]\left[\frac{\left(\beta+\alpha\epsilon_{t-1}^2\right)\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^2\sigma_t\left(\sigma_t^2-1\right)}+\frac{\left(\beta+\alpha\epsilon_{t-1}^2\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\left(\sigma_t^2-1\right)\sigma_t^3}-\frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2}{\sigma_{t-1}^6\sigma_t^{-1}}\right] \\
&-\frac{2\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^4} \\
&+\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\sigma_{t-1}^4\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^6} \\
&-\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^4}\Bigg],
\end{aligned}$$

for $\forall z_1,z_2=\alpha,\beta$, where we have re-arranged the terms and used the fact that $\sigma_{t-1}^2=\frac{(\sigma_t^2-1)}{(\beta+\alpha\epsilon_{t-1}^2)}$. Applying Lemma 9 of Jensen and Rahbek (2004b) directly again, we get that the previous expectation is bounded. The proof for expression (c) is trivial. For the proof of expressions (e) (f) and (j), we need to consider the extra ratios

$$\frac{\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2\partial\rho}}{\sigma_t^2},\ \text{and}\ \frac{\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z\partial\rho}}{\sigma_t^2}.$$

For the first term, we use the law of iterated expectations and the fact that

$$\left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^2\sigma_t^2}\right]\left[\frac{\rho y_{t-1}^2}{\sigma_{t-1}^2}\right],$$

and

$$\left[\frac{\left(\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t+\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^5\sigma_t^3}\right]\left[\frac{\rho y_{t-1}^2}{\sigma_{t-1}^2}\right],$$

as well as

$$\frac{2\rho y_{t-1}^2}{\sigma_{t-1}^2}\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^4\sigma_t^4},$$

for $\forall z_1, z_2 = \alpha, \beta$ are all bounded. For the second term we need

$$\left[\frac{\rho y_{t-1}^2}{\sigma_{t-1}^2}\right]\left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}-\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right],$$

which is also bounded. For the proof of expressions (h) and (i), we have

$$w_{it}(\boldsymbol{\theta}) = \epsilon_{t-1}^2\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2};\ \ \forall z=\alpha,\beta;\ \ i=8,9.$$

and due to Assumption 1, $\mathrm{E}(w_{it}) < \infty$, $i = 8, 9$. ■

Proof of Lemma 6 The proof follows the same type of argument as Lemma 3 and the proof in Jensen and Rahbek (2004b). Note that Lemma 6 is the same as Lemma 3, but with $i = 1, ..., p$. Define again $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$. Using the law of iterated expectations and the properties of v_t, then $\mathrm{E}(s_{1t}|I_{t-1}) = \mathrm{E}(s_{2t}|I_{t-1}) = \mathrm{E}\left(s_{(2+i)t}|I_{t-1}\right) = 0$. Also, by the same argument as in Lemma 3

$$\mathrm{E}|s_{1t}|<\infty;\qquad \mathrm{E}|s_{2t}|<\infty;\qquad \mathrm{E}\left|s_{(2+i)t}\right|<\infty;\ \forall i=1,...,p.$$

In addition,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{1t}^2|I_{t-1}\right) &\xrightarrow{p} \frac{\zeta}{4\alpha_0^2},\\
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{2t}^2|I_{t-1}\right) &\xrightarrow{p} \frac{\zeta(1+\mu_1)\mu_2}{4\beta_0^2(1-\mu_1)(1-\mu_2)},\\
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{1t}s_{2t}|I_{t-1}\right) &\xrightarrow{p} \frac{\zeta\mu_1}{4\alpha_0\beta_0(1-\mu_1)},\\
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{(2+i)t}^2|I_{t-1}\right) &\xrightarrow{p} \frac{1}{(1-\rho_{i0}^2)}.
\end{aligned}$$

■

Proof of Lemma 7 The proof follows the same type of arguments as in Lemma 4. Note that Lemma 7 is the same as Lemma 4, but for $i = 1, ..., p$. All the results of the proof of Lemma 4 apply here directly. ■

Proof of Lemma 8 The proof follows the same argument as Lemma 5. Note that Lemma 8 is the same as Lemma 5, but instead for $i = 1, ..., p$. All the results of the proof of Lemma 5 apply here directly. ■

Result 1: First order derivatives The first order derivatives are given by

$$\frac{\partial}{\partial z} l_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1 - \frac{\left(y_t - \rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta})}\right)\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})} + \frac{\frac{\partial\left(y_t - \rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)^2}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})}\right); \forall z = \alpha, \beta,$$

with

$$\begin{aligned}
\frac{\partial}{\partial\alpha} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{1t}(\boldsymbol{\theta}), \\
\frac{\partial}{\partial\beta} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{2t}(\boldsymbol{\theta}), \\
\frac{\partial}{\partial\rho} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{3t}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\frac{\left(y_t - \rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}}{\sigma_t(\boldsymbol{\theta})},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\alpha}}{\sigma_t^2(\boldsymbol{\theta})} &= \frac{\sum_{j=1}^{t}\beta^{j-1}y_{t-j}^2}{\sigma_t^2(\boldsymbol{\theta})}, \\
\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\beta}}{\sigma_t^2(\boldsymbol{\theta})} &= \sum_{j=1}^{t}\beta^{j-1}\frac{\sigma_{t-j}^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})}, \\
\frac{\partial\left(y_t - \rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)^2}{\partial z} &= \frac{-\left(y_t - \rho\sigma_t(\boldsymbol{\theta})\sigma_{t-1}^{-1}(\boldsymbol{\theta})y_{t-1}\right)\rho y_{t-1}}{\left(w + \alpha y_{t-2}^2 + \beta\sigma_{t-2}^2(\boldsymbol{\theta})\right)^{3/2}\left(w + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta})\right)^{1/2}} \\
&\quad \times\left(\left(w + \alpha y_{t-2}^2 + \beta\sigma_{t-2}^2(\boldsymbol{\theta})\right)\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial z} - \left(w + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta})\right)\frac{\partial\sigma_{t-1}^2(\boldsymbol{\theta})}{\partial z}\right),
\end{aligned}$$

for $\forall z = \alpha, \beta$.

Result 2: Second order derivatives We have,

$$\begin{aligned}
\frac{\partial^2}{\partial z_1\partial z_2} l_T(\boldsymbol{\theta}) &= \frac{1}{2}\sum_{t=1}^{T}\left(\left(1 - \frac{2\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta})}\right)\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4(\boldsymbol{\theta})} + \left(\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2(\boldsymbol{\theta})} - 1\right)\frac{\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^2(\boldsymbol{\theta})}\right) \\
&\quad + \frac{1}{2}\sum_{t=1}^{T}\left(-\frac{\frac{\partial^2\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2}}{\sigma_t^2} + \frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2} + \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\right)}{\sigma_t^4}\right), \\
\frac{\partial^2}{\partial\rho^2} l_T(\boldsymbol{\theta}) &= -\sum_{t=1}^{T}\sigma_{t-1}^{-2}y_{t-1}^2, \\
\frac{\partial^2}{\partial z\partial\rho} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\left(\frac{\frac{\partial\left[\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z}}{\sigma_t^2} - \frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_{t-1}^{-1}y_{t-1}\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^3}\right),
\end{aligned}$$

where

$$
\begin{aligned}
\frac{\frac{\partial^2 \sigma_t^2}{\partial \alpha^2}}{\sigma_t^2} &= 2\frac{\sum_{j=1}^{t} (j-1)\beta^{j-2} y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^2 \sigma_t^2}{\partial \beta^2}}{\sigma_t^2} &= 2\sum_{j=1}^{t} (j-1)\beta^{j-2} \frac{\sigma_{t-j}^2}{\sigma_t^2}, \\
\frac{\partial\left[\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z} &= \frac{y_t y_{t-1}}{2\sigma_t\sigma_{t-1}}\left(\frac{\partial\sigma_t^2}{\partial z} - \frac{\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right) - \rho y_{t-1}^2\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_{t-1}^2} - \frac{\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^4}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2} &= \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)(-\rho y_{t-1})\left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1} + \sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right] \\
&+ \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)(\rho y_{t-1})\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t + \sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^5\sigma_t^2}\right] \\
&+\rho^2 y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2},
\end{aligned}
$$

for $\forall z, z_1, z_2 = \alpha, \beta$.

Result 3: Third order derivatives We have,

$$
\begin{aligned}
\frac{\partial^3}{\partial z_1^2 \partial z_2} l_T(\boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^3\sigma_t^2}{\partial z_1^2\partial z_2}}{\sigma_t^2} \\
&\quad -\sum_{t=1}^{T}\left(1-\frac{3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6} \\
&\quad +\sum_{t=1}^{T}\frac{\frac{1}{2}\left(\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}+\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2}\right)}{\sigma_t^4} \\
&\quad +\sum_{t=1}^{T}\frac{\left(\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\right)}{\sigma_t^4} \\
&\quad -\sum_{t=1}^{T}\left(2\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}-1\right)\frac{\left(\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\frac{1}{2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^4}-\frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1^2\partial z_2}}{\sigma_t^2} \\
&\quad -\sum_{t=1}^{T}\frac{\left(\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2}\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2+2\frac{\partial\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^6}, \\
\frac{\partial^3}{\partial\rho^3} l_T(\boldsymbol{\theta}) &= 0, \\
\frac{\partial^3}{\partial z_1\partial z_2\partial\rho} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6}+\sum_{t=1}^{T}\frac{\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^4} \\
&\quad +\frac{1}{2}\sum_{t=1}^{T}\left(-\frac{\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2\partial\rho}}{\sigma_t^2}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2\partial\rho}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial\rho}\right)}{\sigma_t^4}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^3}{\partial z\partial\rho^2} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{y_{t-1}^2\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^4}, \\
\frac{\frac{\partial^3\sigma_t^2}{\partial\alpha^3}}{\sigma_t^2} &= 3\frac{\sum_{j=1}^{T}(j-1)(j-2)\beta^{j-3}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^3\sigma_t^2}{\partial\beta^3}}{\sigma_t^2} &= 3\sum_{j=1}^{t}(j-1)(j-2)\beta^{j-3}\frac{\sigma_{t-j}^2}{\sigma_t^2},
\end{aligned}
$$

for $\forall z, z_1, z_2 = \alpha, \beta$ where

$$\begin{aligned}
\frac{\partial^3 \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1^2 \partial z_2} &= -\rho y_{t-1}\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)[\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2} + \frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\sigma_{t-1}^2 - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2} - \sigma_t^2\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t} \\
&\quad -\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(3\sigma_{t-1}\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_t^{-1}\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-1}^6\sigma_t^2} \\
&\quad -\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} + \frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t} \\
&\quad +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(2\sigma_{t-1}^2\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \frac{1}{2}\sigma_t^{-1}\sigma_{t-1}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^8\sigma_t^2} - \frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^2\sigma_t^3} + \frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^{-1}} \\
&\quad +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_t\sigma_{t-1}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t^6} \\
&\quad -\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2\left(3\sigma_{t-1}^4\sigma_t^{-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2} - \sigma_t^{-3}\sigma_{t-1}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^{-2}}] \\
&\quad -\frac{\rho^2 y_{t-1}^2}{2}[\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^4\sigma_t} + \frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-1}^2\sigma_t^3} - \frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2}{\sigma_{t-1}^6\sigma_t^{-1}}\right] \\
&\quad -\frac{2\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2} \\
&\quad +\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\sigma_{t-1}^4\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^4} \\
&\quad -\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z \partial\rho} &= \left[\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1} - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2\right]\left[\frac{\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^3} - \frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t\sigma_{t-1}}\right], \\
\frac{\partial^3 \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1 \partial z_2 \partial\rho} &= \left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1} + \sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2} - \frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right] \\
&\quad \times\left[\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2 - \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}\right] \\
&\quad +\left[\frac{\left(\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1} - \sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t + \sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^6\sigma_t^2}\right] \\
&\quad \times\left[\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2 - \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}\right] \\
&\quad +2\rho y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2},
\end{aligned}$$

again for $\forall z, z_1, z_2 = \alpha, \beta$.

Result 4: First order derivatives The first order derivatives are given by ($\forall z = \alpha, \beta$)

$$\frac{\partial}{\partial z} l_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{(\sigma_t(\boldsymbol{\theta})v_t)^2}{\sigma_t^2(\boldsymbol{\theta})}\right)\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})}+\frac{\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z}}{\sigma_t^2(\boldsymbol{\theta})}\right),$$

with

$$\begin{aligned}\frac{\partial}{\partial\alpha} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{1t}(\boldsymbol{\theta}),\\ \frac{\partial}{\partial\beta} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} s_{2t}(\boldsymbol{\theta}),\end{aligned}$$

and

$$\frac{\partial}{\partial\rho_i} l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} s_{(2+i)t}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\frac{(\sigma_t(\boldsymbol{\theta})v_t)\,\sigma_{t-i}^{-1}(\boldsymbol{\theta})\,y_{t-i}}{\sigma_t(\boldsymbol{\theta})}; \forall i = 1, ..., p,$$

where

$$\begin{aligned}\frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\alpha}}{\sigma_t^2(\boldsymbol{\theta})} &= \frac{\sum_{j=1}^{t}\beta^{j-1}y_{t-j}^2}{\sigma_t^2(\boldsymbol{\theta})},\\ \frac{\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\beta}}{\sigma_t^2(\boldsymbol{\theta})} &= \sum_{j=1}^{t}\beta^{j-1}\frac{\sigma_{t-j}^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})},\end{aligned}$$

$$\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z} = -(\sigma_t(\boldsymbol{\theta})v_t)\left[\sum_{i=1}^{p}\frac{\left(\sigma_{t-i}^2(\boldsymbol{\theta})\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial z}-\sigma_t^2(\boldsymbol{\theta})\frac{\partial\sigma_{t-i}^2(\boldsymbol{\theta})}{\partial z}\right)\rho_i y_{t-i}}{\sigma_{t-i}^3(\boldsymbol{\theta})\sigma_t(\boldsymbol{\theta})}\right].$$

Result 5: Second order derivatives The second order derivatives are given by

$$\begin{aligned}\frac{\partial^2}{\partial z_1\partial z_2} l_T(\boldsymbol{\theta}) &= \frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{2(\sigma_t v_t)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4}+\left(\frac{(\sigma_t v_t)^2}{\sigma_t^2}-1\right)\frac{\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^2}-\frac{\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1\partial z_2}}{\sigma_t^2}\right)\\ &\quad+\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1}}{\sigma_t^4}\right),\\ \frac{\partial^2}{\partial\rho_i\partial\rho_j} l_T(\boldsymbol{\theta}) &= -\sum_{t=1}^{T}\sigma_{t-i}^{-1}y_{t-i}\sigma_{t-j}^{-1}y_{t-j},\\ \frac{\partial^2}{\partial z\partial\rho_i} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{\frac{\partial\left[\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)\sigma_t\sigma_{t-i}^{-1}y_{t-i}\right]}{\partial z}}{\sigma_t^2}-\sum_{t=1}^{T}\frac{(\sigma_t v_t)\,\sigma_{t-i}^{-1}y_{t-i}\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^3},\end{aligned}$$

for $\forall i, j = 1, ..., p$,and $\forall z, z_1, z_2 = \alpha, \beta$, where

$$
\begin{aligned}
\frac{\frac{\partial^2 \sigma_t^2}{\partial \alpha^2}}{\sigma_t^2} &= 2\frac{\sum_{j=1}^{t}(j-1)\beta^{j-2}y_{t-j}^2}{\sigma_t^2},\\
\frac{\frac{\partial^2 \sigma_t^2}{\partial \beta^2}}{\sigma_t^2} &= 2\sum_{j=1}^{t}(j-1)\beta^{j-2}\frac{\sigma_{t-j}^2}{\sigma_t^2},\\
\frac{\partial^2(\sigma_t v_t)^2}{\partial z_1 \partial z_2} &= -(\sigma_t v_t)\left[(\rho_1 y_{t-1})\left(\frac{\frac{\partial \sigma_{t-1}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-1}^2\frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2}}{\sigma_{t-1}^3\sigma_t}\right)\right]\\
&\quad -...-(\sigma_t v_t)\left[(\rho_p y_{t-p})\left(\frac{\frac{\partial \sigma_{t-p}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-p}^2\frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-p}^2}{\partial z_1} - \sigma_t^2\frac{\partial^2 \sigma_{t-p}^2}{\partial z_1 \partial z_2}}{\sigma_{t-p}^3\sigma_t}\right)\right]\\
&\quad +(\sigma_t v_t)\left[(\rho_1 y_{t-1})\frac{\left(\sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial \sigma_{t-1}^2}{\partial z_1}\right)\left(3\frac{\partial \sigma_{t-1}^2}{\partial z_2}\sigma_t + \sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^5\sigma_t^2}\right]\\
&\quad +...+(\sigma_t v_t)\left[(\rho_p y_{t-p})\frac{\left(\sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial \sigma_{t-p}^2}{\partial z_1}\right)\left(3\frac{\partial \sigma_{t-p}^2}{\partial z_2}\sigma_t + \sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-p}^5\sigma_t^2}\right]\\
&\quad +\rho_1^2 y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial \sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial \sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2}\\
&\quad +...+\rho_p^2 y_{t-p}^2\frac{\left[\sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2\frac{\partial \sigma_{t-p}^2}{\partial z_1}\right]\left[\sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2\frac{\partial \sigma_{t-p}^2}{\partial z_2}\right]}{2\sigma_{t-p}^6\sigma_t^2},
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial\left[\left(y_t - \sum_{i=1}^{p}\rho_i\sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)\sigma_t\sigma_{t-i}^{-1}y_{t-i}\right]}{\partial z} &= -\frac{y_t y_{t-i}}{2}\left[\sigma_t\sigma_{t-i}^{-3}\frac{\partial \sigma_{t-i}^2}{\partial z} - \sigma_t^{-1}\sigma_{t-i}^{-1}\frac{\partial \sigma_t^2}{\partial z}\right]\\
&\quad -\rho_1 y_{t-1}y_{t-i}\left[\sigma_{t-1}^{-1}\sigma_{t-i}^{-1}\frac{\partial \sigma_t^2}{\partial z} - \frac{\sigma_t^2}{2}\left[\sigma_{t-1}^{3}\sigma_{t-i}^{-1}\frac{\partial \sigma_{t-1}^2}{\partial z} + \sigma_{t-1}^{-1}\sigma_{t-i}^{-3}\frac{\partial \sigma_{t-i}^2}{\partial z}\right]\right]\\
&\quad -...-\rho_p y_{t-p}y_{t-i}\left[\sigma_{t-p}^{-1}\sigma_{t-i}^{-1}\frac{\partial \sigma_t^2}{\partial z} - \frac{\sigma_t^2}{2}\left[\sigma_{t-p}^{3}\sigma_{t-i}^{-1}\frac{\partial \sigma_{t-p}^2}{\partial z} + \sigma_{t-p}^{-1}\sigma_{t-i}^{-3}\frac{\partial \sigma_{t-i}^2}{\partial z}\right]\right].
\end{aligned}
$$

Result 6: Third order derivatives The third order derivatives are given by

$$
\begin{aligned}
\frac{\partial^3}{\partial z_1^2 \partial z_2} l_T(\boldsymbol{\theta}) &= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{(\sigma_t v_t)^2}{\sigma_t^2}\right)\frac{\frac{\partial^3 \sigma_t^2}{\partial z_1^2 \partial z_2}}{\sigma_t^2} - \sum_{t=1}^{T}\left(2\frac{(\sigma_t v_t)^2}{\sigma_t^2}-1\right)\frac{\left(\frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2}\frac{\partial \sigma_t^2}{\partial z_1}+\frac{1}{2}\frac{\partial^2 \sigma_t^2}{\partial z_1^2}\frac{\partial \sigma_t^2}{\partial z_2}\right)}{\sigma_t^4} \\
&\quad -\sum_{t=1}^{T}\left(1-\frac{3(\sigma_t v_t)^2}{\sigma_t^2}\right)\frac{\left(\frac{\partial \sigma_t^2}{\partial z_1}\right)^2 \frac{\partial \sigma_t^2}{\partial z_2}}{\sigma_t^6} - \frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1^2 \partial z_2}}{\sigma_t^2} \\
&\quad +\sum_{t=1}^{T}\left(\frac{\left(\frac{1}{2}\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1^2}\frac{\partial \sigma_t^2}{\partial z_2}+\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1 \partial z_2}\frac{\partial \sigma_t^2}{\partial z_1}\right)}{\sigma_t^4}\right) \\
&\quad +\sum_{t=1}^{T}\left(\frac{\left(\frac{1}{2}\frac{\partial^2 \sigma_t^2}{\partial z_1^2}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_2}+\frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1}\right)}{\sigma_t^4}\right) \\
&\quad -\sum_{t=1}^{T}\frac{\left(\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_2}\left(\frac{\partial \sigma_t^2}{\partial z_1}\right)^2+2\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_2}\right)}{\sigma_t^6}, \\
\frac{\partial^3}{\partial z_1 \partial z_2 \partial \rho_i} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{2(\sigma_t v_t)\sigma_t y_{t-i}\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_2}}{\sigma_{t-i}\sigma_t^6}+\sum_{t=1}^{T}\frac{(\sigma_t v_t)\sigma_t y_{t-i}\frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2}}{\sigma_{t-i}\sigma_t^4}-\sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1 \partial z_2 \partial \rho_i}}{2\sigma_t^2} \\
&\quad +\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\left(\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_2 \partial \rho_i}+\frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i \sigma_t(\boldsymbol{\theta})\sigma_{t-i}^{-1}(\boldsymbol{\theta})y_{t-i}\right)^2}{\partial z_1 \partial \rho_i}\right)}{\sigma_t^4}\right), \\
\frac{\partial^3}{\partial z \partial \rho_i \partial \rho_j} l_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T}\frac{y_{t-i}y_{t-j}\left(\sigma_{t-i}\frac{\partial \sigma_{t-j}}{\partial z}+\sigma_{t-j}\frac{\partial \sigma_{t-i}}{\partial z}\right)}{\sigma_{t-i}^2\sigma_{t-j}^2}, \\
\frac{\partial^3}{\partial \rho_i \partial \rho_j \partial \rho_k} l_T(\boldsymbol{\theta}) &= 0,
\end{aligned}
$$

for $\forall z_1, z_2 = \alpha, \beta$ and $\forall i, j, k = 1, ..., p$ where

$$
\begin{aligned}
\frac{\frac{\partial^3 \sigma_t^2}{\partial \alpha^3}}{\sigma_t^2} &= 3\frac{\sum_{j=1}^{t}(j-1)(j-2)\beta^{j-3}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^3 \sigma_t^2}{\partial \beta^3}}{\sigma_t^2} &= 3\sum_{j=1}^{t}(j-1)(j-2)\beta^{j-3}\frac{\sigma_{t-j}^2}{\sigma_t^2},
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^3\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t\left(\boldsymbol{\theta}\right)\sigma_{t-i}^{-1}\left(\boldsymbol{\theta}\right)y_{t-i}\right)^2}{\partial z_1^2\partial z_2} = & -\rho_1 y_{t-1}\left(\sigma_t v_t\right)[\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}+\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\sigma_{t-1}^2-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t} \\
& -\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(3\sigma_{t-1}\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_t^{-1}\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-1}^6\sigma_t^2}-\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t} \\
& +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(2\sigma_{t-1}^2\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\frac{1}{2}\sigma_t^{-1}\sigma_{t-1}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^8\sigma_t^2}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^2\sigma_t^3}+\frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^{-1}} \\
& +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_t\sigma_{t-1}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t^6}-\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2\left(3\sigma_{t-1}^4\sigma_t^{-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}-\sigma_t^{-3}\sigma_{t-1}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^{-2}}] \\
& -\ldots-\rho_p y_{t-p}\left(\sigma_t v_t\right)[\frac{\left(\frac{\partial\sigma_{t-p}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}+\frac{\partial^3\sigma_{t-p}^2}{\partial z_1^2\partial z_2}\sigma_{t-p}^2-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^3\sigma_{t-p}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-p}^3\sigma_t} \\
& -\frac{\left(\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}\right)\left(3\sigma_{t-p}\sigma_t\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_t^{-1}\sigma_{t-p}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-p}^6\sigma_t^2}-\frac{\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^4\sigma_t} \\
& +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(2\sigma_{t-p}^2\sigma_t\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\frac{1}{2}\sigma_t^{-1}\sigma_{t-p}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^8\sigma_t^2}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^2\sigma_t^3}+\frac{3\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^6\sigma_t^{-1}} \\
& +\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_t\sigma_{t-p}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^4\sigma_t^6}-\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)^2\left(3\sigma_{t-p}^4\sigma_t^{-1}\frac{\partial\sigma_{t-p}^2}{\partial z_2}-\sigma_t^{-3}\sigma_{t-p}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^{12}\sigma_t^{-2}}] \\
& -\frac{\rho_1^2 y_{t-1}^2}{2}[\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^4\sigma_t}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-1}^2\sigma_t^3}-\frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2}{\sigma_{t-1}^6\sigma_t^{-1}}\right] \\
& -\frac{2\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2} \\
& +\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\sigma_{t-1}^4\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^4}-\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}] \\
& -\ldots-\frac{\rho_p^2 y_{t-p}^2}{2}[\left[\frac{\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right)}{\sigma_{t-p}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)}{\sigma_{t-p}^4\sigma_t}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-p}^2\sigma_t^3}-\frac{3\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)^2}{\sigma_{t-p}^6\sigma_t^{-1}}\right] \\
& -\frac{2\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-p}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2} \\
& +\frac{\left(\sigma_{t-p1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(3\sigma_{t-p}^4\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_{t-p}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^{12}\sigma_t^4}+\frac{\left(\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}\right)\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right)}{\sigma_{t-p}^6\sigma_t^2}],
\end{aligned}
$$

for $\forall z_1, z_2 = \alpha, \beta$. Furthermore,

$$\begin{aligned}
\frac{\partial^2 \left(y_t - \sum_{i=1}^{p} \rho_i \sigma_t(\boldsymbol{\theta}) \sigma_{t-i}^{-1}(\boldsymbol{\theta}) y_{t-i}\right)^2}{\partial z \partial \rho} &= \left[(\sigma_t v_t) y_{t-1} - \rho_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1}^2\right] \left[\frac{\sigma_t \frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^3} - \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t \sigma_{t-1}}\right] \\
&+ ... + \left[(\sigma_t v_t) y_{t-p} - \rho_p \sigma_t \sigma_{t-p}^{-1} y_{t-p}^2\right] \left[\frac{\sigma_t \frac{\partial \sigma_{t-p}^2}{\partial z}}{\sigma_{t-p}^3} - \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t \sigma_{t-p}}\right], \\
\frac{\partial^3 (\sigma_t v_t)^2}{\partial z_1 \partial z_2 \partial \rho} &= \left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2}}{\sigma_{t-1}^3 \sigma_t}\right] \\
&\times \left[\rho_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1}^2 - (\sigma_t v_t) y_{t-1}\right] \\
&+ ... + \left[\frac{\frac{\partial \sigma_{t-p}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-p}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-p}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-p}^2}{\partial z_1 \partial z_2}}{\sigma_{t-p}^3 \sigma_t}\right] \\
&\times \left[\rho_p \sigma_t \sigma_{t-p}^{-1} y_{t-p}^2 - (\sigma_t v_t) y_{t-p}\right] \\
&+ \left[\frac{\left(\sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-1} \frac{\partial \sigma_{t-1}^2}{\partial z_2} \sigma_t + \sigma_{t-1}^3 \frac{\partial \sigma_t^2}{\partial z_2} \sigma_t^{-1}\right)}{2\sigma_{t-1}^6 \sigma_t^2}\right] \\
&\times \left[\rho_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1}^2 - (\sigma_t v_t) y_{t-1}\right] \\
&+ ... + \left[\frac{\left(\sigma_t^2 \frac{\partial \sigma_{t-p}^2}{\partial z_1} - \sigma_{t-p}^2 \frac{\partial \sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-p} \frac{\partial \sigma_{t-p}^2}{\partial z_2} \sigma_t + \sigma_{t-p}^3 \frac{\partial \sigma_t^2}{\partial z_2} \sigma_t^{-1}\right)}{2\sigma_{t-p}^6 \sigma_t^2}\right] \\
&\times \left[\rho_p \sigma_t \sigma_{t-p}^{-1} y_{t-p}^2 - (\sigma_t v_t) y_{t-p}\right] \\
&+ 2\rho_1 y_{t-1}^2 \frac{\left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6 \sigma_t^2} \\
&+ ... + 2\rho_p y_{t-p}^2 \frac{\left[\sigma_{t-p}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-p}^2}{\partial z_1}\right]\left[\sigma_{t-p}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-p}^2}{\partial z_2}\right]}{2\sigma_{t-p}^6 \sigma_t^2},
\end{aligned}$$

for $\forall z, z_1, z_2 = \alpha, \beta$.

Research Papers 2009

CREATES
Center for Research in Econometric Analysis of Time Series