Theis Lange

Asymptotic Theory in Financial Time Series Models with Conditional Heteroscedasticity

Ph.D. Thesis
2008

Thesis Advisor:	**Professor Anders Rahbek**
	University of Copenhagen
Thesis Committee:	**Professor Thomas Mikosch**
	University of Copenhagen
	Professor H. Peter Boswijk
	Universiteit van Amsterdam
	Associate Professor Christian Dahl
	University of Aarhus

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

Preface

This thesis is written in partial fulfillment of the requirements for achieving the Ph.D. degree in mathematical statistics at the Department of Mathematical Sciences under the Faculty of Science at the University of Copenhagen. The work has been completed from May 2005 to May 2008 under the supervision of Professor Anders Rahbek, University of Copenhagen.

The overall topic of the present thesis is econometrics and especially the field of volatility modeling and non-linear cointegration. The work is almost exclusively theoretical, but both the minor included empirical studies as well as the potential applications are to financial data. The thesis is composed of four separate papers suitable for submission to journals on theoretical econometrics. Even though all four papers concern volatility modeling they are quite different in terms of both scope and choice of perspective. In many ways this mirrors the many inspiring, but different people I have met during the last three years. As a natural consequence of this dispersion of focus there are some notational discrepancies among the four papers. Each paper should therefore be read independently.

Financial support from the Danish Social Sciences Research Council grant no. 2114-04-0001, which have made this work possible is gratefully acknowledge. In addition I thank the Danish Ministry of Science, Technology, and Innovation for awarding me the 2007 EliteForsk travel grant. I have on several occasions visited the Center for Research in Econometric Analysis of Time Series (CREATES) at the University of Aarhus and I thank them for their hospitality and support.

I would like to take this opportunity to thank my supervisor Professor Anders Rahbek for generously sharing his deep knowledge of the field and for countless hours of inspiring and rewarding conversations. Furthermore, I owe much thanks to Professor Tim Bollerslev, Duke University for his hospitality during my visit to Duke University in the spring of 2007. I highly appreciate the support and interest shown to me by many of my colleagues and indeed everybody at the Department of Mathematical Sciences. I wish to mention in particular Anders Tolver Jensen, Søren Tolver Jensen, and Søren Johansen. Finally, I would like to thank friends and family and especially Tilde Hellsten, Lene Lange, and Kjeld Sørensen.

Theis Lange
Copenhagen, May 2008

Abstract

The present thesis deals with asymptotic analysis of financial time series models with conditional heteroscedasticity. It is well-established within financial econometrics that most financial time series data exhibit time varying conditional volatility, as well as other types of non-linearities. Reflecting this, all four essays of this thesis consider models allowing for time varying conditional volatility, or heteroscedasticity.

Each essay is described in detail below. In the first essay a novel estimation technique is suggested to deal with estimation of parameters in the case of heavy tails in the autoregressive (AR) model with autoregressive conditional heteroscedastic (ARCH) innovations. The second essay introduces a new and quite general non-linear multivariate error correction model with regime switching and discusses a theory for inference. In this model cointegration can be analyzed with multivariate ARCH innovations. In the third essay properties of the much applied heteroscedastic robust Wald test statistic is studied in the context of the AR-ARCH model with heavy tails. Finally, in the fourth essay, it is shown that the stylized fact that almost all financial time series exhibit integrated GARCH (IGARCH), can be explained by assuming that the true data generating mechanism is a continuous time stochastic volatility model.

Lange, Rahbek & Jensen (2007): **Estimation and Asymptotic Inference in the AR-ARCH Model**. This paper studies asymptotic properties of the quasi-maximum likelihood estimator (QMLE) and of a suggested modified version for the parameters in the AR-ARCH model.

The modified QMLE (MQMLE) is based on truncation of the likelihood function and is related to the recent so-called self-weighted QMLE in Ling (2007b). We

show that the MQMLE is asymptotically normal irrespectively of the existence of finite moments, as geometric ergodicity alone suffice. Moreover, our included simulations show that the MQMLE is remarkably well-behaved in small samples. On the other hand the ordinary QMLE, as is well-known, requires finite fourth order moments for asymptotic normality. But based on our considerations and simulations, we conjecture that in fact only geometric ergodicity and finite second order moments are needed for the QMLE to be asymptotically normal. Finally, geometric ergodicity for AR-ARCH processes is shown to hold under mild and classic conditions on the AR and ARCH processes.

Lange (2008*a*): **First and second order non-linear cointegration models**. This paper studies cointegration in non-linear error correction models characterized by discontinuous and regime-dependent error correction and variance specifications. In addition the models allow for ARCH type specifications of the variance. The regime process is assumed to depend on the lagged disequilibrium, as measured by the norm of linear stable or cointegrating relations. The main contributions of the paper are: *i*) conditions ensuring geometric ergodicity and finite second order moment of linear long run equilibrium relations and differenced observations, *ii*) a representation theorem similar to Granger's representations theorem and a functional central limit theorem for the common trends, *iii*) to establish that the usual reduced rank regression estimator of the cointegrating vector is consistent even in this highly extended model, and *iv*) asymptotic normality of the parameters for fixed cointegration vector and regime parameters. Finally, an application of the model to US term structure data illustrates the empirical relevance of the model.

Lange (2008*b*): **Limiting behavior of the heteroskedastic robust Wald-test when the underlying innovations have heavy tails**. This paper establishes that the usual OLS estimator of the autoregressive parameter in the first order AR-ARCH model has a non-standard limiting distribution with a non-standard rate of convergence if the innovations have non-finite fourth order moments. Furthermore, it is shown that the robust t- and Wald test statistics of White (1980) are still consistent and have the usual rate of convergence, but a non-standard limiting distribution when the innovations have non-finite fourth order moment. The critical values for the non-standard limiting distribution are found to be higher than the usual N(0,1) and χ^2_1 critical values, respectively, which implies that an acceptance of the hypothesis using the standard robust t- or Wald tests remains valid even in the fourth order moment condition is not met. However, the size of the test might be higher than the nominal size. Hence the analysis presented in this paper extends the usability of the robust t- and Wald tests of White (1980). Finally, a small empirical study illustrates the results.

Jensen & Lange (2008): **On IGARCH and convergence of the QMLE for misspecified GARCH models**. We address the IGARCH puzzle by which we understand the fact that a GARCH(1,1) model fitted by quasi maximum likelihood estimation to virtually any financial dataset exhibit the property that $\hat{\alpha} + \hat{\beta}$ is close to one. We prove that if data is generated by certain types of continuous time stochastic volatility models, but fitted to a GARCH(1,1) model one gets that $\hat{\alpha} + \hat{\beta}$ tends to one in probability as the sampling frequency is increased. Hence, the paper suggests that the IGARCH effect could be caused by misspecification. The result establishes that the stochastic sequence of QMLEs do indeed behave as the deterministic parameters considered in the literature on filtering based on misspecified ARCH models, see e.g. Nelson (1992). An included study of simulations and empirical high frequency data is found to be in very good accordance with the mathematical results.

Resume

Denne afhandling omhandler asymptotisk teori for finansielle tidsrække modeller med tidsvarierende betinget varians. Det er velkendt indenfor finansiel økonometri, at de fleste finansielle tidsrækker udviser tidsafhængig betinget varians og andre type af ikke-lineariteter. I lyset af dette, analyser alle fire artikler i denne afhandling modeller, der tillader sådanne.

Hver artikel er beskrevet mere uddybende i de følgende afsnit. I den første artikel introduceres en ny estimationsteknik, der kan håndtere parameterestimation under tungt halede innovationer i den autoregressive (AR) model med autoregressiv betinget heteroskedastisitet (ARCH). Den anden artikel foreslår en ny og ganske generel ikke-lineær multivariat fejlkorrektionsmodel og diskuterer desuden asymptotisk teori. I denne model kan kointegration analyseres med multivariate ARCH innovationer. Med udgangspunkt i AR-ARCH modellen studeres i den tredje artikel egenskaberne ved det meget anvendte heteroskedastisk robuste Wald test. Den fjerde artikel udspringer af det såkaldte *stylized fact*, at stort set alle finansielle tidsrækker udviser integreret GARCH (IGARCH) egenskaben. I artiklen demonstreres det, at diskrete stikprøver fra kontinuert tids stokastiske volatilitets modeller kan producere IGARCH effekten.

Lange, Rahbek & Jensen(2007): **Estimation and Asymptotic Inference in the AR-ARCH Model**. Denne artikel studerer de asymptotiske egenskaber ved quasi maksimum likelihood estimatoren (QMLE) og ved en foreslået modificeret version for parametrene i AR-ARCH modellen.

Den modificerede QMLE (MQMLE) er baseret på trunkering af likelihood funktionen og er relateret til den nyeligt foreslåede self-weighted QMLE i Ling (2007b). Artiklen etablerer at MQMLE'en er asymptotisk normalfordelt uanset om inno-

vationerne har endelige momenter af en bestemt orden, i det geometrisk ergodicitet alene er tilstrækkeligt. Det inkluderede simulationsstudie viser desuden at MQMLE'en har bemærkelsesværdige fine egenskaber for korte dataserier. Endelig udledes simple og klassiske betingelser på AR og ARCH paramenterne, der garanterer at processer genereret af modellen er geometrisk ergodiske.

Lange (2008*a*): **First and second order non-linear cointegration models**. Artiklen omhandler kointegration i ikke-lineære fejl-korrektions modeller med diskontinuær og regime afhængig fejl-korrektion samt variansspecifikation. Desuden tillader modellen ARCH specifikation af variansen. Regime processen antages at afhænge af tidligere observationer. Artiklens hovedbidrag er: *i*) betingelser der sikrer geometrisk ergidicitet og endeligt andet moment af lineære langtidsligevægtsrelationer og tilvækster, *ii*) en repræsentationssætning svarende til Granger's repræsentationssætning og en variant af Donsker's sætning for de delte stokastiske trends, *iii*) at etablere at den sædvanlige reducerede rank regressions estimater af kointegrationsvektoren er konsistent selv i denne udvidede model og *iv*) asymptotisk normalitet af parameterestimaterne for fast kointegrationsvektor og regimeparametre. Den empiriske relevans af resultaterne illustreres med en anvendelse på amerikanske rentedata.

Lange (2008*b*): **Limiting behavior of the heteroskedastic robust Wald-test when the underlying innovations have heavy tails**. Artiklen etablerer at den sædvanlige OLS estimater af den autoregressive parameter i første ordens AR-ARCH modellen har en ikke-standard grænsefordeling med en ikke-standard konvergensrate, når innovationerne ikke har endeligt fjerde moment. Desuden vises det, at de robuste t- and Wald teststørrelser (se White (1980)) er konsistente med standard konvergensrate, men ikke-standard grænsefordeling når innovationer ikke har endeligt fjerde moment. De kritiske værdier for den etablerede grænsefordeling er højere end de tilsvarende for $N(0, 1)$ og χ_1^2 fordelingerne, hvilket implicerer at en hypotese accepteret ved hjælp af et standard robust t- eller Wald test forbliver accepteret selv når innovationerne ikke har endeligt fjerde moment. Det skal dog bemærkes, at testets størrelse kan være højere end den nominelle størrelse. Resultaterne præsenteret i denne artikel udvider dermed anvendelsen af de robuste t- og Wald tests introduceret i White (1980). Et kort empirisk studie illustrer resultaterne.

Jensen & Lange (2008): **On IGARCH and convergence of the QMLE for misspecified GARCH models**. Vi adresserer IGARCH effekten, ved hvilken vi forstår det faktum, at en GARCH(1,1), model fittet ved hjælp af quasi maksimum likelihood estimation til så godt som ethvert finansielt datasæt besidder den egenskab at $\hat{\alpha} + \hat{\beta}$ er tæt på én. Vi beviser, at hvis data er genereret af bestemte typer af kontinuer tids stokastiske volatilitets modeller, men fittet til en GARCH(1,1) model vil $\hat{\alpha}+\hat{\beta}$ konvergere til én i sandsynlighed når datafrekvensen går mod uendelig. Dermed indikerer artiklen, at IGARCH effekten kan være forårsaget af misspecifikation. Resultatet etablerer også at følgen af stokastiske QMLE'ere opfører sig som de deterministiske parametre betragtet i litteraturen omhandlende filtrering baseret på misspecificerede ARCH modeller, se f.eks. Nelson (1992). Det inkluderede studie af simulationer og højfrekvent empirisk data er i imponerende god overensstemmelse med de matematiske resultater.

Content

Estimation and Asymptotic Inference in the AR-ARCH Model

By Theis Lange, Anders Rahbek, and Søren Tolver Jensen

Department of Mathematical Sciences, University of Copenhagen
E-mail: *lange@math.ku.dk, rahbek@math.ku.dk, and soerent@math.ku.dk*

Abstract: This paper studies asymptotic properties of the quasi-maximum likelihood estimator (QMLE) and of a suggested modified version for the parameters in the autoregressive (AR) model with autoregressive conditional heteroskedastic (ARCH) errors.
The modified QMLE (MQMLE) is based on truncation of the likelihood function and is related to the recent so-called self-weighted QMLE in Ling (2007b). We show that the MQMLE is asymptotically normal irrespectively of the existence of finite moments, as geometric ergodicity alone suffices. Moreover, our included simulations show that the MQMLE is remarkably well-behaved in small samples. On the other hand the ordinary QMLE, as is well-known, requires finite fourth order moments for asymptotic normality. But based on our considerations and simulations, we conjecture that in fact only geometric ergodicity and finite second order moments are needed for the QMLE to be asymptotically normal. Finally, geometric ergodicity for AR-ARCH processes is shown to hold under mild and classic conditions on the AR and ARCH processes.

Keywords: ARCH; Asymptotic theory; Geometric ergodicity; QMLE; Modified QMLE.

1 Introduction

This paper considers likelihood based inference in a general stable autoregressive model with autoregressive conditional heteroskedastic errors, the AR-ARCH model. The aim of the paper is to contribute towards relaxing the moment restrictions currently found in the literature, which are often not met in empirical findings as noted in Francq & Zakoïan (2004), Ling & McAleer (2003), Ling & Li (1998), and Weiss (1984). Common to all these is the need for law of large number type theorems, which in turn induces the need for moment restrictions. In the pure ARCH model (no conditional mean part) Jensen & Rahbek (2004*b*) show how the parameter region in which the QMLE is asymptotically normal at the usual root T rate, can be expanded to include even non-stationary explosive processes. Adapting these techniques to the AR-ARCH model leads to the study of estimators based on two objective functions. One is the likelihood function and one is a censored version of the likelihood function based on censoring extreme terms of the log-likelihood function. The first estimator is the well known quasi-maximum likelihood function (QMLE) while the second is new and is denoted the modified quasi-maximum likelihood function (MQMLE).

Recently, and independently of our work, Ling (2007*b*) provides a theoretical study of a closely related general estimator denoted self-weighted QMLE. When computing the self-weighted QMLE the individual terms of the log-likelihood function are weighted such that the impact of the largest terms is decreased. The weights suggested in Ling (2007*b*) are fairly complex functions of the previous observations. This contrast our censoring scheme, which can be viewed as weighting with zero-one weights. Apart from the simplicity of our censoring scheme the main differences are: (*i*) We do not assume that the process has been initiated in the stationary distribution, but instead allow for any initial distribution. (*ii*) We have included a simulation study, which can provide specific advice on selecting the proper censoring. The included simulation study shows that the new estimator performs remarkably well and in many respects is superior to the ordinary QMLE. Finally our paper also establishes mild conditions for geometric ergodicity of AR-ARCH processes.

The presence of ARCH type effects in financial and macro economic time series is a well established fact. The combination of the ARCH specification for the conditional variance and the AR specification for the conditional mean has many appealing features, including a better specification of the forecast variance and the possibility of testing the presence of momentum in stock returns in a well specified model. Recently the AR-ARCH type models have been used as the basic "building blocks" for Markov switching and mixture models as in e.g. Lanne & Saikkonen (2003) and Wong & Li (2001).

The linear ARCH model model was originally introduced by Engle (1982) and asymptotic inference for this and other ARCH models have been studied in, e.g. Strumann & Mikosch (2006), Kristensen & Rahbek (2005), Medeiros & Veiga (2004) Berkes, Horváth & Kokoszka (2003), Lumsdaine (1996), Lee & Hansen (1994), and Weiss (1986). Common to these is as mentioned the assumption that the ARCH process is suitably ergodic or stationary. Recently

Jensen & Rahbek (2004b) have showed that the maximum likelihood estimator of the ARCH parameter is asymptotically normal with the same rate of convergence even in the non-stationary explosive case. However, as we demonstrate these results do not carry over to the AR-ARCH model due to the conditional mean part. Asymptotic inference in the AR-ARCH model is also considered in Francq & Zakoïan (2004), Ling & McAleer (2003), and Ling & Li (1998). The first of these establishes consistency of the QMLE assuming only existence of a stationary solution, but as noted the asymptotic normality results in all three references rely on the assumption of at least finite fourth order moment.

Note finally, that a related model is the so called DAR or double autoregressive model studied in Ling (2007a), Chan & Peng (2005), and Ling (2004), sometimes confusingly referred to as the AR-ARCH model. It differs from the AR-ARCH model, by not allowing the ARCH effect in the errors to vary independently of the level of the process. In particular, a unit root in the AR mean part of the DAR process, does not necessarily imply non-stationarity.

The paper proceeds as follows. In Section 2 the model and some important properties including geometric ergodicity of processes generated by the AR-ARCH model are discussed. Section 3 introduces the estimators and states the main results. In Section 4 we use Monte Carlo methods to compare the finite sample properties of the estimators and provide advice on how to estimate in practice. Finally Section 5 concludes. The Appendix contains all proofs.

2 Properties of the AR-ARCH model

In this section we present the model and discuss application of a law of large numbers to functions of the process, which is a critical tool in the asymptotic inference. The AR(r)-ARCH(p) model is given by

$$\begin{aligned} y_t &= \sum_{i=1}^{r} \rho_i y_{t-i} + \varepsilon_t(\theta) = \rho' \bar{y}_{t-1} + \varepsilon_t(\theta), \quad \rho = (\rho_1, ..., \rho_r) && (1) \\ \varepsilon_t(\theta) &= \sqrt{h_t(\theta)} z_t && (2) \\ h_t(\theta) &= \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2(\theta), \quad \alpha = (\alpha_1, ..., \alpha_p)' && (3) \end{aligned}$$

with $t = 1, ..., T$, $\bar{y}_t = (y_t, ..., y_{t-r+1})'$, and z_t an i.i.d.(0,1) sequence of random variables following a distribution P. Furthermore ρ and α denotes r and p dimensional vectors, respectively. For future reference define $\tilde{\varepsilon}_t(\theta) = (\varepsilon_t(\theta), ..., \varepsilon_{t-p+1}(\theta))'$. As to initial values estimation and inference is conditional on $(y_0, ..., y_{1-r-p})$, which is observed. The parameter vector is denoted $\theta = (\rho', \alpha', \omega)'$ and the true parameter θ_0 with α_0 and ω_0 strictly positive and the roots of the characteristic polynomial corresponding to (1) outside the unit circle. For notational ease we adopt the convention $\varepsilon_t(\theta_0) =: \varepsilon_t$ and $h_t(\theta_0) =: h_t$.

Corresponding to the model, all results regarding inference hold independently of the values of initial values. In particular, we do not assume that the initial values are initiated from an invariant distribution. Instead, similar to Kristensen & Rahbek (2005) where pure ARCH models are considered, we establish geometric ergodicity of the AR-ARCH process; see also Tjøstheim (1990) for a formal discussion of geometric ergodicity. Geometric ergodicity ensures that there exists an invariant distribution, but also as shown in Jensen & Rahbek (2007) that the law of large numbers apply to any measurable function of current and past values of the geometric ergodic process, independently of initial values, see (4) in Lemma 1 below. The application of the law of large numbers is a key part of the derivations in the next section when considering the behavior of the score and the information. The next lemma states sufficient (and mild) conditions for geometric ergodicity of the Markov chain $x_t = (\bar{y}_{t-1}, \tilde{\varepsilon}_t)'$ which appears in the score and information expressions. Note that the choice of $\bar{y}_{t-1}$ and $\tilde{\varepsilon}_t$ in the stacked process is not important, and for instance the same result holds with x_t defined as $(y_t, ..., y_{t-r-p+1})'$ instead. Initially two sets of assumptions corresponding to the general model and the first order model, respectively, are stated.

Assumption 1. *Assume that the roots of the characteristic polynomial corresponding to (1) evaluated at the true parameters are outside the unit circle and that*

$$\sum_{i=1}^{p} \alpha_{0,i} < 1.$$

Assumption 2. *Assume that* $r = p = 1$*, so* α *and* ρ *are scalars, and that*

$$E\left[\log(\alpha_0 z_t^2)\right] < 0 \text{ and } |\rho_0| < 1.$$

Lemma 1. *If either Assumption 1 or 2 holds and if z_t has a density f with respect to the Lebesgue measure on $\mathbb{R}$, which is bounded away from zero on compact sets then the process $x_t = (\bar{y}'_{t-1}, \tilde{\varepsilon}'_t)'$ generated by the AR-ARCH model, is geometrically ergodic. In particular there exists a stationary version and moreover if $E|g(x_t, ..., x_{t+k})| < \infty$ where expectation is taken with respect to the invariant distribution, the Law of Large Numbers given by*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} g(x_t, ..., x_{t+k}) \overset{a.s.}{=} E\left[g(x_t, ..., x_{t+k})\right], \tag{4}$$

holds irrespectively of the choice of initial distribution.

Note that the formulation of the lemma allows the application of the law of large numbers to summations involving functions of the Markov chain x_t even when the x_t has a non-finite expectation. The proof which utilizes the drift criterion can be found in the Appendix. Note that in recent years evermore general conditions for geometric ergodicity for generalized ARCH type processes have been derived, see e.g. Francq & Zakoïan (2006), Meitz & Saikkonen (2006), Kristensen (2005), Liebscher (2005), and the many references therein. Common to these is however, that they do not allow for an autoregressive mean part or belong to the category of DAR models. To the best of our knowledge the only results regarding geometric ergodicity of processes generated by the AR-ARCH model can be found in Cline & Pu (2004), Meitz & Saikkonen (2008), and Cline (2007), but their conditions are considerably more restrictive than the above since the very general setup employed does not utilize the exact specification of the simple AR-ARCH model.

With regards to the asymptotic theory the main contribution of Lemma 1 is to enable the use of the law of large numbers. Since the conditions of Assumption 1 imply the existence of finite second order moment, which it not needed for the first order model, it seems to be overly restrictive. We therefore state the following high order condition, which simply enables the use of the law of large numbers.

Assumption 3. *Assume that z_t has a density f with respect to the Lebesgue measure on $\mathbb{R}$, which is bounded away from zero on compact sets, that there* *exists an invariant distribution for the Markov chain* $x_t = (\bar{y}'_{t-1}, \tilde{\varepsilon}'_t)'$*, and that*

$$\frac{1}{T}\sum_{t=1}^{T} g(x_t, ..., x_{t+k}) \xrightarrow{P} E\left[g(x_t, ..., x_{t+k})\right] \quad as \quad T \to \infty,$$

for any measurable functions satisfying $E\left[g(x_t, ..., x_{t+k})\right] < \infty$.

This mild assumption is trivially satisfied if the drift criterion is used to establish stability of the chain. In the following we will discuss estimation and asymptotic theory under either of the three assumptions.

3 Estimation and Asymptotic Theory

In this section we study two estimators for the parameter θ in the AR-ARCH model. The first is the classical quasi maximum likelihood estimator (QMLE). Second, we propose a different estimator (the MQMLE) based on a modification of the Gaussian likelihood function which censors a few extreme observations. We show that both estimators are consistent and asymptotically normally distributed, and illustrate this by simulations. The proofs are based on verifying classical asymptotic conditions given in Lemma A.1 of the appendix. This involves asymptotic normality of the first derivative of the likelihood functions evaluated at the true values, convergence of the second order derivative evaluated at the true values and finally a uniform convergence result for the second order derivatives in a neighborhood around the true value, conditions $(A.1)$, $(A.2)$, and $(A.4)$, respectively. For both estimators we verify conditions $(A.1)$ and $(A.2)$ under the assumption of only second order moments of the ARCH process for the QMLE, and no moments (but under Assumption 3) for the MQMLE in Lemma 2. The uniform convergence is established for the MQMLE without any moment requirements and only the assumption of geometric ergodicity of the AR-ARCH process is therefore needed for this estimator to be asymptotically normal. The uniform convergence for the QMLE we can establish under the assumption of finite fourth order moment as in Ling & Li (1998). However, based on simulations, this assumption seems not essential at all and the result is conjectured to hold for the QMLE with only second order moments assumed to be finite.

Thus for the MQMLE consistency and normality holds independently of existence of any finite moments, only existence of a stationary invariant distribution is needed. In addition, the MQMLE have some nice finite sample properties as studied in the simulations. In particular, for the estimator of the autoregressive parameter ρ the finite sample distribution corresponding to the MQMLE approximates more rapidly the asymptotic normal one than the finite sample distribution of the QMLE of ρ. Furthermore the bias when estimating the ARCH parameter α is smaller when using the MQMLE than when using the classical QMLE. Of course since we are ignoring potentially useful information by censoring, the asymptotic variance for the MQMLE will be higher than for the QMLE.

We will consider the estimators based on minimizing the following functions

$$L_T^i(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t^i(\theta) \text{ where } l_t^i(\theta) = \gamma_t^i \left(\log h_t(\theta) + \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right),$$

for $i = 0, 1$ and with

$$\gamma_t^0 = 1, \quad \text{and} \quad \gamma_t^1 = 1_{\{|y_{t-1}|<M,\ldots,|y_{t-r-p}|<M\}} \tag{5}$$

for any positive constant M. The QMLE denoted $\hat{\theta}_T^0$ and the MQMLE denoted $\hat{\theta}_T^1$ will be the estimators based on minimizing L_T^0 and L_T^1, respectively.

The MQMLE estimator differs from the QMLE by introducing censoring. Clearly, the role of the censoring depends on the tail behavior of y_t. Davis & Mikosch (1998) show that under the assumptions of Lemma 1 the invariant distribution for ε_t is regularly varying with some index λ, and by Lange (2006) the invariant distribution for y_t is regularly varying with the same index. The interpretation of the tail index is, that the AR-ARCH process has finite moments of all orders below λ, but $E|y_t|^\lambda = \infty$ or, equivalently, that the density of the invariant distribution of y_t behaves like $|y_t|^{-\lambda-1}$ for $|y_t|$ large. Hence the probability of getting extreme observations is closely related to moment restrictions on the ARCH process. And since large observations provide the most precise estimates of the autoregressive parameter ρ, we have that if the probability of getting extreme observations becomes too large the QMLE has a non-standard (faster) rate of convergence. This is confirmed by the fact that when the second order moment of ε_t tends to infinity the asymptotic variance of the QMLE tends to zero (the exact expressions can be found in Conjecture 1). Unlike the QMLE the MQMLE censors away these extreme observations and is therefore asymptotically normal without any moment restrictions (see Theorem 1).

In practice, based on the simulations, we propose to use a censoring constant M which corresponds to censoring away at most 5% of the terms in the likelihood function (see Section 4 for further discussion). This choice is similar to the choice in the threshold- and change-point literature where for testing a priori certain quantiles of the observations are assumed to be in each of the regimes, see Hansen (1996, 1997). Note that if M is chosen in a data dependent fashion it may formally only depend on some finite number of observations. While this is crucial from a mathematical point of view, it is of no importance in practice.

The last part of this section contains the formal versions of our results.

Lemma 2. *Under either Assumption 1, 2, or 3 and the additional assumption that z_t has a a symmetric distribution with $E\left[(z_t^2-1)^2\right]=\kappa<\infty$ and density with respect to the Lebesgue measure, which is bounded on compact sets, the score and the observed information satisfy*

$$\begin{aligned}\sqrt{T}DL_T^1(\theta_0) &\overset{D}{\to} N(0,\Omega_S^1)\\ D^2L_T^1(\theta_0) &\overset{P}{\to} \Omega_I^1.\end{aligned}$$

If the true parameter θ_0 is such that in addition to the above the ARCH process has finite second order moment it holds that

$$\begin{aligned}\sqrt{T}DL_T^0(\theta_0) &\overset{D}{\to} N(0,\Omega_S^0)\\ D^2L_T^0(\theta_0) &\overset{P}{\to} \Omega_I^0.\end{aligned}$$

The matrices Ω_S^i and $\Omega_I^i>0$ are positive definite block diagonal and the exact expressions can be found in the appendix.

The notation defined in this lemma will be used throughout the rest of the paper. We can now state our main results regarding the MQMLE. Note that the proof can be found in the Appendix.

Theorem 1. *Under the assumptions of Lemma 2 regarding L_T^1 there exists a fixed open neighborhood $U=U(\theta_0)$ of θ_0 such that with probability tending to one as $T\to\infty$, $L_T^1(\theta)$ has a unique minimum point $\hat{\theta}_T^1$ in U. Furthermore $\hat{\theta}_T^1$ is* *consistent and asymptotically Gaussian,*

$$\sqrt{T}(\hat{\theta}_T^1 - \theta_0) \xrightarrow{D} N(0, (\Omega_I^1)^{-1}\Omega_S^1(\Omega_I^1)^{-1}).$$

If z_t is indeed Gaussian we have $\kappa = 2$ and therefore

$$(\Omega_I^i)^{-1}\Omega_S^i(\Omega_I^i)^{-1} = 2(\Omega_I^i)^{-1}$$

for $i = 0, 1$. Note that Theorem 1 is a local result in the sense that it only guarantees the existence of a small neighborhood around the true parameter value in which the function $L_T^1(\theta)$ has a unique minimum point, denoted $\hat{\theta}_T^1$, which is consistent and asymptotically Gaussian. In contrast to this Ling (2007*b*) establishes consistency and asymptotic normality over an arbitrary compact set. However, unlike Ling (2007*b*) we do not work with a compact parameter set during the estimation and hence our focus is on local behavior.

In the next section we provide numerical results, which indicate that the QMLE is asymptotically normal with an asymptotic variance given by Lemma A.1 and Lemma 2 as long as the ARCH process has finite second order moment. The required uniform convergence for the QMLE we can establish under the assumption of finite fourth order moment as in Ling & Li (1998). However, based on simulations, this assumption seems not essential at all and the result is conjectured to hold for the QMLE with only second order moments assumed to be finite. Hence we put forward the following conjecture.

Conjecture 1. *Under the assumptions of Lemma 2 regarding L_T^0 there exists a fixed open neighborhood $U = U(\theta_0)$ of θ_0 such that with probability tending to one as $T \to \infty$, the likelihood function $L_T^0(\theta)$ has a unique minimum point $\hat{\theta}_T^0$ in U. Furthermore $\hat{\theta}_T^0$ is consistent and asymptotically Gaussian,*

$$\sqrt{T}(\hat{\theta}_T^0 - \theta_0) \xrightarrow{D} N(0, (\Omega_I^0)^{-1}\Omega_S^0(\Omega_I^0)^{-1}).$$

It should be noted that consistency of the QMLE has been established in Francq & Zakoïan (2004) in which they also discuss (p. 613) whether the QMLE might indeed be asymptotically normal under the mild assumption of finite second order moment of the innovations. However, the result has still not been formally established.

4 Simulation Study

In this section we examine the finite sample properties of the two estimators by Monte Carlo simulation methods. Furthermore we provide advice on how to estimate AR-ARCH models in applications. We generate data from the DGP given by (1) - (3), with $r = p = 1$ and $z_t \sim$ i.i.d.$N(0,1)$, setting $\omega_0 = 1$ with no loss of generality[1]. The autoregressive parameter ρ_0 will be kept fixed at 0.5. Other values of this parameter were also considered, but these led to the same qualitative results as long as the absolute value of ρ_0 was not very close to unity. In the first part of this section we investigate the case where $\alpha_0 = 0.8$, corresponding to finite second order moment but non-finite fourth order moment of the ARCH process. With these parameter values the model does not meet the moment restrictions employed in the literature, but the model does satisfy the conditions of Conjecture 1 and Theorem 1. In the second part of this section we consider the case where $\alpha_0 = 1.5$, corresponding to non-finite second order moment of the ARCH process. With these parameter values the conditions of Conjecture 1 are not meet, but the conditions of Theorem 1 are. This part therefore serves as an illustration of the robustness of the MQMLE. Using the notation of the previous sections, we investigate the impacts of varying the sample size T, among $T = 250, 500$, 1,000, 4,000 and the truncation constant M, among $M = 2, 3, 5$.

Table 1 reports the bias of the estimators, sample standard deviation of $\sqrt{T}(\hat{\theta}_T^i - \theta_0)$ and in parentheses the deviation between the sample standard deviation and the true asymptotic standard deviation (from Conjecture 1 and Theorem 1 obtained by a different simulation study using 10^7 replications) in percent of the true asymptotic standard deviation. The table also reports skewness and excess kurtosis of the estimators normalized by their asymptotic standard deviation and finally the average truncation frequency. Note that $M = \infty$ corresponds to the QMLE.

Figure 1 reports QQ-plots of the two estimators ($\sqrt{T}(\theta_T^i - \theta_0)$) normalized by their respective true asymptotic variances (from Conjecture 1 and Theorem 1) against a standard normal distribution. The dotted lines correspond to (point-by-point) 95% confidence bands and are constructed using the empirical distribution

[1] All experiments were programmed using the random-number generator of the matrix programming language Ox 3.40 of Doornik (1998) over $N =$ 10,000 Monte Carlo replications.

functions. The normalization allows one to compare how close the finite sample distribution is to the asymptotic distribution directly between the MQMLE and the QMLE.

We will first consider the properties of the QMLE of the autoregressive parameter. Recall that known asymptotic results only guarantees consistency, see Francq & Zakoïan (2004), but not asymptotic normality since the ARCH process has non-finite fourth order moment when $\alpha_0 = 0.8$. However, both the QQ-plot and the numeric results of Table 1 indicate that the estimator based on L_T^0 (the maximum likelihood estimator) is asymptotically normal distributed with the claimed asymptotic variance. This is in good accordance with Lemma 2, which states that both the first- and second derivatives of L_T^0 evaluated at the true values have the right limits as long as the ARCH process has finite second order moment. This forms the motivation for Conjecture 1. The plots and tables also confirm that the QMLE of the ARCH parameters α and ω are asymptotically Gaussian.

Next we will compare the performance of the two estimators of the autoregressive parameter. From Table 1 it is noted that the observed standard deviation, skewness, and excess kurtosis of the normalized estimator $\hat{\rho}_T^1$ are consistently closer to their true asymptotic values than those of the maximum likelihood estimator. Furthermore from Figure 1 it is evident that the finite sample distribution of the MQMLE is "closer" to the claimed normal distribution than the finite sample distribution of the QMLE. Note that the left part of the confidence bands for the two estimators are non-overlapping, which indicates that the observed difference is statistically significant. This is true for all values of the truncation constant M, but is most evident when M is small. From Table 1 it it also clear that the asymptotic variance for $\hat{\rho}_T^1$ increases as the censoring constant is decreased, this is due to the fact that the censoring in effect ignores useful information. However, for $M = 5$, which in this case corresponds to ignoring around 5% of the terms of the likelihood function, the asymptotic standard deviation is only around 15% larger than that of the maximum likelihood estimator.

When comparing the estimators of the ARCH parameter α, the conclusions become less clear cut. Table 1 and Figure 1 indicate that unlike when estimating the autoregressive parameter, the traditional QMLE is the one that approaches its asymptotic distribution fastest (both when measured by the sample standard

	T = 250				T = 500			
	$M = 2$	$M = 3$	$M = 5$	$M = \infty$	$M = 2$	$M = 3$	$M = 5$	$M = \infty$
$\hat{\rho}$ Bias	0.001	0.000	-0.002	-0.003	0.000	-0.001	0.000	-0.001
$\hat{\alpha}$ Bias	0.001	-0.002	-0.004	-0.016	-0.005	-0.005	0.000	-0.007
$\hat{\omega}$ Bias	0.001	0.003	0.003	0.008	0.002	0.001	0.002	0.004
$\hat{\rho}$ Std.Dev. (%)	1.553(3.66)	1.099(2.55)	0.873(2.46)	0.754(7.3)	1.546(3.56)	1.082(0.91)	0.865(1.26)	0.737(5.39)
$\hat{\alpha}$ Std.Dev. (%)	4.641(9.82)	3.249(4.29)	2.678(2.94)	2.446(3.54)	4.423(4.66)	3.190(2.41)	2.643(1.64)	2.401(1.66)
$\hat{\omega}$ Std.Dev. (%)	3.060(5.95)	2.719(3.47)	2.573(2.82)	2.593(3.66)	2.966(2.86)	2.678(1.82)	2.543(1.62)	2.493(1.81)
$\hat{\rho}$ Skewness	0.077	-0.026	-0.133	-0.199	0.015	-0.010	-0.069	-0.159
$\hat{\alpha}$ Skewness	0.344	0.251	0.219	-0.004	0.206	0.203	0.131	-0.011
$\hat{\omega}$ Skewness	0.441	0.373	0.354	0.357	0.273	0.239	0.271	0.243
$\hat{\rho}$ Excess kurtosis	0.234	0.133	0.062	0.344	0.151	0.113	0.001	0.235
$\hat{\alpha}$ Excess kurtosis	0.432	0.153	0.191	0.042	0.094	0.030	0.142	0.006
$\hat{\omega}$ Excess kurtosis	0.481	0.295	0.170	0.216	0.158	0.192	0.202	0.170
Mean censoring freq.	0.364	0.182	0.066	0.000	0.363	0.180	0.062	0.000

	T = 1,000			
$\hat{\rho}$ Bias	0.000	0.001	0.000	0.000
$\hat{\alpha}$ Bias	-0.002	-0.003	0.000	-0.004
$\hat{\omega}$ Bias	0.000	0.003	0.000	0.002
$\hat{\rho}$ Std.Dev. (%)	1.499(0.25)	1.078(0.48)	0.862(0.99)	0.724(3.07)
$\hat{\alpha}$ Std.Dev. (%)	4.323(-2.28)	3.120(-0.12)	2.605(-0.23)	2.352(-0.39)
$\hat{\omega}$ Std.Dev. (%)	2.910(-0.91)	2.648(-0.78)	2.501(-0.13)	2.453(-0.18)
$\hat{\rho}$ Skewness	0.006	-0.018	-0.074	-0.084
$\hat{\alpha}$ Skewness	0.165	0.133	0.101	-0.005
$\hat{\omega}$ Skewness	0.237	0.179	0.191	0.169
$\hat{\rho}$ Excess kurtosis	0.025	0.044	0.014	0.165
$\hat{\alpha}$ Excess kurtosis	0.144	0.029	0.137	0.042
$\hat{\omega}$ Excess kurtosis	0.163	0.122	0.123	0.097
Mean censoring freq.	0.362	0.180	0.060	0.000

Table 1: Results of the simulation study with $\theta_0 = (0.5, 0.8, 1)'$ based on 10,000 Monte Carlo replications. Note that $M = \infty$ corresponds to the QMLE. Bias is defined as the sample average between $\hat{\theta}^i_T$ and θ_0. Std. Dev. is the sample standard deviation of $\sqrt{T}(\hat{\theta}^i_T - \theta_0)$ and in parentheses this is compared to the asymptotic standard deviation from Conjecture 1 and Theorem 1. Skewness and excess kurtosis are calculated from $\sqrt{T}(\hat{\theta}^i_T - \theta_0)$ normalized by their asymptotic standard deviation.

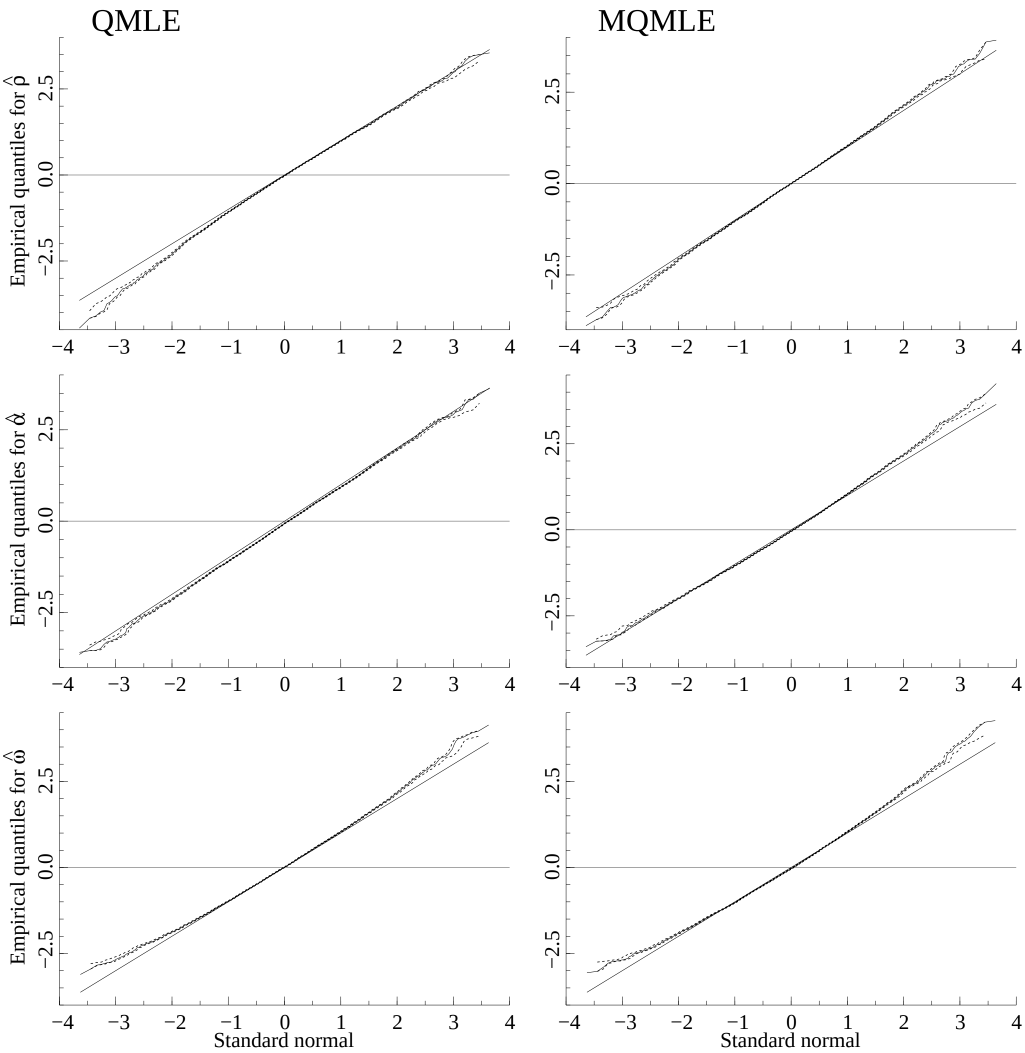

Figure 1: QQ-plots of $\sqrt{T}(\hat{\theta}_T^i - \theta_0)$ normalized by their asymptotic standard deviation (from Conjecture 1 and Theorem 1) against a standard normal distribution. The left column corresponds to the QMLE and the right to the MQMLE (with $M = 2$). The parameters are kept fixed at $\theta_0 = (0.5, 0.8, 1)'$ and $T = 500$ and the plot is based on 10,000 Monte Carlo replications. The dotted lines correspond to 95% confidence bands based on the empirical distribution function.

deviation, skewness, and excess kurtosis, and when inspected graphically). However, the MQMLE seems to have a lower bias than the QMLE.

Finally Table 1 and the bottom row of Figure 1 show that the estimation of the scale parameter ω is relatively unaffected by the choice of estimator and censoring constant.

Hence the choice of how to estimate in the AR-ARCH model depends on which parameters that are of most interest to the problem at hand. All in all we would suggest using the MQMLE, because it avoids the need for moment restrictions, and selecting the censoring constant such that around 5% of the observations are censored away, as this makes the price in the form of higher asymptotic standard deviation fairly small.

In the following we will consider the case where $\alpha_0 = 1.5$, which corresponds to non-finite second order moment of the ARCH process. It should be noted that in this case the asymptotic variance associated with $\hat{\rho}_T^0$ cannot be guaranteed to be finite, which makes the rescaling used in Figure 1 meaningless. Hence Figure 2 reports QQ-plots of the two estimators against a normal distribution with mean zero and the same variance as $\sqrt{T}(\theta_T^i - \theta_0)$. When varying the sample length T this approach allows one to see directly whether $\sqrt{T}$ is the right rate of convergence. The confidence intervals are constructed as in Figure 1.

From Figure 2 and Figure 3 the most striking feature is the bended shape of the curve corresponding to the QMLE estimator of the autoregressive parameter. The hypothesis that the QMLE has a non-standard rate of convergence is further strengthed by observing that the sample standard deviation decreases as the sample size increases. This is in good accordance with the fact that the asymptotic variance in Conjecture 1 is zero when the ARCH process has non-finite second order moment. Hence it does not seem reasonable to assume that the conditions of Conjecture 1 can be relaxed any further. It is also noted that the asymptotic normality of the QMLE estimators of the ARCH parameter α and the scale parameter ω seems to hold even though the ARCH process has non-finite second order moment. This is in accordance with Jensen & Rahbek (2004*b*). Finally Figure 2 and Figure 3 confirm the asymptotic normality of the MQMLE claimed in Theorem 1.

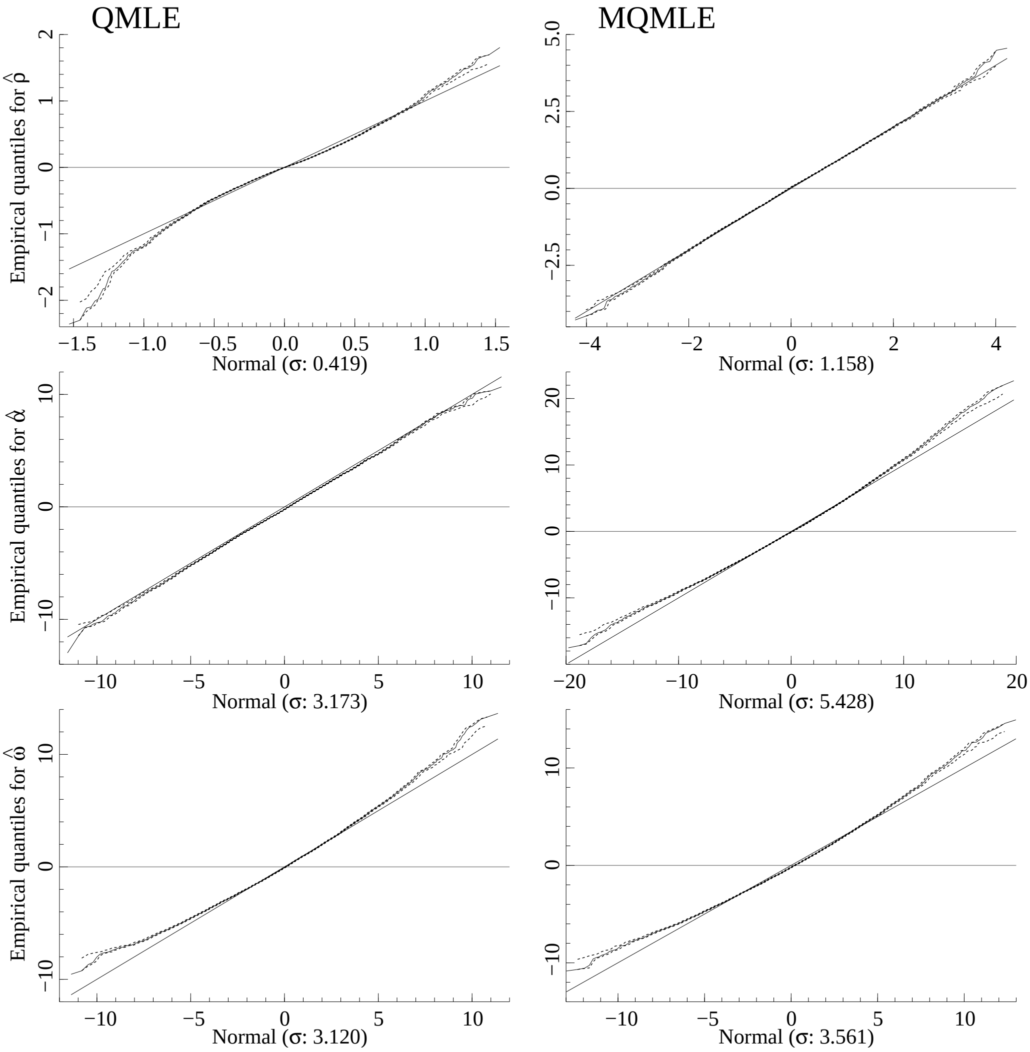

Figure 2: QQ-plots of $\sqrt{T}(\hat{\theta}_T^i - \theta_0)$ against a standard normal distribution with mean zero and the same variance as $\sqrt{T}(\hat{\theta}_T^i - \theta_0)$. The left column corresponds to the QMLE and the right to the MQMLE (with $M = 3$). The parameters are kept fixed at $\theta_0 = (0.5, 1.5, 1)'$ and $T = 500$ and the plot is based on 10,000 Monte Carlo replications. The dotted lines correspond to 95% confidence bands based on the empirical distribution function.

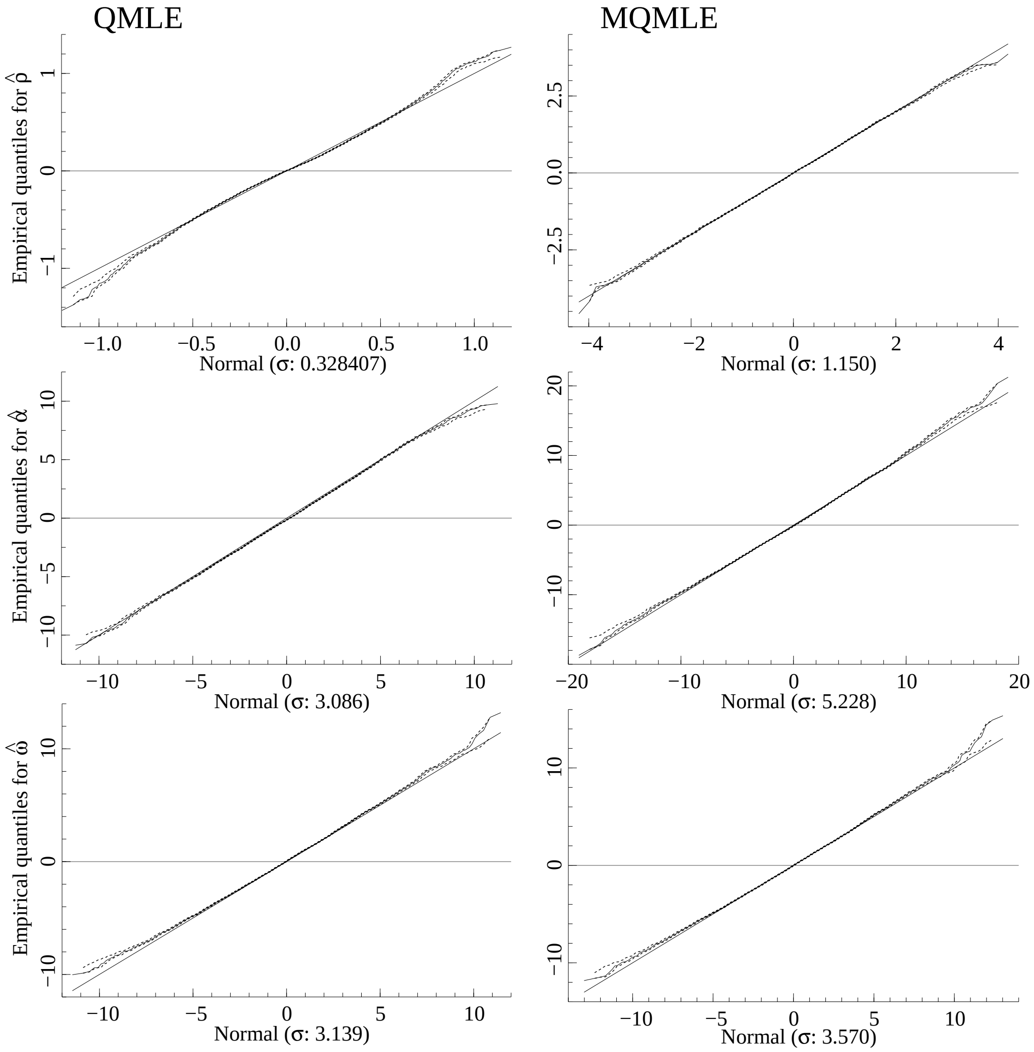

Figure 3: QQ-plots of $\sqrt{T}(\hat{\theta}_T^i - \theta_0)$ against a standard normal distribution with mean zero and the same variance as $\sqrt{T}(\hat{\theta}_T^i - \theta_0)$. The left column corresponds to the QMLE and the right to the MQMLE (with $M = 3$). The parameters are kept fixed at $\theta_0 = (0.5, 1.5, 1)'$ and T = 4,000 and the plot is based on 10,000 Monte Carlo replications. The dotted lines correspond to 95% confidence bands based on the empirical distribution function.

5 Implications and Summary

We have initially derived minimal conditions under which processes generated by the AR-ARCH model are geometrically ergodic. For the maximum likelihood estimator in this model we have conjectured that the parameter region for which the estimator is asymptotically normal can be extended from the fourth order moment condition of Ling & Li (1998) to a second order moment condition. As mentioned similar considerations have been made in Francq & Zakoïan (2004) and Ling (2007*a*). The paper also suggests a different estimator (MQMLE) which we prove to be asymptotically normal without any moment restrictions. By a Monte Carlo study we show that the MQMLE of the autoregressive parameter approximates its asymptotic distribution faster than the maximum likelihood estimator and that its asymptotic variance is only slightly larger. For the estimator of the ARCH parameter α the gain from using the MQMLE is a slightly lower bias, while the estimator of the scale parameter ω is unaffected by the choice of estimator.

On the basis of our results we suggest to implement the MQMLE choosing a censoring constant such that the observed censoring frequency is around 5%[2]. In our view this provides a good balance between low standard deviation on the estimator and a good normal approximation for the sample lengths usually encountered in financial econometrics, however, one should also consider the use of the estimates when deciding the estimation procedure (see the discussion in the previous section), since the two procedures have different strengths.

A Appendix

Proof of Lemma 1. This proof can be seen as verifying the high level conditions (CM.1)-(CM.4) in Kristensen (2005), for geometric ergodicity of general nonlinear state space models. Under Assumption 1 the result follows by combining two well known drift criterions for autoregressive- and ARCH processes, respectively. A detailed derivation can be found in Lange (2008*a*). The remaining part of the proof will therefore focus on Assumption 2 where α and ρ are both scalars.

Note first that x_t is a Markov chain. Using (1) - (3) twice one can express x_t

[2] Ox code for employing both estimators discussed in this paper can be downloaded from www.math.ku.dk/~lange.

in terms of x_{t-2} and the two innovations z_t and z_{t-1} as

$$x_t = \begin{pmatrix} \rho_0^2 y_{t-3} + z_{t-1}(\omega_0 + \alpha_0 \varepsilon_{t-2}^2)^{1/2} + \rho_0 \varepsilon_{t-2} \\ z_t(\omega_0 + \alpha_0(\omega_0 + \alpha_0 \varepsilon_{t-2}^2) z_{t-1}^2)^{1/2} \end{pmatrix}.$$

Conditional on x_{t-2} the map from (z_{t-1}, z_t) to x_t is a bijective and all points are regular (the determinant of the Jacobian matrix is non-zero for all points). Since the pair (z_{t-1}, z_t) has a density with respect to the two dimensional Lebesgue measure, which is strictly positive on compacts, a classical result regarding transformations of probability measures with densities yields that the two step transition kernel for the chain x_t has strictly positive density on compact sets. By Chan & Tong (1985) the 2-step chain is aperiodic, Lebesgue-irreducible, and all compact sets are small.

Below we establish that the 1-step chain satisfies a drift criterion, which for a drift function $V(x)$ can be formulated as

$$E[V(x_t) \mid x_{t-1} = x] \leq aV(x) + b,$$

with $0 < a < 1$ and $b > 0$. For the 2-step chain the law of the iterated expectation yields

$$E[V(x_t) \mid x_{t-2} = x] \leq E[aV(x_{t-1}) + b \mid x_{t-2} = x] \leq a^2 V(x) + ab + b.$$

Hence the 2-step chain also satisfies a drift criterion and it is therefore geometrically ergodic by Tjøstheim (1990). By Lemma 3.1 of Tjøstheim (1990) it therefore holds that the 1-step chain is geometrically ergodic as well.

In order to establish the drift criterion for the 1-step chain define the drift function

$$V(x_t) = 1 + |y_{t-1}|^\delta + C|\varepsilon_t|^\delta,$$

where $C > 0$ and $1 > \delta > 0$. Since V is continuous x_t is geometrically ergodic by the drift criterion of Tjøstheim (1990) if

$$\frac{E\left[V(x_t) \mid x_{t-1}\right]}{V(x_{t-1})} < 1, \tag{6}$$

for all x_{t-1} outside some compact set K. Simple calculations yield

$$\begin{aligned}\frac{E\left[V(x_t) \mid x_{t-1}\right]}{V(x_{t-1})} &= \frac{1+|\rho_0 y_{t-2}+\varepsilon_{t-1}|^\delta + C(\omega_0+\alpha_0\varepsilon_{t-1}^2)^{\delta/2}E\left[|z_t|^\delta\right]}{V(x_{t-1})} \\ &\leq \frac{1+C\omega_0^{\delta/2}}{V(x_{t-1})} + \frac{|\rho_0|^\delta|y_{t-2}|^\delta + (1+C\alpha_0^{\delta/2}E\left[|z_t|^\delta\right])|\varepsilon_{t-1}|^\delta}{V(x_{t-1})}.\end{aligned}$$

With $K(r) = \{x \in \mathbb{R}^2 \mid \|x\| < r\}$ the first fraction can be made arbitrarily small outside K by choosing r large enough. Next define the function $h(\delta) = \alpha_0^{\delta/2}E\left[|z_t|^\delta\right]$ and note that $h(0) = 1$ and $h'(0) = E\left[\log(\alpha_0 z_t^2)\right]/2$. The existence of the derivative from the right is guaranteed by Lebesgue's Dominated Convergence Theorem and the finite second order moment of z_t. Hence by assumption there exists a $\delta \in]0,1[$ such that $h(\delta) < 1$ and $|\rho_0| < 1$. Therefore the constant C can be chosen large enough such that (6) holds for all x_{t-1} outside the compact set K.

Finally the law of large numbers (4) follows from Theorem 1 of Jensen & Rahbek (2007). This completes the proof of Lemma 1. $\square$

All our asymptotic results are based on applying Lemma A.1, which follows. Note that conditions $(A.1)$ - $(A.4)$ are similar to conditions stated in the literature on asymptotic likelihood-based inference, see, e.g. Jensen & Rahbek (2004*a*) Lemma 1; Lehmann (1999) Theorem 7.5.2. The difference is that $(A.1)$ - $(A.4)$ avoid making assumptions on the third derivatives of the estimating function.

Lemma A.1. *Consider $\ell_T(\phi)$, which is a function of the observations $X_1, ..., X_T$ and the parameter $\phi \in \Phi \subseteq \mathbb{R}^k$. Introduce furthermore ϕ_0, which is an interior point of Φ. Assume that $\ell_T(\cdot) : \mathbb{R}^k \to \mathbb{R}$ is two times continuously differentiable in ϕ and that*

$(A.1)$ *As* $T \to \infty, \sqrt{T}\partial\ell_T(\phi_0)/\partial\phi \xrightarrow{D} N(0, \Omega_S),\ \Omega_S > 0.$

$(A.2)$ *As* $T \to \infty, \partial^2\ell_T(\phi_0)/\partial\phi\partial\phi' \xrightarrow{P} \Omega_I > 0.$

$(A.3)$ *There exists a continuous function* $F : \mathbb{R}^k \to \mathbb{R}^{k\times k}$ *such that* $\partial^2\ell_T(\phi)/\partial\phi\partial\phi' \xrightarrow{P} F(\phi)$ *for all* $\phi \in N(\phi_0)$.

$(A.4)$ $\sup_{\phi\in N(\phi_0)} \|\partial^2\ell_T(\phi)/\partial\phi\partial\phi' - F(\phi)\| \xrightarrow{P} 0,$

where $N(\phi_0)$ is a neighborhood of ϕ_0. Then there exists a fixed open neighborhood $U(\phi_0) \subseteq N(\phi_0)$ of ϕ_0 such that

$(B.1)$ *As* $T \to \infty$ *it holds that*

$$\begin{aligned} &P(\text{there exists a minimum point } \hat{\phi}_T \text{ of } \ell_T(\phi) \text{ in } U(\phi_0)) \to 1 \\ &P(\ell_T(\phi) \text{ is convex in } U(\phi_0)) \to 1 \\ &P(\hat{\phi}_T \text{ is unique and solves } \partial \ell_T(\phi)/\partial\phi = 0) \to 1 \end{aligned}$$

$(B.2)$ *As* $T \to \infty, \hat{\phi}_T \xrightarrow{P} \phi_0$.

$(B.3)$ *As* $T \to \infty, \sqrt{T}(\hat{\phi}_T - \phi_0) \xrightarrow{D} N(0, \Omega_I^{-1}\Omega_S\Omega_I^{-1})$.

Note that assumptions $(A.3)$ and $(A.4)$ could have been stated as a single condition, but for ease of exposition in the following proofs we have chosen this formulation.

Proof of Lemma A.1. By definition the continuous function $\ell_T(\phi)$ attains its minimum on any compact set $K(\phi_0, r) = \{\theta \,|\, \|\phi - \phi_0\| \leq r\} \subseteq N(\phi_0)$. With $v_\phi = (\phi - \phi_0)$, and ϕ^* on the line from ϕ to ϕ_0, Taylor's formula gives

$$\begin{aligned} \ell_T(\phi) - \ell_T(\phi_0) &= D\ell_T(\phi_0)v_\phi + \frac{1}{2}v_\phi' D^2\ell_T(\phi^*)v_\phi \\ &= D\ell_T(\phi_0)v_\phi + \frac{1}{2}v_\phi'[\Omega_I + (D^2\ell_T(\phi_0) - \Omega_I) \\ &\quad +(D^2\ell_T(\phi^*) - D^2\ell_T(\phi_0))]v_\phi. \end{aligned} \tag{7}$$

Note that

$$\begin{aligned} &\|D^2\ell_T(\phi^*) - D^2\ell_T(\phi_0)\| \\ = \; &\|D^2\ell_T(\phi^*) - F(\phi^*) + (F(\phi^*) - F(\phi_0)) + (F(\phi_0) - D^2\ell_T(\phi_0))\| \\ \leq \; &\|D^2\ell_T(\phi^*) - F(\phi^*)\| + \|F(\phi^*) - F(\phi_0)\| + \|F(\phi_0) - D^2\ell_T(\phi_0)\| \\ \leq \; &2 \sup_{\phi \in K(\phi_0, r)} \|D^2\ell_T(\phi) - F(\phi)\| + \sup_{\phi \in K(\phi_0, r)} \|F(\phi) - F(\phi_0)\|. \end{aligned}$$

The first term converges to zero as T tends to infinity by $(A.4)$ and the last term can be made arbitrarily small by the continuity of F. The remaining part of the proof is identically to the proof of Lemma 1 in Jensen & Rahbek (2004*a*). The only exception is that the upper bound on $\|D^2\ell_T(\phi^*) - D^2\ell_T(\phi_0)\|$ is not linear in r, but is a function which decreases to zero as r tends to zero. □

Proof of Lemma 2. We will begin by proving the part of the lemma regarding the log-likelihood function L_T^0. For exposition only we initially focus on the autoregressive parameter $\rho \in \mathbb{R}^r$. The derivations regarding the ARCH parameter α and the scale parameter ω are simple when compared with the ones with respect to ρ and are outlined in the last part of the proof. It is also there that the asymptotic results for the joint parameters are given.

Initially introduce some notation

$$\tilde{\varepsilon}_t(\theta) = (\varepsilon_t(\theta), ..., \varepsilon_{t-p+1}(\theta))', \quad \bar{y}_t = (y_t, ..., y_{t-r+1})', \quad \bar{\bar{y}}_t = (\bar{y}_t, ..., \bar{y}_{t-p+1})', \tag{8}$$

and A as the r by r matrix with the ARCH parameters (α) on the diagonal. The first and second derivatives of L_T^0 with respect to the autoregressive parameter are given below.

$$\begin{aligned}
\frac{\partial L_T^0(\theta)}{\partial \rho} &= \frac{1}{T}\sum_{t=1}^{T} \frac{2\eta_t(\theta)}{h_t(\theta)}\tilde{\varepsilon}_{t-1}(\theta)' A \bar{\bar{y}}_{t-2} - \frac{2\varepsilon_t(\theta)}{h_t(\theta)}\bar{y}_{t-1} \\
&=: \frac{1}{T}\sum_{t=1}^{T} s_t^{(\rho)}(\theta)' \\
\frac{\partial^2 L_T^0(\theta)}{\partial \rho^2} &= \frac{1}{T}\sum_{t=1}^{T} \frac{8\eta_t(\theta)}{h_t^2(\theta)}\bar{\bar{y}}_{t-2}' A \tilde{\varepsilon}_{t-1}(\theta)\tilde{\varepsilon}_{t-1}(\theta)' A \bar{\bar{y}}_{t-2} \\
&\quad + \frac{4}{h_t^2(\theta)}\bar{\bar{y}}_{t-2}' A \tilde{\varepsilon}_{t-1}(\theta)\tilde{\varepsilon}_{t-1}(\theta)' A \bar{\bar{y}}_{t-2} \\
&\quad - \frac{2\eta_t(\theta)}{h_t(\theta)}\bar{\bar{y}}_{t-2}' A \bar{\bar{y}}_{t-2} - \frac{8\varepsilon_t(\theta)}{h_t^2(\theta)}\bar{y}_{t-1}\tilde{\varepsilon}_{t-1}(\theta)' A \bar{\bar{y}}_{t-2} + \frac{2}{h_t(\theta)}\bar{y}_{t-1}\bar{y}_{t-1}', \\
&=: \frac{1}{T}\sum_{t=1}^{T} s_t^{(\rho\rho)}(\theta),
\end{aligned}$$

where $\eta_t(\theta) = \frac{\varepsilon_t^2(\theta)}{h_t(\theta)} - 1$. Recall the notational convention that $\varepsilon_t(\theta_0) =: \varepsilon_t$ etc. Note that the sequence $(s_t^{(\rho)})_{t=1}^T$ is a Martingale difference sequence with respect to the filtration $\mathbb{F}_t = \mathbb{F}(y_t, y_{t-1}, ..., y_0, y_{-1})$ and applying a standard CLT for Martingale differences (e.g. from Brown (1971)) leads to consider first the conditional second order moment

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} E\left[s_t^{(\rho)} s_t^{(\rho)\prime} \mid \mathbb{F}_{t-1}\right] \\
&\qquad = \frac{4}{T}\sum_{t=1}^{T}\left(\frac{\kappa}{h_t^2}\bar{\bar{y}}_{t-2} A \tilde{\varepsilon}_{t-1}\tilde{\varepsilon}_{t-1}' A \bar{\bar{y}}_{t-2}' + \frac{1}{h_t}\bar{y}_{t-1}\bar{y}_{t-1}'\right) \\
&\qquad \xrightarrow{P} 4\kappa E\left[\frac{1}{h_t^2}\bar{\bar{y}}_{t-2} A \tilde{\varepsilon}_{t-1}\tilde{\varepsilon}_{t-1}' A \bar{\bar{y}}_{t-2}'\right] + 4E\left[\frac{1}{h_t}\bar{y}_{t-1}\bar{y}_{t-1}'\right],
\end{aligned} \tag{9}$$

where the convergence is due to either Lemma 1 or Assumption 3. Note that the cross products vanish since we have assumed a symmetric distribution for z_t. Before addressing the Lindeberg condition of Brown (1971) it is noted that since the parameter region for which an ARCH(p) process has finite second order moment is given by a sharp inequality there exists a small constant, λ, such that it also has finite $2+\lambda$ moment, see Lemma 3. Hence the Lindeberg condition is satisfied as

$$\frac{1}{T}\sum_{t=1}^{T} E\left[\|s_t^{(\rho)}\|^2 1_{\left\{\|s_t^{(\rho)}\|>\delta\sqrt{T}\right\}} \mid \mathbb{F}_{t-1}\right] \quad \leq \quad \frac{1}{T^{1+\lambda/2}\delta^{\lambda}}\sum_{t=1}^{T} E\left[\|s_t^{(\rho)}\|^{2+\lambda} \mid \mathbb{F}_{t-1}\right] \to 0,$$

for all $\delta > 0$ as $T \to \infty$, which holds since $E\left[\|s_t^{(\rho)}\|^{2+\lambda}\right] < \infty$ by Lemma 3.

By Lemma 1 it holds that

$$\frac{1}{T}\sum_{t=1}^{T} s_t^{(\rho\rho)} \ \xrightarrow{P}\ 4E\left[\frac{1}{h_t^2}\bar{\tilde{y}}_{t-2}' A\tilde{\varepsilon}_{t-1}\tilde{\varepsilon}_{t-1}' A\bar{\tilde{y}}_{t-2}\right] + 2E\left[\frac{1}{h_t}\bar{y}_{t-1}\bar{y}_{t-1}'\right]. \tag{10}$$

The necessary finiteness of the moments is easily verified by utilizing that $\eta_t(\theta_0) = z_t^2 - 1$ and $\|\tilde{\varepsilon}_{t-1}/h_t^{1/2}\| < p/\min\{\alpha_0\}^{1/2} < \infty$.

The arguments with respect to the ARCH parameter α and the scale parameter ω, and hence the joint variation, are completely analogous to the ones applied above, and use repeatedly the inequalities $\|\tilde{\varepsilon}_{t-1}/h_t^{1/2}\| < p/\min\{\alpha_0\}^{1/2}$ and $1/h_t < 1/\omega_0$. For reference all first and second order derivatives are reported at the end of the Appendix.

Turning to the result regarding L_T^1 one notes that since $\gamma_t^1 = 1_{\{|y_{t-1}|,\ldots,|y_{t-r-p-2}|<M\}}$ the moments of interest can be bounded from above as follows

$$\begin{aligned} E&\left[\frac{\gamma_t^1}{h_t^2}\bar{\tilde{y}}_{t-2}' A\tilde{\varepsilon}_{t-1}\tilde{\varepsilon}_{t-1}' A\bar{\tilde{y}}_{t-2}\right] + E\left[\frac{\gamma_t^0}{h_t}\bar{y}_{t-1}\bar{y}_{t-1}'\right] \\ &\leq \ \frac{\max\{\alpha_0\}^2}{\min\{\alpha_0\}^2}M^2 + \frac{1}{\omega_0}M^2 < \infty. \end{aligned}$$

Hence the expectations in (9) and (10) are finite and the result can be derived using the same arguments as for L_T^0. Direct calculations yield the following expressions for the matrices Ω_I and Ω_S.

$$\Omega_S^i = \begin{pmatrix} 4\kappa\mu_1^i + 4\mu_2^i & 0 & 0 \\ 0 & \kappa\mu_3^i & \kappa\mu_5^{i\prime} \\ 0 & \kappa\mu_5^i & \kappa\mu_4^i \end{pmatrix}, \ \Omega_I^i = \begin{pmatrix} 4\mu_1^i + 2\mu_2^i & 0 & 0 \\ 0 & \mu_3^i & \mu_5^{i\prime} \\ 0 & \mu_5^i & \mu_4^i \end{pmatrix}, \tag{11}$$

where

$$\begin{aligned}
\mu_1^i &= E\left[\tfrac{\gamma_t^i}{h_t^2}\bar{\bar{y}}_{t-2}A\tilde{\varepsilon}_{t-1}\tilde{\varepsilon}_{t-1}'A\bar{\bar{y}}_{t-2}'\right] & \mu_2^i &= E\left[\tfrac{\gamma_t^i}{h_t}\bar{y}_{t-1}\bar{y}_{t-1}'\right] \\
\mu_3^i &= E\left[\tfrac{\gamma_t^i}{h_t^2}(\tilde{\varepsilon}_{t-1}\odot\tilde{\varepsilon}_{t-1})(\tilde{\varepsilon}_{t-1}\odot\tilde{\varepsilon}_{t-1})'\right] & \mu_4^i &= E\left[\tfrac{\gamma_t^i}{h_t^2}\right] \\
\mu_5^i &= E\left[\tfrac{\gamma_t^i}{h_t^2}(\tilde{\varepsilon}_{t-1}\odot\tilde{\varepsilon}_{t-1})'\right],
\end{aligned}$$

for $i = 0, 1$ where $\odot$ denote element by element multiplication and expectation is taken with respect to the invariant distribution.

In order to show that Ω_S^i is positive definite one must establish that for any non-zero vector $\xi = (\xi_1', \xi_2')' \in \mathbb{R}^{r+p}$ it holds that $\xi'\Omega_S^i\xi$ is strictly positive. Initially note that under Assumption 1 or Assumption 2 the drift criterion ensures that the invariant distribution for the Markov chain $x_t = (\bar{y}_{t-1}', \tilde{\varepsilon}_t')'$ has a density with respect to the Lebesgue measure on $\mathbb{R}^{r+p}$. Since the roots of a non-trivial polynomial has zero Lebesgue measure it holds that under the invariant measure $P(\xi_1'\bar{y}_{t-1}\bar{y}_{t-1}'\xi_1 = 0) = 0$. Furthermore, since the set $]-M, M[^r \times \mathbb{R}^p$ has strictly positive Lebesgue measure for all $M > 0$ it holds that under the invariant distribution $P(\gamma_t^i = 1) > 0$. Finally, note that $h_t \geq \omega_0 > 0$, which implies that under the invariant measure it holds that $P(\xi_1'\frac{\gamma_t^i}{h_t}\bar{y}_{t-1}\bar{y}_{t-1}'\xi_1 > 0) > 0$. Hence it can be concluded that under the invariant measure

$$E[\xi_1'\frac{\gamma_t^i}{h_t}\bar{y}_{t-1}\bar{y}_{t-1}'\xi_1] > 0$$

and μ_2^i is therefore positive definite. By symmetry μ_1^i is positive semi-definite and it can therefore be concluded that $4\kappa\mu_1^i + 4\mu_2^i$ is also positive definite. Likewise it can be shown that the matrix

$$\kappa\begin{pmatrix}\mu_3^i & \mu_5^{i\prime} \\ \mu_5^i & \mu_4^i\end{pmatrix}$$

is positive definite. It can therefore holds that under either Assumption 1 or Assumption 2 the matrix Ω_S^i is positive definite. Finally, one notes that the preceding arguments also guarantee that Ω_I^i is positive definite.

Under Assumption 3 one notes that since the innovations, z_t, have a strictly positive density with respect to the Lebesgue measure there exists an integer n such that the n-step transition kernel for the Markov chain x_t has strictly positive density with respect to the Lebesgue measure. This implies that the invariant measure also has a strictly positive density with respect to the Lebesgue measure. By combining this with the previous arguments one has that both Ω_S^i and Ω_I^i are positive definite. This completes the proof of Lemma 2. $\square$

Proof of Theorem 1. The proof of Theorem 1 is based on applying Lemma A.1. By Lemma 2 conditions $(A.1)$ and $(A.2)$ are satisfied. As in the proof of Lemma 2 we will focus on the vector of autoregressive parameters (ρ), and only briefly discuss the other parameters at the end of the proof.

Introduce lower and upper values for each parameter in θ, $\alpha_L < \alpha_0 < \alpha_U$ (element by element), $\omega_L < \omega_0 < \omega_U$, and a positive constant δ. In terms of these, the neighborhood $N(\theta_0)$ around the true value θ_0 is defined as

$$N(\theta_0) = \{(\rho, \alpha, \omega)' \in \mathbb{R}^{r+p+1} \mid \|\rho - \rho_0\| < \delta, \alpha_L < \alpha < \alpha_U, \omega_L < \omega_0 < \omega_U\}.$$

As in the previous proof set $s_t^{(\rho\rho)}(\theta) = \partial^2 l_t^1(\theta)/\partial\rho\partial\rho'$. In the following the existence of a stochastic variable $u_t^{(\rho\rho)}$, which satisfies

$$\sup_{\theta \in N(\theta_0)} \|s_t^{(\rho\rho)}(\theta)\| < u_t^{(\rho\rho)} \tag{12}$$

and has finite expectation with respect to the invariant measure is established. All terms of $s_t^{(\rho\rho)}(\theta)$ can be treated using similar arguments and we therefore only report the derivations regarding the most involved term. Note that

$$\begin{aligned} \sup_{\theta \in N(\theta_0)} & \frac{\gamma_t^1 \varepsilon_t^2(\theta)}{h_t^3(\theta)} \|\bar{\tilde{y}}_{t-2}' A \tilde{\varepsilon}_{t-1}(\theta) \tilde{\varepsilon}_{t-1}(\theta)' A \bar{\tilde{y}}_{t-2}\| \\ \leq & \; 2r^2 \frac{\max\{\alpha_U\}^2}{\min\{\alpha_L\}^2} M^2 \gamma_t^1 (\varepsilon_t(\theta_0)^2 + 2(\rho_0 \odot \rho_0 + \delta^2)'(\bar{y}_{t-1} \odot \bar{y}_{t-1})) \\ \leq & \; z_t C_1 + C_2 =: u_t, \end{aligned}$$

for some positive constants C_1 and C_2. By assumption $E|u_t| < \infty$ and by the triangle inequality there exists a constant K such that (12) holds with $u_t^{(\rho\rho)} = K u_t$. Hence by Lemma 1 one can define a function F as the point-by-point probability limit of the average of the $s_t^{(\rho\rho)}(\theta)$'s. Next by dominated convergence and the continuity of the function $s_t^{(\rho\rho)}(\theta)$ in θ the function F is also continuous, so $(A.3)$ holds.

Finally we must verify $(A.4)$. By $(A.1)$ - $(A.3)$ and Theorem 4.2.1 of Amemiya (1985) it suffices to show

$$E\left[\sup_{\rho\in N(\rho_0)} \|s_t^{(\rho\rho)}(\theta)\|\right] < \infty,$$

which follows follows from (12). Note that Theorem 4.2.1 is applicable in our setup by Amemiya (1985) p. 117 as the law of large numbers apply by Lemma 1. By inspecting the remaining derivatives, reported below, it is evident that the arguments above can be applied to all other derivatives, and applying Lemma A.1 completes the proof of Theorem 1. □

Lemma 3. *Under Assumption 1, 2, or 3, the assumptions from Lemma 1, and the additional assumption that the ARCH(p) process ε_t has finite second order moment, that is $\sum_{i=1}^{p}\alpha_{0,i} < 1$, it holds that there exists a constant $\lambda > 0$ such that $E[|\varepsilon_t|^{2+\lambda}]$ and $E[|y_t|^{2+\lambda}]$ are both finite.*

Proof. With out lose of generality we give the proof for the case where $p = 2$. From earlier results the drift criterion is applicable. Corresponding to the ARCH process we therefore define the Markov chain $x_t = (\varepsilon_t, ..., \varepsilon_{t-p+1})'$ and the drift function

$$g(x_t) = 1 + |\varepsilon_t|^{2+\lambda} + b_1|\varepsilon_{t-1}|^{2+\lambda},$$

for some positive constant λ. Verifying (2.4) and (2.5) of Lu (1996) establishes the first part of the lemma. Direct calculations yield

$$\begin{aligned}
& E[g(x_t) \mid x_{t-1} = (x_1, x_2)'] \\
= \; & 1 + E[|z_t|^{2+\lambda}](\omega_0 + \alpha_{0,1}x_1 + \alpha_{0,2}x_2)^{1+\lambda/2} + b_1|x_1|^{2+\lambda} \\
\leq \; & 1 + E[|z_t|^{2+\lambda}]\frac{(\omega_0 + \alpha_{0,1} + \alpha_{0,2})^{1+\lambda/2}}{\omega_0 + \alpha_{0,1} + \alpha_{0,2}}(\omega_0 + \alpha_{0,1}|x_1|^{2+\lambda} + \alpha_{0,2}|x_2|^{2+\lambda}) + b_1|x_1|^{2+\lambda} \\
= \; & 1 + c(\lambda)\omega_0 + (b_1 + c(\lambda)\alpha_{0,1})|x_1|^{2+\lambda} + c(\lambda)\alpha_{0,2}|x_2|^{2+\lambda}.
\end{aligned}$$

Where the inequality holds as the mapping $x \mapsto |x|^{1+\lambda}$ is convex. Since the innovations z_t have finite fourth order moment it holds by bounded convergence that $c(\lambda)$ tends to 1 as λ tends to zero. Hence there exists a $\lambda > 0$ such that $c(\lambda)(\alpha_{0,1} + \alpha_{0,2}) < 1$ and by mimicking the proof of Theorem 1 in Lu (1996) conditions (2.4) and (2.5) can be verified. The second half of the lemma follows directly from the infinite representation of the autoregressive process y_t. □

Derivatives: Recall the definition $\eta_t(\theta) = \frac{\varepsilon_t^2(\theta)}{h_t(\theta)} - 1$ and the definitions in (8). In the following $A \odot B$ denotes element by element multiplication of two matrices of identical dimensions and $1_{r\times p}$ a r by p matrix of ones. The first order partial derivatives of L_T^0 and L_T^1 are stated below.

$$
\begin{aligned}
\frac{\partial L_T^i(\theta)}{\partial\theta} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t^i(\theta)}{\partial\theta},\\
\frac{\partial l_t^i(\theta)}{\partial\rho} &= \frac{2\gamma_t^i\eta_t(\theta)}{h_t(\theta)}\tilde{\varepsilon}_{t-1}(\theta)' A\bar{\tilde{y}}_{t-2} - \frac{2\gamma_t^i\varepsilon_t(\theta)}{h_t(\theta)}\bar{y}_{t-1},\\
\frac{\partial l_t^i(\theta)}{\partial\alpha} &= \frac{-\gamma_t^i\eta_t(\theta)}{h_t(\theta)}(\tilde{\varepsilon}_{t-1}(\theta)\odot\tilde{\varepsilon}_{t-1}(\theta))',\\
\frac{\partial l_t^i(\theta)}{\partial\omega} &= \frac{-\gamma_t^i\eta_t(\theta)}{h_t(\theta)}.
\end{aligned}
$$

The second derivatives:

$$
\begin{aligned}
\frac{\partial^2 l_t^i(\theta)}{\partial\rho\partial\rho'} &= \frac{8\gamma_t^i\eta_t(\theta)}{h_t^2(\theta)}\bar{\tilde{y}}_{t-2}A\tilde{\varepsilon}_{t-1}(\theta)\tilde{\varepsilon}_{t-1}(\theta)'A\bar{\tilde{y}}_{t-2}' + \frac{4\gamma_t^i}{h_t^2(\theta)}\bar{\tilde{y}}_{t-2}'A\tilde{\varepsilon}_{t-1}(\theta)\tilde{\varepsilon}_{t-1}(\theta)'A\bar{\tilde{y}}_{t-2}\\
&\quad -\frac{2\gamma_t^i\eta_t(\theta)}{h_t(\theta)}\bar{\tilde{y}}_{t-2}'A\bar{\tilde{y}}_{t-2} - \frac{8\gamma_t^i\varepsilon_t(\theta)}{h_t^2(\theta)}\bar{y}_{t-1}\tilde{\varepsilon}_{t-1}(\theta)'A\bar{\tilde{y}}_{t-2} + \frac{2\gamma_t^i}{h_t(\theta)}\bar{y}_{t-1}\bar{y}_{t-1}',\\
\frac{\partial^2 l_t^i(\theta)}{\partial\alpha\partial\alpha'} &= (\frac{2\gamma_t^i\eta_t(\theta)}{h_t^2(\theta)} + \frac{\gamma_t^i}{h_t^2(\theta)})(\tilde{\varepsilon}_{t-1}(\theta)\odot\tilde{\varepsilon}_{t-1}(\theta))(\tilde{\varepsilon}_{t-1}(\theta)\odot\tilde{\varepsilon}_{t-1}(\theta))',\\
\frac{\partial^2 l_t^i(\theta)}{\partial\omega^2} &= \frac{2\gamma_t^i}{h_t^2(\theta)}\eta_t(\theta) + \frac{\gamma_t^i}{h_t^2(\theta)},\\
\frac{\partial^2 l_t^i(\theta)}{\partial\rho\partial\alpha'} &= \frac{2\gamma_t^i\eta_t(\theta)}{h_t(\theta)}(\tilde{\varepsilon}_{t-1}(\theta)\otimes 1_{1\times r})\odot\bar{\tilde{y}}_{t-2} - \frac{4\gamma_t^i\eta_t(\theta)}{h_t^2(\theta)}(\tilde{\varepsilon}_{t-1}(\theta)\odot\tilde{\varepsilon}_{t-1}(\theta))\tilde{\varepsilon}_{t-1}'A\bar{\tilde{y}}_{t-2}\\
&\quad -\frac{2\gamma_t^i}{h_t^2(\theta)}(\tilde{\varepsilon}_{t-1}(\theta)\odot\tilde{\varepsilon}_{t-1}(\theta))\tilde{\varepsilon}_{t-1}'A\bar{\tilde{y}}_{t-2} + \frac{2\gamma_t^i\varepsilon_t(\theta)}{h_t^2(\theta)}(\tilde{\varepsilon}_{t-1}(\theta)\odot\tilde{\varepsilon}_{t-1}(\theta))\bar{y}_{t-1}',\\
\frac{\partial^2 l_t^i(\theta)}{\partial\rho\partial\omega} &= \frac{-4\gamma_t^i\eta_t(\theta)}{h_t^2(\theta)}\tilde{\varepsilon}_{t-1}'A\bar{\tilde{y}}_{t-2} + \frac{2\gamma_t^i\varepsilon_t(\theta)}{h_t^2(\theta)}\bar{y}_{t-1}' - \frac{2\gamma_t^i}{h_t^2(\theta)}\tilde{\varepsilon}_{t-1}'A\bar{\tilde{y}}_{t-2},\\
\frac{\partial^2 l_t^i(\theta)}{\partial\alpha\partial\omega} &= (\frac{2\gamma_t^i\eta_t(\theta)}{h_t^2(\theta)} + \frac{\gamma_t^i}{h_t^2(\theta)})(\tilde{\varepsilon}_{t-1}(\theta)\odot\tilde{\varepsilon}_{t-1}(\theta))',
\end{aligned}
$$

where $i = 0, 1$.

First and second order non-linear cointegration models

By Theis Lange

Department of Mathematical Sciences, University of Copenhagen
E-mail: *lange@math.ku.dk*

Abstract: This paper studies cointegration in non-linear error correction models characterized by discontinuous and regime-dependent error correction and variance specifications. In addition the models allow for autoregressive conditional heteroscedasticity (ARCH) type specifications of the variance. The regime process is assumed to depend on the lagged disequilibrium, as measured by the norm of linear stable or cointegrating relations. The main contributions of the paper are: *i*) conditions ensuring geometric ergodicity and finite second order moment of linear long run equilibrium relations and differenced observations, *ii*) a representation theorem similar to Granger's representations theorem and a functional central limit theorem for the common trends, *iii*) to establish that the usual reduced rank regression estimator of the cointegrating vector is consistent even in this highly extended model, and *iv*) asymptotic normality of the parameters for fixed cointegration vector and regime parameters. Finally, an application of the model to US term structure data illustrates the empirical relevance of the model.

Keywords: Cointegration, Non-linear adjustment, Regime switching, Multivariate ARCH.

1 Introduction

Since the 1980's the theory of cointegration has been hugely successful. Especially Granger's representation theorem, see Johansen (1995), which provides conditions under which non-stationary vector autoregressive (VAR) models can exhibit stationary, stable linear combinations. This very intuitive concept of stable relations is probably the main reason why cointegration models have been so widely applied (even outside the world of economics). For an up to date discussion see the survey Johansen (2007).

However, recent empirical studies suggest that the adjustments to the stable relations might not be adequately described by the linear specification employed in the traditional cointegration model. When modeling key macroeconomic variables such as GNP, unemployment, real exchange rates, or interest rate spreads, non-linearities can be attributed to transaction costs, which induces a band of no disequilibrium adjustment. For a more thorough discussion see e.g. Dumas (1992), Sercu, Uppal & Van Hulle (1995), Anderson (1997), Hendry & Ericsson (1991), and Escribano (2004). Furthermore, policy interventions on monetary or foreign exchange markets may also cause non-linear behavior, see Ait-Sahalia (1996) and Forbes & Kofman (2000) among others. Such non-linearities can also explain the problem of seemingly non-constant parameters encountered in many applications of the usual linear models. To address this issue Balke & Fomby (1997) suggested the threshold cointegration model, where the adjustment coefficients may switch between a specific set of values depending on the cointegrating relations. Generalizations of this model has lead to the smooth transition models, see Kapetanios, Shin & Snell (2006) and the references therein and the stochastically switching models, see e.g. Bec & Rahbek (2004), and Dufrenot & Mignon (2002) and the many references therein.

Parallel to this development the whole strain of literature devoted to volatility modeling has documented that non-linearities should also be included in the specification of the variance of the innovations. A large, and ever growing, number of autoregressive conditional heteroscedasticity (ARCH) type models, originally introduced by Engle (1982) and generalized by Bollerslev (1986), has been suggested, see e.g. Bauwens, Laurent & Rombouts (2006) for a recent discussion of multivariate generalized ARCH models.

Motivated by these findings, this paper proposes a cointegration model, which allows for non-linearities in both the disequilibrium adjustment and the variance specifications. The model will be referred to as the first and second order non-linear cointegration vector autoregressive (FSNL-CVAR) model. The adjustments to the stable relations are assumed to be switching according to a threshold state process, which depends on past observations. Thus, the model extends the concept of threshold cointegration as suggested in Balke & Fomby (1997). The main novelty of the FSNL-CVAR model is to adopt a more general variance specification in which the conditional variance is allowed to depend on both the current regime as well as lagged values of the innovations, herby including an important feature of financial time series.

Constructing a model which embeds many of the previously suggested models opens up the use of likelihood based tests to assess the relative importance of these models. For instance, does the inclusion of a regime dependent covariance matrix render the traditional ARCH specification obsolete or vice versa? Furthermore, since both the mean- and variance parameters depend on the current regime a test for no regime effect in for example the mean equation can be conducted as a simple χ^2-test, since the issue of vanishing parameters under the null hypothesis, and resulting non-standard limiting distributions see e.g. Davies (1977), has been resolved by retaining the dependence on the regime process in the variance specification.

The present paper derives easily verifiable conditions ensuring geometric ergodicity, and hence the existence of a stationary initial distribution, of the first differences of the observations and of the linear cointegrating relationships. Stability and geometric ergodicity results form the basis for law of large numbers theorems and are therefore an important step not only towards an understanding of the dynamic properties of the model, but also towards the development of an asymptotic theory. The importance of geometric ergodicity has recently been emphasized by Jensen & Rahbek (2007), where a general law of large numbers is shown to be a direct consequence of geometric ergodicity. It should be noted that the conditions ensuring geometric ergodicity do not involve the parameters of adjustment in the inner regime, corresponding to the band of no action in the example above. The paper also derives a representation theorem corresponding to the well known Granger representation theorem and establishes a functional central limit theorem (FCLT) for the common trends. Finally, asymptotic normality of the parameter estimates is shown to hold under the assumption of known cointegration vector and threshold parameters. The results are applied to US term structure data. The empirical analysis finds clear evidence indicating that the short-term and long-term rates only adjusts to one another when the spread is above a certain threshold. In order to achieve a satisfactory model fit the inclusion of ARCH effects is paramount. Hence the empirical analysis support the need for cointegration models, which are non-linear in both the mean and the variance. Finally, the empirical study shows that adjustments occurs through the short rate only, which is in accordance with the expectation hypothesis for the term structure.

The rest of the paper is structured as follows. Section 2 presents the model and the necessary regularity conditions. Next Section 3 contains the results regarding stability and order of integration. Estimation and asymptotic theory is discussed in Section 4 and the empirical study presented in Section5. Conclusions are presented in Section 6 and all proofs can be found in Appendix.

The following notation will be used throughout the paper. For any vector $\|\cdot\|$ denotes the Euclidian vector norm and I_p a p-dimensional unit matrix. For some $p\times r$ matrix β of rank $r \leq p$, define the orthogonal complement $\beta_\perp$ as the $p\times(p-r)$-dimensional matrix with the property $\beta'\beta_\perp = 0$. The associated orthogonal projections are given by $I_p = \bar{\beta}\beta' + \bar{\beta}_\perp\beta_\perp$ with $\bar{\beta} = \beta(\beta'\beta)^{-1}$. Finally $\varepsilon_{i,t}$ denotes the i'th coordinate of the vector ε_t. In Section 2 and 3 and the associated proofs only the true parameters will be considered and the usual subscript 0 on the true parameters will be omitted to avoid an unnecessary cumbersome notation.

2 The first and second order non-linear cointegration model

In this section the model is defined and conditions for geometric ergodicity of process generated according to the model are stated. As discussed the model is non-linear in both the mean- and variance specification, which justifies referring to the model as the first and second order non-linear cointegration vector autoregressive (FSNL-CVAR) model.

2.1 Non-linear adjustments

Let X_t be a p-dimensional observable stochastic process. The process is driven by both an unobservable i.i.d. sequence ν_t and a zero-one valued state process s_t. It is assumed that the distribution of the latter depends on lagged values of the observable process and that ν_t is independent of s_t. The evolution of the observable process is governed by the following generalization of the usual CVAR model, see e.g. Johansen (1995).

$$
\begin{aligned}
\Delta X_t &= s_t \left(a^{(1)} \beta' X_{t-1} + \sum_{j=1}^{q-1} \Gamma_j \Delta X_{t-j} \right) \\
&\quad + (1 - s_t) \left(a^{(0)} \beta' X_{t-1} + \sum_{j=1}^{q-1} G_j \Delta X_{t-j} \right) + \varepsilon_t \qquad (1) \\
\varepsilon_t &= H^{1/2}(s_t, \varepsilon_{t-1}, ..., \varepsilon_{t-q}) \nu_t = H_t^{1/2} \nu_t,
\end{aligned}
$$

where $a^{(0)}, a^{(1)}$, and β are $p \times r$ matrices, $(\Gamma_j, G_j)_{j=1,...,q-1}$ are $p \times p$ matrices, and ν_t an i.i.d.$(0, I_p)$ sequence. By letting the covariance matrix H_t depend on lagged innovations $\varepsilon_{t-1}, ..., \varepsilon_{t-q}$ the model allows for a very broad class of ARCH type specifications. The exact specification of the covariance matrix will be addressed in the next section, but by allowing for dependence of the lagged innovations the suggested model permits traditional ARCH type dynamics of the innovations. Saikkonen (2008) has suggested to use lagged values of the observed process X_t in the conditional variance specification, however, this leads to conditions for geometric ergodicity, which cannot be stated independently for the mean- and variance parameters and a less clear cut definition of a unit root.

As indicated in the introduction the proposed model allows for non-linear and discontinuous equilibrium correction. The state process could for instance be specified such that if the deviation from the stable relations, measured by $\|\beta' X_{t-1}\|$, is below some predefined threshold adjustment to the stable relations occurs through $a^{(0)}$ and as a limiting case no adjustment occurs, which could reflect transaction costs. However, if $\|\beta' X_{t-1}\|$ is large adjustment will take place through $a^{(1)}$. For applications along theses lines, see Akram & Nymoen (2006), Chow (1998), and Krolzig, Marcellino & Mizon (2002).

2.2 Switching autoregressive heteroscedasticity

Depending on the value of the state process at time t the covariance matrix is given by

$$\begin{aligned} H_t &= D_t^{1/2}\Lambda^{(l)}D_t^{1/2} && (2)\\ D_t &= \mathrm{diag}(\Pi_t) \\ \Pi_t &= (\pi_{1,t},...,\pi_{p,t})' \\ \pi_{i,t} &= 1 + g_i(\varepsilon_{i,t-1},...,\varepsilon_{i,t-q}), \quad i = 1,...,p && (3) \end{aligned}$$

with Λ_l a positive definite covariance matrix, $g_i(\cdot)$ a function onto the non-negative real numbers for all $i = 1,...,p$, and $l = 0,1$ corresponds to the possible values of the state process.

The factorization in (2) isolates the effect of the state process into the matrix $\Lambda^{(l)}$ and the ARCH effect into the diagonal matrix D_t. This factorization implies that all information about correlation is contained in the matrix $\Lambda^{(l)}$, which switches with the regime process. In this respect the variance specification is related to the constant conditional correlation (CCC) model of Bollerslev (1990) and can be viewed as a mixture generalization of this model.

For example, suppose that $p = 2$, $q = 1$, $g_i(\varepsilon_{i,t-1}) = \alpha_i\varepsilon^2_{i,t-1}$, and $s_t = 1$ almost surely for all t. Then the conditional correlation between $X_{1,t}$ and $X_{2,t}$ is given by the off-diagonal element of Λ_1, which illustrates that the model in this case is reduced to the traditional cointegration model with the conditional variance specified according to the CCC model.

Since the functions $g_1,...,g_p$ allow for a feedback from past realizations of the innovations to the present covariance matrix it is necessary to impose some regularity conditions on these functions in order to discuss stability of the cointegrating relations $\beta'X_t$ and ΔX_t.

Assumption 1. (i) *For all* $i = 1,...,p$ *there exists constants, denoted* $\alpha_{i,1},...,\alpha_{i,q}$, *such that for* $\|(\varepsilon'_{t-1},...,\varepsilon'_{t-q})'\|$ *sufficiently large it holds that* $g_i(\varepsilon_{i,t-1},...,\varepsilon_{i,t-q}) \leq \sum_{j=1}^{q}\alpha_{i,j}\varepsilon^2_{i,t-j}$.

(ii) *For all* $i = 1,...,p$ *the sequence of constants satisfies* $\max_{l=0,1}\Lambda^{(l)}_{i,i}\sum_{j=1}^{q}\alpha_{i,j} < 1$.

The assumption essentially ensures that as the lagged innovations became large the covariance matrix responds no more vigorously than an ARCH(q) process with finite second order moment. However, for smaller shocks the assumption allows for a broad range of non-linear responses.

2.3 The State Process

Initially recall that the state or switching variable s_t is zero-one valued. Next define the $r + p(q-1)$-dimensional variable z_t as

$$z_t = (X'_{t-1}\beta, \Delta X'_{t-1}, ..., \Delta X'_{t-q+1})'. \tag{4}$$

By assumption the dynamics of the state process are given by the conditional probability

$$P(s_t = 1 \mid X_{t-1}, ..., X_0, s_{t-1}, ..., s_0) = P(s_t = 1 \mid z_t) \equiv p(z_t). \tag{5}$$

Some theoretical results regarding univariate switching autoregressive models where the regime process is similar to (5) can be found in Gourieroux & Robert (2006).

The transition function $p(\cdot)$ will be assumed to be an indicator function taking the value one outside a bounded set as suggested by Balke & Fomby (1997). In the transaction cost example it is intuitively clear that as the distance from the stable relation increases so does the probability of adjustment to the stable relation. This leads to:

Assumption 2. *The transition probability $p(z)$ defined in (5) is zero-one valued and tends to one as $\|z\| \to \infty$.*

On the basis of Assumption 2 it is natural to refer to the regime where $s_t = 0$ as the inner regime and the other as the outer regime. As in Bec & Rahbek (2004) additional inner regimes can be added without affecting the validity of the results, the only difference being a more cumbersome notation. Extending the model to include additional outer regimes, as in Saikkonen (2008), leads inevitable to regularity conditions expressed in terms of the joint spectral radius of a class of matrices, which are in most cases impossible to verify.

It follows that the FSNL-CVAR model allows for epochs of equilibrium adjustments and epochs without. Furthermore the model allows these epochs to be characterized by different correlation structures and have general ARCH type variance structure. In the next section stationarity and geometric ergodicity of $\beta' X_t$ and $\beta'_\perp \Delta X_t$ are discussed.

3 Stability and order of integration

In the first part of this section conditions which ensure geometric ergodicity of $\beta' X_t$ and ΔX_t will be derived. In the second part of the section we derive a representation theorem corresponding to the well known Granger representation theorem and establish a FCLT for the common trends. The results presented in this section rely on by now classical Markov chain techniques, see Meyn & Tweedie (1993) for an introduction and definitions.

3.1 Geometric Ergodicity

Note initially that the process X_t generated by (1), (2), and (5) is not in itself a Markov chain due to the time varying components of the covariance matrix. Define therefore the process $V_t = (X_t'\beta, \Delta X_t'\beta_\perp)'$ where the orthogonal projection $I_p = \bar{\beta}\beta' + \bar{\beta}_\perp \beta'_\perp$, has been used. Furthermore define the stacked processes

$$\bar{V}_t = (V_t', ..., V_{t-q+1}')', \quad \bar{\varepsilon}_t = (\varepsilon_t', ..., \varepsilon_{t-q+1}')',$$

and the selection matrix

$$\varphi = (I_p, 0_{p\times(q-1)p})'.$$

By construction the process $\bar{V}_t$ is generated according to the VAR(1) model given by

$$\bar{V}_t = s_t \mathbf{A} \bar{V}_{t-1} + (1 - s_t)\mathbf{B}\bar{V}_{t-1} + \eta_t, \tag{6}$$

where $\eta_t = \varphi(\beta', \beta'_\perp)'\varepsilon_t$ and hence is a mean zero random variable with variance $(\beta', \beta'_\perp)' H_{t,s_t}(\beta', \beta'_\perp)$. The matrices $\mathbf{A}$ and $\mathbf{B}$ are implicitly given by (1), see (19) in the appendix for details. Finally the Markov chain to be considered can be defined as $Y_t = (\bar{\varepsilon}'_t, \bar{V}'_{t-q})'$. Note that $V_t, ..., V_{t-q+1}$ and $s_t, ..., s_{t-q+1}$ are computable from Y_t, which can be seen by first computing s_{t-q+1} then V_{t-q+1} and repeating.

As hinted earlier it suffices to assume that the usual cointegration assumption is satisfied for the parameters of the outer regime;

Assumption 3. *Assume that the rank of $a^{(1)}$ and β equals r and furthermore that there are exactly $p - r$ roots at $z = 1$ for the characteristic polynomial*

$$A(z) = I_p(1 - z) - a^{(1)}\beta' z - \sum_{i=1}^{q-1} \Gamma_i (1 - z) z^i, \quad z \in \mathbb{C} \tag{7}$$

while the remaining roots are larger than one in absolute value.

Next the main result regarding geometric ergodicity of the Markov chain Y_t will be stated, with the central conditions expressed in terms of the matrices $\mathbf{A}$ and $\mathbf{B}$. In the subsequent corollaries some important special cases are considered.

Theorem 1. *Assume that:*

(i) *For some $qp \times n$, $qp \geq n \geq 0$ matrix μ of rank n, it holds that $(\mathbf{A} - \mathbf{B})\mu = 0$.*

(ii) *The largest in absolute value of the eigenvalues of $\mathbf{A}$ is smaller than one.*

(iii) *The stochastic state variable s_t is zero-one valued and the state probability is zero-one valued and satisfies $p(\gamma'\bar{v}) \to 0$ as $\|\gamma'\bar{v}\| \to \infty$ with $\gamma = \mu_\perp$ and $\bar{v} \in \mathbb{R}^{pq}$*

(iv) *ν_t is i.i.d.(0,I_p) and has a continuous and strictly positive density with respect to the Lebesgue measure on $\mathbb{R}^p$.*

(v) *The generalized ARCH functions $g_1, ..., g_p$ satisfy the conditions listed in Assumption 1.*

Then Y_t is a geometrically ergodic process, which has finite second order moment.

Furthermore there exists a distribution for Y_0 such that the sequence $(Y_t)_{t=1}^T$ is stationary.

The formulation of the FSNL-CVAR model is in essence a combination of the non-linear cointegration model of Bec & Rahbek (2004) and the constant conditional correlation model of Bollerslev (1990). It is therefore not surprising that the following corollaries show that the stability conditions for the FSNL-CVAR model are simply a combination of the stability conditions associated with the these two "parent" models.

Corollary 1. *Under Assumption 1, 2, and 3, and* (iv) *of Theorem 1 it holds that Y_t is a geometrically ergodic process with finite second order moment and that there exists a distribution for Y_0 such that the sequence $(Y_t)_{t=1}^T$ is stationary.*

In the special case where the transition probability in (5) only depends on $\beta' X_t$ Assumption 2 does not hold. However, using Theorem 1 the following result can be established.

Corollary 2. *Consider the case where the transition probability in (5) only depends on $z_t = \beta' X_t$ and that $\Gamma_j = G_j$ for $j = 1, ..., q-1$. Under Assumption 1, 2, and 3 and* (iv) *of Theorem 1 it holds that Y_t is a geometrically ergodic process with finite second order moment and that there exists a distribution for Y_0 such that the sequence $(Y_t)_{t=1}^T$ is stationary.*

3.2 Non-stationarity

When considering linear VAR models the concept of I(1) processes is well defined, see Johansen (1995). This is in contrast to non-linear models, such as the FSNL-CVAR model, where there still exists considerable ambiguity as to how to define I(1) processes. In this paper we follow Corradia, Swanson & White (2000) and Saikkonen (2005) and simply define an I(1) process as a process for which a functional central limit theorem (FCLT) applies. In Theorem 2 below we establish conditions for which the $(p-r)$ common trends of X_t have a non-degenerate long-run variance and a FCLT applies.

Theorem 2. *Under the assumptions of Theorem 1 the process X_t given by (1), (2), and (5) has the representation*

$$X_t = C\sum_{i=1}^{t}(\varepsilon_t + (\Phi^{(0)} - \Phi^{(1)})u_t) + \tau_t, \tag{8}$$

where the processes τ_t, see 20, and $u_t = (1 - s_t)z_t$ are stationary and z_t is defined in (4). Furthermore the parameters $C, \Phi^{(0)}$, and $\Phi^{(1)}$ are defined by

$$\begin{aligned} C &= \beta_\perp({a^{(1)}}'_\perp(I_p - \sum_{i=1}^{k-1}\Gamma_i)\beta_\perp)^{-1}{a^{(1)}}'_\perp, \quad \Phi^{(0)} = (a^{(0)}, G_1, ..., G_{k-1}), \\ \Phi^{(1)} &= (a^{(1)}, \Gamma_1, ..., \Gamma_{k-1}). \end{aligned}$$

The $(p - r)$ common trends of X_t are given by $\sum_{i=1}^{t} c_i$, where $c_t = {a^{(1)}}'_\perp(\varepsilon_t + (\Phi^{(0)} - \Phi^{(1)})u_t)$. A FCLT applies to $c_t - Ec_T$, if

$$\Upsilon = \psi' \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{pmatrix} \psi > 0, \quad \textit{where} \ \ \psi' = {a^{(1)}}'_\perp(I_p, \Phi^{(0)} - \Phi^{(1)}). \tag{9}$$

The Σ matrices are the long run variances and the exact expression can be found in (21).

Note that sufficient conditions for Υ being positive definite are $\mathrm{sp}(\Phi^{(0)}) = \mathrm{sp}(\Phi^{(1)})$ or ${a^{(1)}}'_\perp\Sigma_{\epsilon\beta} = 0$.

4 Estimation and asymptotic normality

In this section it is initially established that the usual estimator in the linear cointegrated VAR model of the cointegration vector β, which is based on reduced rank regression (RRR), see Johansen (1995), is consistent even when data is generated by the much more general FSNL-CVAR model. The second part of this section considers estimation and asymptotic theory of the remaining parameters.

Define $\hat{\beta}$ as the usual RRR estimator of the cointegration vector defined in Johansen (1995) Theorem 6.1. Introduce the normalized estimator $\tilde{\beta}$ given by

$\tilde{\beta} = \hat{\beta}(\bar{\beta}'\hat{\beta})^{-1}$, where $\bar{\beta} = \beta(\beta'\beta)^{-1}$. Note that this normalization is clearly not feasible in practice as the matrix β is not know, however, the purpose of the normalized version is only to facilitate the formulation of the following consistency result.

Theorem 3. *Under the assumptions of Theorem 1 and the additional assumption that Υ defined in (9) is positive definite, $\tilde{\beta}$ is consistent and $\tilde{\beta} - \beta = o_p(T^{1/2})$.*

Theorem 3 suggests that once the cointegrating vector β has been estimated using RRR the remaining parameters can be estimated by quasi-maximum likelihood using numerical optimization. To further reduce the curse of dimensionality the parameters $\Lambda^{(1)}$ and $\Lambda^{(0)}$ can be concentrated out of the log-likelihood function as discussed in Bollerslev (1990). An Ox implementation of the algorithm can be downloaded from www.math.ku.dk/$\sim$lange.

In order to discuss asymptotic theory restrict the variance specification in (2) to linear ARCH(q), that is replace (3) by

$$\pi_{i,t} = 1 + \sum_{j=1}^{q} \alpha_{i,j}\varepsilon_{i,t-j}^2 \tag{10}$$

and define the parameter vectors

$$\theta^{(1)} = \text{vec}(\Phi^{(1)}, \Phi^{(0)}), \qquad \theta^{(2)} = (\alpha_{1,1}, ..., \alpha_{p,1}, \alpha_{2,1}, ..., \alpha_{p,q}),$$
$$\theta^{(3)} = (\text{vech}(\Lambda^{(1)})', \text{vech}(\Lambda^{(0)})')',$$

and $\theta = (\theta^{(1)\prime}, \theta^{(2)\prime}, \theta^{(3)\prime})'$. As is common let θ_0 denote the true parameter value. If the cointegration vector β and the threshold parameters are assumed known the realization of state process is computable and the quasi log-likelihood function to be optimized is, apart from a constant, given by

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = -\log(|H_t(\theta)|)/2 - \varepsilon_t(\theta)'H_t(\theta)^{-1}\varepsilon_t(\theta)/2, \tag{11}$$

where $\varepsilon_t(\theta)$ and $H_t(\theta)$ are given by (1) and (2), respectively. The assumption of known β and λ is somewhat unsatisfactory, but at present necessary to establish the result. Furthermore, the assumption can be partly justified by recalling that estimators of both the cointegration vector and the threshold parameter are usually super consistent. The proof of the following asymptotic normality result can be found in the appendix.

Theorem 4. *Under the assumptions of Corollary 1 and the additional assumption that there exists a constant $\delta > 0$ such that $E[\|\varepsilon_t\|^{4+\delta}]$ and $E[\|\nu_t\|^{4+\delta}]$ are both finite and $\theta^{(2)} > 0$ it holds that when β and the parameters of the regime process are kept fixed at true values there exists a fixed open neighborhood around the true parameter $N(\theta_0)$ such that with probability tending to one as T tends to infinity, $L_T(\theta)$ has a unique minimum point $\hat{\theta}_T$ in $N(\theta_0)$. Furthermore, $\hat{\theta}_T$ is consistent and satisfies*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0, \Omega_I^{-1}\Omega_S\Omega_I^{-1}),$$

where $\Omega_S = E[(\partial l_t(\theta_0)/\partial\theta)(\partial l_t(\theta_0)/\partial\theta')]$ *and* $\Omega_S = E[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta']$.

The proof is given in the appendix, where precise expressions for the asymptotic variance are also stated.

5 An application to the interest rate spread

In this section an analysis of the spread between the long and the short U.S. interest rates using the FSNL-CVAR model is presented. The analysis is similar to the analysis of German interest rate spreads presented in Bec & Rahbek (2004). However, since the FSNL-CVAR model allows for heteroscedasticity the present analysis will employ daily data unlike the analysis in Bec & Rahbek (2004), which is based on monthly averages. The well-known expectations hypothesis of the term structure implies that, under costless and instantaneous portfolio adjustments and no arbitrage the spread between the long and the short rate can be represented as

$$R(k,t) - R(1,t) = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{i} E_t[\Delta R(1,t+j)] + L(k,t), \tag{12}$$

where $R(k,t)$ denotes the k-period interest rate at time t, $L(k,t)$ represents the term premium, accounting for risk and liquidity premia, and $E_t[\cdot]$ the expectation conditional on the information at time t, see e.g. Bec & Rahbek (2004) for details. Clearly, the right hand side is stable or stationary provided interest rate changes and the the term premium are stationary (see Hall, Anderson & Granger (1992)). In fact portfolio adjustments are neither costless nor instantaneous. It is therefore reasonable to assume that the spread $S(k,1,t) = R(k,t) - R(1,t)$, will temporarily depart from its equilibrium value given by (12). However, once portfolio adjustments have taken place (12) will again hold. Hence, the long and the short interest rate should be cointegrated with a cointegration vector of $\beta = (1,-1)'$. Testing this implication of the expectations hypothesis of the term structure has been the focus for many empirical papers, however, the results are not clear cut. Indeed the U.S. spread is found to be stationary in e.g. Campbell & Shiller (1987), Stock & Watson (1988), Anderson (1997), and Tzavalis & Wickens (1998), but integrated of order 1 in e.g. Evans & Lewis (1994), Enders & Siklos (2001), and Bec, Guayb & Guerre (2008). Note however, that when allowing for a stationary non-linear alternative the last two papers reject the hypothesis of non-stationarity of the U.S. spread. Indeed Anderson (1997) establishes that if one considers homogeneous transaction costs which reduces the investors yield on a bond by a constant amount, say λ, then one expects that the yield spread is stationary, but non-linear, since portfolio adjustments will only occur when the difference between the actual spread $S(k,1,t)$ and the value predicted by (12) is larger in absolute value than λ.

According to Anderson's argument the joint dynamics of short-term and long-term interest rates could be described by the non-linear error correction model given by (1):

$$\Delta X_t = (s_t a^{(1)} + (1-s_t)a^{(0)})\beta' X_t + \sum_{j=1}^{k-1} \Delta X_{t-j} + \varepsilon_t, \tag{13}$$

where $X_t = (R_t^S, R_t^L)$, denotes the short and the long rates and the transition function is defined in accordance with Anderson's argument. However, as it is a well established fact that daily interest rates exhibit considerable heteroscedasticity the model must include time dependent variance as in (10).

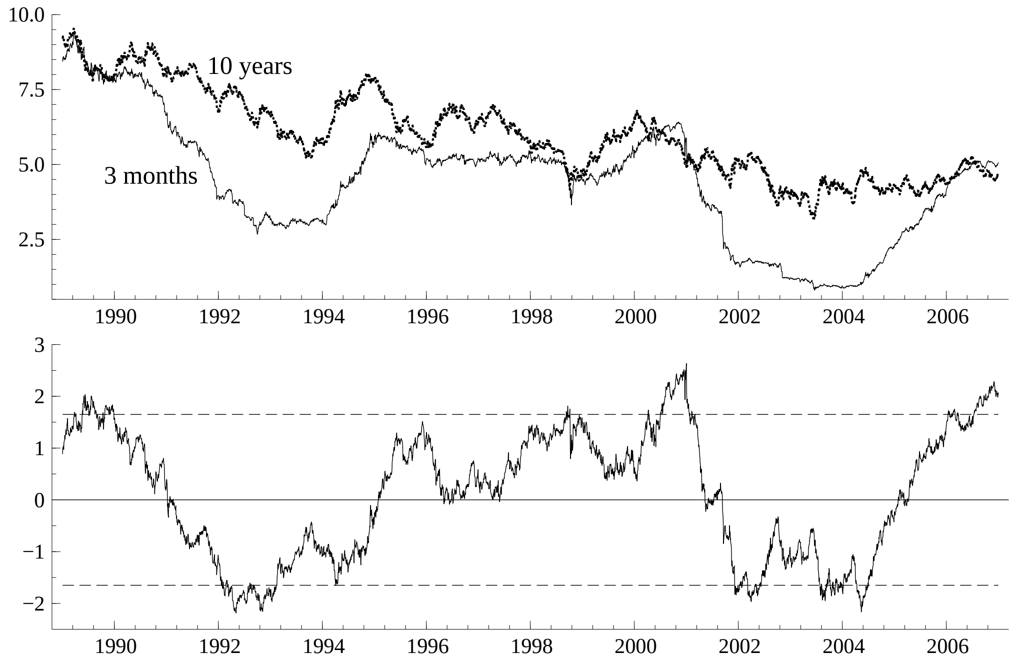

Figure 1: The 3-month and 10-year interest rates (top panel) and the spread between the two series adjusted for their mean (bottom panel). The dashed lines indicate the threshold $\lambda = 1.65$.

In the following the proposed FSNL-CVAR model will be applied to daily recordings of the U.S. 3-Month Treasury Constant Maturity Rate and the U.S. 10-Year Treasury Constant Maturity Rate spanning the period from 1/1-1988 to 1/1-2007 yielding a total of 4,500 observations. Data have been downloaded from the webpage of the Federal Reserve Bank of St. Louis. Following Bec & Rahbek (2004) both series are corrected for their average and the state process is therefore given by $s_t = 1_{\{|S^G_{t-1}| \geq \lambda\}}$, with $S^G_{t-1} = \beta' X_{t-1}$. This amounts to approximate the long-run equilibrium given by (12) by the average of the actual spread, as is common in the literature. Figure 1 depicts the data.

Initially a self-exiting threshold autoregressive (SETAR) model was fitted to the series S^G_t, which indicated a threshold parameter of $\lambda = 1.65$. This value is very close to the threshold parameter value of 1.7 reported in Bec & Rahbek (2004) for a similar study based on monthly German interest rate data. For the remaining part of the analysis the threshold parameter will be kept fixed at 1.65. However, it should be noted that by determining the threshold parameter in such a data dependent way the conditions for the asymptotic results given in Theorem 4 are formally not met. In this respect, recall from the vast literature on univariate threshold models that the threshold parameter is super-consistent and hence can be treated as fixed when making inference on the remaining parameters, we would expect this to hold in this case as well. Furthermore, as can be seen from (13) the short-term parameters Γ_i are assumed to be identical over the two regimes, the estimators and covariances in Theorem 4 should be adjusted accordingly.

Concerning the specification of lag lengths in (13), additional lags were included until there were no evidence of neither autocorrelation nor additional heteroscedasticity in the residuals. This lead us to retain seven lags in the mean equation and six lags in the variance equation. The choice of lag specification was confirmed by both the AIC as well as statistical test indicating that additional lags were not statistically significant at the 5% level.

The parameter estimates of the mean equation are reported in the first two columns of Table 1. Initially it is noted that the estimated parameters seem to confirm our conjecture that when the spread is below the threshold value no adjustment towards the equilibrium occurs. This is confirmed by testing the hypothesis that $a^{(0)} = (0,0)'$, which is accepted with a p-value of 0.60 using the LR test. In addition the estimates of $a^{(1)}$ indicate that long-term rates do not seem to adjust to disequilibrium. This is confirmed by the LR test of the hypothesis $a_1^{(0)} = a_2^{(0)} = a_2^{(1)} = 0$ which cannot be rejected. The test statistic equals 2.2 corresponding to a p-value of 0.53. The result implies that big spreads significantly affects the short-term rate only, which is in accordance with the expectation hypothesis for the term structure. This conclusion as well as the sign of the estimate of $a_1^{(1)}$ coincides with the findings of Bec & Rahbek (2004). Estimates of this restricted model are reported in the last two columns of Table 1 for the parameters of the mean equation and Table 2 for the parameters of the variance equation.

Table 2 reports the estimates of the variance equations. In order to ease comparison with traditional ARCH models the parametrization has been changed slightly from the one presented in (2) to directly reporting the coefficients of the equation $\Lambda_{1,1}^{(1)}\pi_{1,t} = \Lambda_{1,1}^{(1)} + \sum_{j=1}^{6} \Lambda_{1,1}^{(1)}\alpha_{i,j}\varepsilon_{1,t-j}^2$, which gives the conditional variance for the first element of ε_t when $s_t = 1$ and likewise for the other cases. It

	Unrestricted		Restricted	
	ΔR_t^S	ΔR_t^L	ΔR_t^S	ΔR_t^L
$1_{\{\|S_{t-1}^G\|\geq 1.65\}} S_{t-1}^G$	**-0.00116**	0.000996	**-0.00136**	-
	(-2.26)	(1.09)	(-2.86)	
$1_{\{\|S_{t-1}^G\|<1.65\}} S_{t-1}^G$	0.000478	-9.81e-005	-	-
	(0.92)	(-0.11)		
ΔR_{t-1}^S	**0.0687**	0.00626	**0.0687**	0.00693
	(3.77)	(0.30)	(3.79)	(0.34)
ΔR_{t-1}^L	-0.0163	**0.0636**	-0.0163	**0.0629**
	(-1.82)	(3.80)	(-1.83)	(3.84)
ΔR_{t-2}^S	**-0.0717**	-0.00262	**-0.0713**	-0.00175
	(-3.91)	(-0.13)	(-3.88)	(-0.09)
ΔR_{t-2}^L	8.56e-005	-0.0295	-0.000117	-0.0301
	(0.01)	(-1.80)	(-0.01)	(-1.86)
ΔR_{t-3}^S	**-0.0587**	0.0107	**-0.0582**	0.0115
	(-2.71)	(0.53)	(-2.85)	(0.58)
ΔR_{t-3}^L	0.00217	**-0.0531**	0.00212	**-0.0537**
	(0.24)	(-3.28)	(0.24)	(-3.35)
ΔR_{t-4}^S	**0.0583**	-0.0177	**0.0589**	-0.0165
	(3.14)	(-0.89)	(3.17)	(-0.82)
ΔR_{t-4}^L	-0.0111	-0.0224	-0.0111	-0.0231
	(-1.18)	(-1.37)	(-1.19)	(-1.41)
ΔR_{t-5}^S	**0.158**	**0.0476**	**0.158**	**0.0483**
	(8.47)	(2.46)	(8.54)	(2.50)
ΔR_{t-5}^L	-0.0162	-0.0116	-0.0159	-0.0121
	(-1.71)	(-0.66)	(-1.69)	(-0.70)
ΔR_{t-6}^S	**-0.0747**	-0.0282	**-0.0749**	-0.0278
	(-4.29)	(-1.43)	(-4.30)	(-1.45)
ΔR_{t-6}^L	**0.0270**	-0.00558	**0.0271**	-0.00614
	(3.07)	(-0.34)	(3.10)	(-0.37)
LM(ARCH 1-8)	0.6966	0.6596	0.6963	0.6598
[p-value]	[0.69]	[0.73]	[0.70]	[0.73]
LM(AR 1-8)	34.99		34.79	
[p-value]	[0.32]		[0.34]	
log-L	14,909.10		14,908.00	

Table 1: Model (13) estimates. t-statistics are reported in parentheses. LM tests of no remaining ARCH and no vector autocorrelation, respectively. Statistically significant parameters are indicated in bold.

	$s_t = 1$		$s_t = 0$	
	$\Lambda_{1,1}^{(1)}\pi_{1,t}$	$\Lambda_{2,2}^{(1)}\pi_{2,t}$	$\Lambda_{1,1}^{(0)}\pi_{1,t}$	$\lambda_{2,2}^{(0)}\pi_{2,t}$
Intercept	0.000412	0.00195	0.000475	0.00226
	(0.02)	(0.12)	(0.02)	(0.25)
ε_{t-1}^2	**0.148**	**0.0270**	**0.171**	**0.0313**
	(18.30)	(2.80)	(21.87)	(2.79)
ε_{t-2}^2	**0.159**	**0.0430**	**0.183**	**0.0499**
	(18.14)	(3.54)	(21.51)	(3.60)
ε_{t-3}^2	**0.235**	0.0241	**0.271**	0.0279
	(24.48)	(1.72)	(35.49)	(1.72)
ε_{t-4}^2	**0.112**	**0.0540**	**0.129**	**0.0626**
	(18.44)	(4.24)	(22.30)	(4.34)
ε_{t-5}^2	**0.170**	**0.115**	**0.196**	**0.134**
	(18.63)	(7.25)	(22.70)	(7.65)
ε_{t-6}^2	**0.0603**	**0.0558**	**0.0695**	**0.0647**
	(8.84)	(4.70)	(9.22)	(4.74)
Correlation	**0.425**		**0.396**	
	(16.56)		(29.75)	

Table 2: Estimates of the variance parameters. The parameters in e.g. the first column correspond to the coefficients in the variance equation $\Lambda_{1,1}^{(1)}\pi_{1,t} = \Lambda_{1,1}^{(1)} + \sum_{j=1}^{6} \Lambda_{1,1}^{(1)}\alpha_{i,j}\varepsilon_{1,t-j}^2$. t-statistics are reported in parentheses and were computed by the delta-method based on the expressions for the asymptotic variance in Theorem 4. Statistically significant parameters are indicated in bold.

should be noted that the change of parametrization has been performed after estimating the parameters using the parametrization of (2), which is preferable when performing the numerical optimization of the log-likelihood function as the parameters in $\Lambda^{(1)}$ and $\Lambda^{(1)}$ can be concentrated out. The reported t-statistics have therefore been computed using the delta-method.

The reported parameter estimates clearly demonstrates the presence of heteroscedasticity in the residuals. Examining the covariance matrix of the parameters collected in $\theta^{(3)}$ (covariance matrix not reported) indicates that $\Lambda_{i,i}^{(1)}$ and $\Lambda_{i,i}^{(1)}$ are statistically different for both the short- and the long rate. As expected the correlation is highest when adjustment to disequilibrium occurs ($s_t = 1$), but the hypothesis that the correlations are identical cannot be rejected at the 5% level. Hence the parameter estimates indicates the overall level of variance is highest in the regime where no adjustment occurs, but the correlation might be the same.

Since the parameters of the switching mechanism can be identified based solely on the variance specification the central hypothesis $a^{(1)} = a^{(0)}$, corresponding to no switching in the mean equation, can be tested by the standard LR test statistic, which will be asymptotically χ^2 distributed with two degrees of freedom. Thus avoiding the usual problems of unidentified parameters under the null, see e.g. Davies (1977), Davies (1987), and the survey Lange & Rahbek (2008), often encountered when testing no-switching hypothesis. The test statistic is 18.78 and the hypothesis of no switching in the mean equation is therefore clearly rejected.

It should be noted that the sum of the ARCH coefficients in each column of Table 2 are very close to one, which violates the fourth order moment condition of Theorem 4. However, as argued in Lange, Rahbek & Jensen (2007) based on a univariate model, we expect the asymptotic normality to hold even under a weaker second order moment condition.

6 Conclusion

In this paper we have suggested a cointegrated vector error correction model with a non-linear specification of both adjustments to disequilibrium and variance characterized by regime switches. Since the FSNL-CVAR model embeds many previously suggested models, see the discussion in the introduction, it provides a framework for assessing the relative importance of these models in a likelihood based setup. Furthermore tests of hypothesis such as linearity of the mean equation, which previously led to non-standard limiting distributions, can be conducted as standard χ^2-tests in the FSNL-CVAR model since the state process can be identified through the variance specification.

Using Markov chain results we derive easily verifiable conditions under which $\beta' X_t$ and ΔX_t are stable with finite second order moment and can be embedded in a Markov chain, which is geometrically ergodic. The usefulness of this result is enhanced by the recent work of Jensen & Rahbek (2007), which provides a general law of large numbers assuming only geometric ergodicity. Furthermore, a representation theorem corresponding to Granger's representation theorem has been derived and a functional central limit theorem for the common trends established. This is utilized to show that the usual RRR estimator of Johansen (1995) for the cointegrating vector, β, is robust to the model extensions suggested by the FSNL-CVAR model.

Finally, we establish asymptotic normality of the estimated parameters for fixed and known cointegration vector and threshold parameters. Applying the model to daily recordings of the US term structure documents the empirical relevance of the FSNL-CVAR model and the empirical results are in accordance with the expectation hypothesis of the term structure. Specifically it is found that small interest rate spreads are not corrected, while big ones have a significant influence on the short rate only.

Appendix

Proof of Theorem 1

Consider initially the homogenous Markov chain $Y_t = (\bar{\varepsilon}_t', \bar{V}_{t-q}')'$. Before applying the drift criterion, see e.g. Tjøstheim (1990) or Meyn & Tweedie (1993) it must be verified that the Markov chain Y_t is irreducible, aperiodic, and that compact sets are small. In order to do so it will be verified that the $2q$-step transition kernel has a density with respect to the Lebesgue measure, which is positive and bounded away from zero on compact sets.

Note that $V_t, ..., V_{t-q+1}$ and $s_t, ..., s_{t-q+1}$ are computable from Y_t, which can be seen by first computing s_{t-q+1} then V_{t-q+1} and repeating this procedure. With $h(\cdot \mid \cdot)$ denoting a generic conditional density with respect to an appropriate measure the $2q$-step transition kernel can be rewritten as follows (for exposition only the derivations for $q = 2$ are presented).

$$\begin{aligned} h(Y_t \mid Y_{t-4}) &= h(\varepsilon_t \mid \varepsilon_{t-1}, V_{t-2}, V_{t-3}, Y_{t-4}) h(\varepsilon_{t-1} \mid V_{t-2}, V_{t-3}, Y_{t-4}) \\ &\quad h(V_{t-2} \mid V_{t-3}, Y_{t-4}) h(V_{t-3} \mid Y_{t-4}) \\ &= h(\varepsilon_t \mid s_t, \varepsilon_{t-1}, \varepsilon_{t-2}) h(\varepsilon_{t-1} \mid s_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}) \\ &\quad h(V_{t-2} \mid s_{t-2}, V_{t-3}, V_{t-4}, \varepsilon_{t-3}, \varepsilon_{t-4})) \\ &\quad h(V_{t-3} \mid s_{t-3}, V_{t-4}, V_{t-5}, \varepsilon_{t-4}, \varepsilon_{t-5})). \end{aligned}$$

By (iv) and since the conditional covariance matrices are always positive definite the last four densities are Lebesgue densities, strictly positive, and bounded away from zero on compact sets. Hence the $2q$-step transition kernel for Y_t will also have a strictly positive density, which is bounded away from zero on compact sets. This establishes that the Markov chain Y_t is irreducible, aperiodic, and that compact sets are small and we proceed by applying the drift criterion of Tjøstheim (1990).

Denote a generic element of the Markov chain Y_t by $y = (\bar{v}', \bar{e}')'$ and define the drift function f by

$$f(Y_t) = 1 + k\bar{V}_{t-q}' D\bar{V}_{t-q} + \sum_{i=1}^{p}\sum_{j=0}^{q-1} k_{i,j}\varepsilon_{i,t-j}^2, \quad D \equiv \sum_{i=0}^{\infty} \mathbf{A}^{j\prime}\mathbf{A}^j, \tag{14}$$

where D is well defined as $\rho(\mathbf{A}\otimes\mathbf{A}) < 1$ by assumption and the positive constants $k, k_{i,j}$ will be specified later. The choice of the matrix D follows Feigin & Tweedie (1985) and it implies the existence of second-order moments of Y_t. Using (6) and following Bec & Rahbek (2004) it holds that

$$\begin{aligned}
& E[\bar{V}_{t-q}' D\bar{V}_{t-q} \mid Y_{t-1} = y] \\
&= \bar{v}'\mathbf{A}'D\mathbf{A}\bar{v} + (1 - p(\gamma'\bar{v}))\left\{\bar{v}'\mathbf{B}'D\mathbf{B}\bar{v} - \bar{v}'\mathbf{A}'D\mathbf{A}\bar{v}\right\} + E[\eta_{t-q}' D\eta_{t-q} \mid Y_{t-1} = y] \\
&= \bar{v}'\mathbf{A}'D\mathbf{A}\bar{v} + (1 - p(\gamma'\bar{v}))\left\{\bar{v}'(\mathbf{A}-\mathbf{B})'D(\mathbf{A}-\mathbf{B})\bar{v} - 2\bar{v}'\mathbf{A}'D(\mathbf{A}-\mathbf{B})\bar{v}\right\} \\
&\quad + E[\eta_{t-q}' D\eta_{t-q} \mid Y_{t-1} = y] \\
&= \bar{v}'D\bar{v} - \bar{v}'\bar{v} + (1 - p(\gamma'\bar{v}))\left\{(\bar{v}'\gamma)\bar{\gamma}'(\mathbf{A}-\mathbf{B})'D(\mathbf{A}-\mathbf{B})\bar{\gamma}(\gamma'\bar{v})\right. \\
&\quad \left. - 2\bar{v}'\mathbf{A}'D(\mathbf{A}-\mathbf{B})\bar{\gamma}(\gamma'\bar{v})\right\} + E[\eta_{t-q}' D\eta_{t-q} \mid Y_{t-1} = y].
\end{aligned} \tag{15}$$

In the last equality the projection $I_{pq} = \bar{\gamma}\gamma' + \bar{\mu}\mu'$ and (i) have been used. Define for some $\lambda_c > 1$ the compact set

$$C_v = \{\bar{v} \in \mathbb{R}^{pq} \mid \bar{v}'D\bar{v} \leq \lambda_c\}.$$

On the complement of C_v it holds that

$$\frac{\bar{v}'D\bar{v} - \bar{v}'\bar{v}}{\bar{v}'D\bar{v}} = 1 - \frac{\bar{v}'\bar{v}}{\bar{v}'D\bar{v}} \leq 1 - \inf_{\bar{v}\neq 0}\frac{\bar{v}'\bar{v}}{\bar{v}'D\bar{v}} \leq 1 - \frac{1}{\rho(D)},$$

where $\rho(\cdot)$ denotes the spectral radius of a square matrix. Furthermore note that

$$(\bar{v}'\gamma)\bar{\gamma}'(\mathbf{A}-\mathbf{B})'D(\mathbf{A}-\mathbf{B})\bar{\gamma}(\gamma'\bar{v}) \asymp \|\gamma'\bar{v}\|^2$$

and that

$$2\bar{v}'\mathbf{A}'D(\mathbf{A}-\mathbf{B})\bar{\gamma}(\gamma'\bar{v}) \asymp \|\bar{v}\|\,\|\gamma'\bar{v}\|,$$

where $h_1(x) \asymp h_2(x)$ denotes that $h_1(x)/h_2(x)$ tends to a non-zero constant as $\|x\|$ tends to infinity. However, by assumption the drift function satisfies $f(y) \asymp \|y\|^2 = \|\bar{\gamma}\gamma'\bar{v} + \bar{\mu}\mu'\bar{v}\|^2 + \|\bar{e}\|^2$ and since $(1-p(\gamma'\bar{v})) \to 0$ as $\|\gamma'\bar{v}\| \to \infty$ it can be concluded that

$$\frac{k(1-\tilde{p}(\gamma'\bar{v}))\left\{(\bar{v}'\gamma)\bar{\gamma}'(\mathbf{A}-\mathbf{B})'D(\mathbf{A}-\mathbf{B})\bar{\gamma}(\gamma'\bar{v}) - 2\bar{v}'\mathbf{A}'D(\mathbf{A}-\mathbf{B})\bar{\gamma}(\gamma'\bar{v})\right\}}{f(y)} \to 0,$$

as $\|\bar{v}\| \to \infty$. Or in other words, for λ_c adequately large it holds that on the complement of C_v will

$$kE[\bar{V}_{t-q}D\bar{V}_{t-q} \mid Y_{t-1}=y] \leq \frac{(1-\delta^*)k\bar{v}'D\bar{v}}{f(y)}f(y) + kE[\eta_{t-q}'D\eta_{t-q} \mid Y_{t-1}=y]. \quad (16)$$

The constant δ^* should be chosen such that $\delta^* \in]0,1[$ and $\frac{1}{\rho(D)} > \delta^* > 0$. Next consider the final term of (16). By construction it will be positive and there exists positive constants c, c_i where $i=1,...,p$ such that

$$\begin{aligned} kE[\eta_{t-q}'D\eta_{t-q} \mid Y_{t-1}=y] &= kE[\varepsilon_{t-q}'(\beta',\beta_\perp')\varphi'D\varphi(\beta',\beta_\perp')'\varepsilon_{t-q} \mid Y_{t-1}=y] \\ &\leq ck + k\sum_{i=1}^{p} c_i e_{i,q}^2. \end{aligned} \quad (17)$$

As previously define the compact set $C_e = \{\bar{e} \in \mathbb{R}^{pq} \mid \|\bar{e}\|^2 \leq \lambda_c\}$. Furthermore if

λ_c is chosen large enough (v) yields that on the complement of C_e it holds that

$$
\begin{aligned}
& c_i k e_{i,q}^2 + \sum_{j=0}^{q-1} k_{i,j} E[\varepsilon_{i,t-j}^2 \mid Y_{t-1} = y] \\
&= c_i k e_{i,q}^2 + k_{i,0} E[\varepsilon_{i,t}^2 \mid Y_{t-1} = y] + \sum_{j=1}^{q-1} k_{i,j} e_{i,j}^2 \\
&\leq c_i k e_{i,q}^2 + k_{i,0} \bar{\sigma}_i \left(1 + \sum_{j=1}^{q} \alpha_{i,j} e_{i,j}^2\right) + \sum_{j=1}^{q-1} k_{i,j} e_{i,j}^2 \\
&= K + (k_{i,0}\bar{\sigma}_i \alpha_{i,1} + k_{i,1}) e_{i,1}^2 + \sum_{j=2}^{q-1} (k_{i,0}\bar{\sigma}_i \alpha_{i,j} + k_{i,j}) e_{i,j}^2 \\
&\quad + (c_i k + k_{i,0}\bar{\sigma}_i \alpha_{i,q}) e_{i,q}^2 \qquad (18)
\end{aligned}
$$

for all $i = 1, ..., p$ where K is some positive constant and $\bar{\sigma}_i = \max_{l=0,1} \Lambda_{i,i}^{(l)}$. When $\bar{\sigma}_i \sum_{j=1}^{q} \alpha_{i,j} < 1$ the positive constants k and $k_{i,j}$ can be chosen such that the inequalities

$$
\begin{aligned}
k_{i,0}\bar{\sigma}_i \alpha_{i,1} + k_{i,1} &< k_{i,0} \\
k_{i,0}\bar{\sigma}_i \alpha_{i,j} + k_{i,j} &< k_{i,j-1}, \quad j = 2, ..., q-1 \\
c_i k + k_{i,0}\bar{\sigma}_i \alpha_{i,q} &< k_{i,q-1}
\end{aligned}
$$

are all satisfied, which can be seen by setting $k_{i,0} = 1$ and choosing k very small, see Lu (1996) for details. Hence there exists a constant $\delta_i^{**} \in]0, 1[$ such that the coefficient of $e_{i,j+1}^2$ in (18) is smaller than $(1 - \delta_i^{**}) k_{i,j}$ for all $j = 0, ..., q-1$.

By combing (16)-(18) it can be concluded that for y outside the compact set $C = C_v \times C_\varepsilon$ will

$$
\begin{aligned}
E[f(Y_t) \mid Y_{t-1} = y] &\leq \frac{(1-\delta)\bar{v}' D \bar{v} + (1-\delta) \sum_{i=1}^{p} \sum_{j=0}^{q-1} k_{i,j} e_{i,j+1}^2}{1 + \bar{v}' D \bar{v} + \sum_{i=1}^{p} \sum_{j=0}^{q-1} k_{i,j} e_{i,j+1}^2} f(y) \\
&\leq (1-\delta) f(y),
\end{aligned}
$$

where $\delta = \min(\delta^*, \delta_1^{**}, ..., \delta_p^{**}) > 0$. Inside the compact set C the function

$E[f(Y_t) \mid Y_{t-1} = y]$ is continuous and hence bounded. This completes the verification of the drift criterion.

Proof of Corollary 1

Assume without loss of generality that $q = 2$. Under the assumptions listed in the corollary the coefficient matrices of (6) are given by

$$\mathbf{A} = \begin{pmatrix} \beta' a^{(1)} - I_r + \beta'\Gamma_1\bar{\beta} & \beta'\Gamma_1\bar{\beta}_\perp & -\beta'\Gamma_1\bar{\beta} & 0 \\ \beta'_\perp a^{(1)} + \beta'_\perp\Gamma_1\bar{\beta} & \beta'_\perp\Gamma_1\bar{\beta}_\perp & -\beta'_\perp\Gamma_1\bar{\beta} & 0 \\ 0 & 0 & I_r & 0 \\ 0 & 0 & 0 & I_{p-r} \end{pmatrix}, \tag{19}$$

and likewise for $\mathbf{B}$. Hence

$$(\mathbf{A}-\mathbf{B}) = \begin{pmatrix} \beta'(a^{(1)} - a^{(0)}) + \beta'(\Gamma_1 - G_1)\bar{\beta} & \beta'(\Gamma_1 - G_1)\bar{\beta}_\perp & -\beta'(\Gamma_1 - G_1)\bar{\beta} & 0 \\ \beta'_\perp(a^{(1)} - a^{(0)}) + \beta'_\perp(\Gamma_1 - G_1)\bar{\beta} & \beta'_\perp(\Gamma_1 - G_1)\bar{\beta}_\perp & -\beta'_\perp(\Gamma_1 - G_1)\bar{\beta} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

So the matrices γ and μ can be chosen as

$$\gamma = \begin{pmatrix} I_r & 0 & 0 \\ 0 & I_{p-r} & 0 \\ 0 & 0 & I_r \\ 0 & 0 & 0 \end{pmatrix}, \quad \mu = \begin{pmatrix} 0 \\ 0 \\ 0 \\ I_{p-r} \end{pmatrix},$$

which satisfies $\mu = \gamma_\perp$ and the remaining assumptions of Theorem 1. Finally Theorem 1 yields the desired result.

Proof of Corollary 2

Assume again without loss of generality that $q = 2$. In this case the matrices γ and μ can be chosen as

$$\gamma = \begin{pmatrix} I_r \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mu = \begin{pmatrix} 0 & 0 & 0 \\ I_{p-r} & 0 & 0 \\ 0 & I_r & 0 \\ 0 & 0 & I_{p-r} \end{pmatrix},$$

and an application of Theorem 1 completes the proof.

Proof of Theorem 2

The formulation of the theorem as well as the proof owes much to Theorem 4 of Bec & Rahbek (2004). Initially note that the process X_t given by (1), (2), and (5) can be written as

$$A(L)X_t = (\Phi^{(0)} - \Phi^{(1)})u_t + \varepsilon_t,$$

where L denotes the lag-operator and the polynomial $A(\cdot)$ is defined in (7). By the algebraic identity

$$A(z)^{-1} = C\frac{1}{1-z} + C(z),$$

where $C(z) = \sum_{i=0}^{\infty} C_i z^i$ with exponentially decreasing coefficients C_i it holds that

$$\begin{aligned} X_t &= C\sum_{i=1}^{t}((\Phi^{(0)} - \Phi^{(1)})u_i + \varepsilon_i) + C(L)((\Phi^{(0)} - \Phi^{(1)})u_t + \varepsilon_t) \\ &= C\sum_{i=1}^{t}((\Phi^{(0)} - \Phi^{(1)})u_i + \varepsilon_i) + \tau_t. \end{aligned} \tag{20}$$

Next Theorem 1 yields that $\beta' X_t$ and ΔX_{t-i} are stationary and in turn that $u_t = (1 - s_t)z_t$ is stationary. Since $C(L)$ has exponentially decreasing coefficients it therefore holds that τ_t is stationary. Hence the common trends of X_t are given by

$$\sum_{i=1}^{t} c_i = {a^{(1)}}'_{\perp} \sum_{i=1}^{t} ((\Phi^{(0)} - \Phi^{(1)})u_i + \varepsilon_i).$$

Since $\|u_t\| \leq \|z_t\|$ it holds by Theorem 17.4.2 and Theorem 17.4.4 of Meyn & Tweedie (1993) that a FCLT applies to c_t provided the long run variance Υ,

$$\Upsilon = \gamma_{cc}(0) + \sum_{h=1}^{\infty}(\gamma_{cc}(h) + \gamma_{cc}(h)'), \quad \gamma_{cc}(h) = \mathrm{Cov}(c_t, c_{t+h}),$$

is positive definite. Note that the long run variance can be written as

$$\Upsilon = \psi' \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{pmatrix} \psi, \quad \Sigma_{\varepsilon u} = \gamma_{\varepsilon u}(0) + \sum_{h=1}^{\infty}(\gamma_{\varepsilon u}(h) + \gamma_{\varepsilon u}(h)'). \tag{21}$$

With similar expressions for the remaining Σ matrices.

Proof of Theorem 3

By combining Theorem 1 and Theorem 2 one can mimicking the proof of Lemma 13.1 of Johansen (1995) in order to establish the result.

Proof of Theorem 4

Before proving Theorem 4 we initially state and prove some auxiliary lemmas. For notational ease we adopt the convention $\varepsilon_t(\theta_0) = \varepsilon_t$ and likewise for other functions of the parameter vector evaluated in the true parameters. Furthermore, define $\varepsilon_t^d = \mathrm{diag}(\epsilon_t)$ and let $\mathbf{1}_{(d_1 \times d_2)}$ denote a d_1 times d_2 matrix of ones.

Lemma 1. *Under the assumptions of Theorem 4 it holds that*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial l_t(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, \Omega_S),$$

with $\Omega_S = E[(\partial l_t(\theta_0)/\partial\theta)(\partial l_t(\theta_0)/\partial\theta')] > 0$ *as* T *tends to infinity.*

Proof. Initially note that by utilizing the diagonal structure of D_t the first derivatives evaluated at the true parameters can be written as

$$\frac{\partial l_t}{\partial \theta^{(1)}} = -\frac{\partial \varepsilon_t'}{\partial \theta^{(1)}} H_t^{-1}\varepsilon_t - \frac{1}{2}\frac{\partial \Pi_t'}{\partial \theta^{(1)}} D_t^{-1}\xi_t \mathbf{1}_{(p\times 1)} \tag{22}$$

$$\frac{\partial l_t}{\partial \theta^{(2)}} = -\frac{1}{2}\frac{\partial \Pi_t'}{\partial \theta^{(2)}} D_t^{-1}\xi_t \mathbf{1}_{(p\times 1)} \tag{23}$$

$$\begin{aligned} \frac{\partial l_t}{\partial \theta^{(3)}} &= s_t \frac{\partial \mathrm{vec}(\Lambda_1)'}{\partial \theta^{(3)}} \mathrm{vec}(\Lambda_1^{-1} - \Lambda_1^{-1} D_t^{-1/2}\varepsilon_t\varepsilon_t' D_t^{-1/2}\Lambda_1^{-1}) \\ &\quad + (1-s_t)\frac{\partial \mathrm{vec}(\Lambda_0)'}{\partial \theta^{(3)}} \mathrm{vec}(\Lambda_0^{-1} - \Lambda_0^{-1} D_t^{-1/2}\varepsilon_t\varepsilon_t' D_t^{-1/2}\Lambda_0^{-1}) \end{aligned} \tag{24}$$

$$\begin{aligned} \frac{\partial \Pi_t'}{\partial \theta^{(1)}} &= 2\sum_{j=1}^{q} \frac{\partial \varepsilon_{t-j}'}{\partial \theta^{(1)}} A_j \varepsilon_{t-j}^d, \quad A_j = \mathrm{diag}(\alpha_{1,j}, ..., \alpha_{p,j}) \\ \frac{\partial \varepsilon_t'}{\partial \theta^{(1)}} &= s_t J_1 z_t + (1-s_t) J_2 z_t \\ \xi_t &= I_p - \varepsilon_t^d H_t^{-1} \varepsilon_t^d, \end{aligned}$$

where J_1 is a $2p(r + p(q-1))$ times $r + p(q-1)$ matrix with all ones on the first $p(r + p(q-1))$ rows and zeros on the remaining rows and the matrix J_2 the opposite. Finally, the derivative of Π_t' with respect to $\theta^{(2)}$ is a block diagonal matrix with the vectors $(\varepsilon_{1,t-1}, ..., \varepsilon_{1,t-q})'$ to $(\varepsilon_{p,t-1}, ..., \varepsilon_{p,t-q})'$ on the diagonal blocks.

Next, note that since the ARCH parameters are all bounded away from zero there exists a constant $k_1 > 0$ such that $\varepsilon_{i,t-j}^2/\pi_{i,t} \leq 1/\alpha_{i,j} < k_1$ for all $j = 1, ..., q$. By repeating this argument one can conclude that there exists a constant k_2 such that $\|(\partial \Pi_t'/\partial\theta D_t^{-1}\| < k_2$. Combining this with the observations that $E[\|\xi_t\|^2] < \infty$ and $E[\|\nu_t\|^4] < \infty$ yields that $E[\|\partial l_t/\partial\theta\|^2] < \infty$ and $\Omega < \infty$.

For any vector c with same dimension as θ define the sequence $l_t^{(1)} = c'\partial l_t/\partial\theta c$, which is a martingale difference sequence with respect to the natural filtration $\mathbb{F}_t = \sigma(X_t, X_{t-1}, ...)$ since $E[\xi_t \mid \mathbb{F}_{t-1}] = 0$ and s_t is $\mathbb{F}_{t-1}$ measurable. Under the stated conditions Theorem 1, the law of large number for geometrically ergodic time series, and the central limit of Brown (1971) yield that $T^{-1/2}\sum_{t=1}^{T} l_t^{(1)} \xrightarrow{D} N(0, c'\Omega_S c)$ and the Cramér-Wold device establishes the lemma.

The positive definiteness of Ω_S can be established by noting that Theorem 1 guarantees that $P(s_t = 1) > 0$, $P(s_t = 0) > 0$, and that all elements of Y_t have strictly positive densities. Hence is holds that the $c'\Omega c = 0$ if and only if $c = 0$ and Ω_S is therefore positive definite, see Lange et al. (2007) for details. $\square$

Lemma 2. *Under the conditions of Theorem 4 there exists an open neighborhood around the true parameter value $N(\theta_0)$ and a positive constant k_3, such that*

$$\sup_{\theta \in N(\theta_0)} \| \frac{\Pi_t(\theta)'}{\partial \theta^{(1)}} D_t^{-1} \|_{max} < k_3, \quad \sup_{\theta \in N(\theta_0)} \| \frac{\Pi_t(\theta)'}{\partial \theta^{(2)}} D_t^{-1} \|_{max} < k_3,$$

and $\|{\Lambda^{(l)}}^{-1}\|_{max} < k_3$ for $l = 0, 1$, where $\| \cdot \|_{max}$ denotes the max norm.

Proof. Let $N(\theta_0) = \{\theta \in \mathbb{R}^{\dim(\theta_0)} \mid \|\theta_0 - \theta\|_{\max} < \delta\}$. Next, note that by construction any term in $\frac{\partial l_t(\theta)}{\partial \theta^{(1)}}$ will also be in the relevant part of D_t, hence it can be concluded that if δ is sufficiently small there exists a positive constant k_3 such that

$$\sup_{\theta \in N(\theta_0)} \| \frac{\Pi_t(\theta)'}{\partial \theta^{(1)}} D_t^{-1} \|_{\max} \leq \frac{1}{\min_{\theta \in N(\theta_0)} \theta^{(2)}} < k_3$$

and likewise for the derivative with respect to $\theta^{(2)}$. Finally, note that since the true value of both $\Lambda^{(1)}$ and $\Lambda^{(0)}$ are positive definite and the eigenvalues of a matric is a continuous function of the matrix itself it holds that δ can be chosen such that $\|{\Lambda^{(l)}}^{-1}\|_{\max} < k_3$ for $l = 0, 1$.

$\square$

Proof of Theorem 4. The proof is based on a Taylor expansion of the log-likelihood function. To avoid the need for third derivatives we will verify conditions (A.1)-(A.4) of Lemma A.1 in Lange et al. (2007). The asymptotic normality of the score evaluated at the true parameter values has been established in Lemma 1, hence condition (A.1) is satisfied. By directly differentiating (22), (23), and (24) and adopting the notation of Lemma 1 one obtains the following expressions for the second derivatives.

$$\begin{aligned}
&\frac{\partial^2 l_t(\theta)}{\partial\theta^{(1)}\partial\theta^{(1)\prime}} \\
= \quad & -\frac{\partial\varepsilon_t(\theta)'}{\partial\theta^{(1)}}H_t(\theta)^{-1}\frac{\partial\varepsilon_t(\theta)}{\partial\theta^{(1)\prime}} \\
& -\frac{1}{4}\frac{\partial\Pi_t(\theta)'}{\partial\theta^{(1)}}D_t(\theta)^{-1}(\mathrm{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1}\varepsilon_t(\theta)^d\mathbf{1}_{(p\times1)}\} - \xi_t + I_p)D_t(\theta)^{-1}\frac{\partial\Pi_t(\theta)}{\partial\theta^{(1)\prime}} \\
& +\frac{\partial\varepsilon_t(\theta)'}{\partial\theta^{(1)}}H_t(\theta)^{-1}\varepsilon_t(\theta)^d D_t(\theta)^{-1}\frac{\partial\Pi_t(\theta)}{\partial\theta^{(1)\prime}} - \frac{1}{2}\frac{\partial\Pi_t(\theta)'}{\partial\theta^{(1)}}D_t(\theta)^{-2}\mathrm{diag}\{\xi_t(\theta)\mathbf{1}_{(p\times1)}\}\frac{\partial\Pi_t(\theta)}{\partial\theta^{(1)\prime}} \\
& -\sum_{j=1}^{q}\frac{\partial\varepsilon_{t-j}(\theta)'}{\partial\theta^{(1)}}A_j D_t(\theta)^{-1}\mathrm{diag}\{\xi_t(\theta)\mathbf{1}_{(p\times1)}\}\frac{\partial\varepsilon_{t-j}(\theta)}{\partial\theta^{(1)\prime}} \\
& +\frac{1}{2}\frac{\partial\Pi_t(\theta)'}{\partial\theta^{(1)}}D_t(\theta)^{-1}(\mathrm{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1}\mathbf{1}_{(p\times1)}\} + \varepsilon_t^d H_t^{-1})\frac{\partial\varepsilon_t(\theta)}{\partial\theta^{(1)\prime}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta^{(1)}\partial\theta^{(2)\prime}} \quad = \quad & \frac{\partial\varepsilon_t(\theta)'}{\partial\theta^{(1)}}H_t(\theta)^{-1}\varepsilon_t(\theta)^d D_t(\theta)^{-1}\frac{\partial\Pi_t(\theta)}{\partial\theta^{(2)\prime}} \\
& +\frac{1}{4}\frac{\partial\Pi_t(\theta)'}{\partial\theta^{(1)}}D_t^{-1}[2\mathrm{diag}\{\xi_t(\theta)\mathbf{1}_{(p\times1)}\} - \mathrm{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1}\varepsilon_t(\theta)^d\mathbf{1}_{(p\times1)}\} \\
& \qquad -\varepsilon_t(\theta)^d H_t(\theta)^{-1}\varepsilon_t(\theta)^d]D_t^{-1}\frac{\partial\Pi_t(\theta)}{\partial\theta^{(2)\prime}} \\
& -\sum_{j=1}^{q}\frac{\partial\varepsilon_{t-j}(\theta)'}{\partial\theta^{(1)}}\varepsilon_{t-j}^d D_t(\theta)^{-1}\xi_t(\theta)^d\frac{\partial(A_j\mathbf{1}_{(p\times1)})}{\partial\theta^{(2)\prime}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta^{(1)}\partial\theta^{(3)\prime}} \quad = \quad & s_t\frac{\partial\varepsilon_t(\theta)'}{\partial\theta^{(1)}}\big((\varepsilon_t(\theta)'D_t^{-1/2}\Lambda_1^{-1})\otimes(\Lambda_1^{-1}D_t^{-1/2})\big)\frac{\partial\mathrm{vec}(\Lambda_1)}{\partial\theta^{(3)\prime}} \\
& +(1-s_t)\frac{\partial\varepsilon_t(\theta)'}{\partial\theta^{(1)}}\big((\varepsilon_t(\theta)'D_t^{-1/2}\Lambda_0^{-1})\otimes(\Lambda_0^{-1}D_t^{-1/2})\big)\frac{\partial\mathrm{vec}(\Lambda_0)}{\partial\theta^{(3)\prime}} \\
& -\frac{1}{2}\frac{\partial\Pi_t(\theta)'}{\partial\theta^{(1)}}D_t^{-1}\frac{\partial(\xi_t(\theta)\mathbf{1}_{(p\times1)})}{\partial\theta^{(3)\prime}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta^{(2)}\partial\theta^{(2)\prime}} \quad = \quad & \frac{1}{4}\frac{\partial\Pi_t(\theta)'}{\partial\theta^{(2)}}D_t(\theta)^{-1}[2\mathrm{diag}\{\xi_t(\theta)\mathbf{1}_{(p\times1)}\} \\
& \qquad -\mathrm{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1}\varepsilon_t(\theta)^d\mathbf{1}_{(p\times1)}\} + \xi_t - I_p]D_t(\theta)^{-1}\frac{\partial\Pi_t(\theta)}{\partial\theta^{(2)\prime}}
\end{aligned}$$

$$\frac{\partial^2 l_t(\theta)}{\partial\theta^{(2)}\partial\theta^{(3)\prime}} = -\frac{1}{2}\frac{\partial\Pi_t(\theta)'}{\partial\theta^{(2)}}D_t(\theta)^{-1}\frac{\partial(\xi_t(\theta)\mathbf{1}_{(p\times1)})}{\partial\theta^{(3)\prime}}$$

$$\begin{aligned}\frac{\partial^2 l_t(\theta)}{\partial\theta^{(3)}\partial\theta^{(3)\prime}} = {} & s_t\frac{1}{2}\frac{\partial\mathrm{vec}(\Lambda_1)'}{\partial\theta^{(3)}}(\Lambda_1^{-1}\otimes I_p)\big[-(\Lambda_1^{-1}D_t(\theta)^{-1/2}\varepsilon_t(\theta)\varepsilon_t(\theta)'D_t(\theta)^{-1/2})\otimes I_p \\ & -I_p\otimes(\Lambda_1^{-1}D_t(\theta)^{-1/2}\varepsilon_t(\theta)\varepsilon_t(\theta)'D_t(\theta)^{-1/2})-I_p\big](I_p\otimes\Lambda_1^{-1})\frac{\partial\mathrm{vec}(\Lambda_1)}{\partial\theta^{(3)\prime}} \\ & (1-s_t)\frac{1}{2}\frac{\partial\mathrm{vec}(\Lambda_0)'}{\partial\theta^{(3)}}(\Lambda_0^{-1}\otimes I_p)\big[-(\Lambda_0^{-1}D_t(\theta)^{-1/2}\varepsilon_t(\theta)\varepsilon_t(\theta)'D_t(\theta)^{-1/2})\otimes I_p \\ & -I_p\otimes(\Lambda_0^{-1}D_t(\theta)^{-1/2}\varepsilon_t(\theta)\varepsilon_t(\theta)'D_t(\theta)^{-1/2})+I_p\big](I_p\otimes\Lambda_0^{-1})\frac{\partial\mathrm{vec}(\Lambda_0)}{\partial\theta^{(3)\prime}}\end{aligned}$$

$$\begin{aligned}\frac{\partial(\xi_t(\theta)\mathbf{1}_{(p\times1)})}{\partial\theta^{(3)\prime}} = {} & s_t\varepsilon_t(\theta)^d\big((\varepsilon_t(\theta)D_t(\theta)^{-1/2}\Lambda_1^{-1})\otimes(\Lambda_1^{-1}D_t^{-1/2})\big)\frac{\partial\mathrm{vec}(\Lambda_1)}{\partial\theta^{(3)\prime}} \\ & +(1-s_t)\varepsilon_t(\theta)^d\big((\varepsilon_t(\theta)D_t(\theta)^{-1/2}\Lambda_0^{-1})\otimes(\Lambda_0^{-1}D_t^{-1/2})\big)\frac{\partial\mathrm{vec}(\Lambda_0)}{\partial\theta^{(3)\prime}}\end{aligned}$$

By combing Theorem 1 with the law of large numbers for geometrically ergodic time series and Lemma 2 it can be concluded that the

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\xrightarrow{P}\Omega_I,$$

as T tends to infinity. By the same arguments as in the proof of Lemma 1 Ω_I is positive definite. Hence condition (A.2) of Lemma A.1 in Lange et al. (2007) is satisfied.

Next, let $N(\theta_0)=\{\theta\in\mathbb{R}^{\dim(\theta_0)}\mid\|\theta_0-\theta\|_{\max}<\delta\}$ denote an open neighborhood around the true parameter value. By inspecting the second derivatives and utilizing Lemma 2 it is evident that $\delta>0$ can be chosen such that there exists a positive constant k_4 and vector k of positive constants such that

$$E[\sup_{\theta\in N(\theta_0)}\Big\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\Big\|]\le E[k_4(1+\|\nu_t\nu_t'\nu_t\nu_t'\|_{\max})k'(z_t',...,z_{t-q}')'],\tag{25}$$

which is finite by assumption. One can therefore define a function F as the point-by-point (in θ) limit of $T^{-1}\sum_{t=1}^{T}\partial^2 l_t(\theta)/\partial\theta\partial\theta'$ and by dominated convergence the function is continuous. Hence condition (A.3) is also satisfied.

Finally, by Theorem 4.2.1 of Amemiya (1985) the required uniform convergence follows from (25) and condition (A.4) is therefore satisfied. Note that Theorem 4.2.1 is applicable in our setup by Amemiya (1985) p. 117 as the law of large numbers applies due to geometric ergodicity of the Markov chain z_t. This completes the proof. □

Limiting behavior of the heteroskedastic robust Wald test when the underlying innovations have heavy tails

By Theis Lange

Department of Mathematical Sciences, University of Copenhagen

E-mail: *lange@math.ku.dk*

Abstract: This paper initially establishes that the usual OLS estimator of the autoregressive parameter in the first order stable autoregressive model with autoregressive conditional heteroskedastic errors, the AR-ARCH model, has a non-standard limiting distribution with a non-standard rate of convergence when the innovations have non-finite fourth order moment. Furthermore, it is shown that the robust t- and Wald test statistics of White (1980) are still consistent and have the usual rate of convergence, but a non-standard limiting distribution when the innovations have non-finite fourth order moment. The critical values for the non-standard limiting distribution are higher than the usual N(0,1) and χ_1^2 critical values, respectively, which implies that an acceptance of a hypothesis using the standard robust t- or Wald tests remains valid even if the fourth order moment condition is not met. However, the size of the test might be higher than the nominal size. Hence the analysis presented in this paper extends the usability of the robust t- and Wald tests of White (1980). Finally, a small empirical study illustrates the results.

Keywords: ARCH; Robust t- and Wald tests; Heavy tails.

1 Introduction

Given a process $(y_t)_{t=1}^T$ this paper studies the OLS estimator from the regression of y_t on y_{t-1} when the process is assumed to be generated by a stable autoregressive model with autoregressive conditional heteroskedastic errors, the AR-ARCH model. By now the presence of ARCH type effects in financial and macro economic time series is a well established fact. The seminal paper by Engle (1982) in which the linear ARCH model model was originally introduced has been followed by countless papers studying various aspects ARCH type models.

Recognizing that unmodeled heteroscedasticity in the innovations might seriously compromise the validity of traditional t- and Wald tests of the significance of parameters estimated by OLS White (1980) introduced the heteroscedastic robust t- and Wald tests. These tests are now so widely applied that they are routinely reported by many statistical software packages. However, Whites results depend on the innovations to have finite fourth order moment, which is often not met in empirical studies.

In this paper we show that the robust t- and Wald test statistics have the correct normalization, but a non-standard limiting distribution when the innovations have non-finite fourth order moment. The critical values for the non-standard limiting distributions are higher than the usual N(0,1) and χ_1^2 critical values, respectively, which implies that an acceptance of a hypothesis using the standard robust t- or Wald test procedures remains valid even if the fourth order moment condition is not met. However, the size of the test might be higher than the nominal size. Hence the analysis presented in this paper extends the usability of the robust t- and Wald tests of White (1980), which, to our knowledge, has not previously been done in the literature. In addition the paper establishes that the OLS estimator of the autoregressive parameter will have a stable limit with a non-standard rate of convergence. As the tools for handling stable distributions are less evolved than similar tools for normal distributions we are forced to restrict attention to a fairly simple first order model as the true data generating mechanism. Finally, a small empirical study shows how the evidence of no correlation between consecutive movements of interest rates critically depends on our extension of the usability of the robust Wald test.

The paper proceeds as follows. In Section 2 the model and some important properties including geometric ergodicity and tail heaviness are discussed. Section 3 presents the limiting distributions for the OLS estimator and Section 4 states the limiting distributions for the robust t- and Wald test statistics and discusses implications of the results on the standard testing procedures. Finally, Section 5 contains a small empirical study and Section 6 concludes. All proofs are contained in the Appendix.

2 The AR-ARCH Model

The model can be stated as

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \varepsilon_t(\theta), && (1)\\
\varepsilon_t(\theta) &= \sqrt{h_t(\theta)} z_t && (2)\\
h_t(\theta) &= \omega + \alpha \varepsilon_{t-1}^2(\theta) && (3)
\end{aligned}
$$

with $t = 1, ..., T$ and z_t an i.i.d.(0,1) sequence of random variables. The parameter vector is denoted $\theta = (\rho, \alpha, \omega)'$ and the true parameter θ_0. In order to ease notation we adopt the convention $\varepsilon_t = \varepsilon_t(\theta_0)$ etc. for expressions evaluated at the true parameter values. The analysis is conditional on the initial values y_0 and ε_{-1}.

In the context of the AR-ARCH model heavy tails can be introduced either by choosing the value of the ARCH parameter α sufficiently large while keeping the underlying error process z_t light tailed or through the tails of the underlying error process z_t. In this paper the first approach will be explored. The second approach has been investigated in e.g. Davis & Mikosch (1998).

For a fixed value of the ARCH parameter α the tail index, denoted λ, can be found as the unique strictly positive solution to the equation $E[(\alpha z_t^2)^{\lambda/2}] = 1$ as shown in Davis & Mikosch (1998) p. 2062. Note that a tail index of λ has the implication that the ARCH process has finite moments of all orders below λ, but $E[|\varepsilon_t|^\lambda] = \infty$. Figure 1 depicts the correspondence between α and λ when z_t is assumed Gaussian. Using the moment interpretation of the tail index and Figure 1 it is evident that the ARCH process has finite fourth order moment, but non-finite second order moment if the ARCH parameter belongs to the interval $]0.57, 1]$. This part of the parameter space will be the focus for much of the rest of the paper.

The following lemma, which has been proved in Lange, Rahbek & Jensen (2007), establishes minimal conditions under which processes generated by the AR-ARCH model are geometrically ergodic. Geometric ergodicity, and the laws of large numbers implied by this concept, constitutes an important tool when establish-

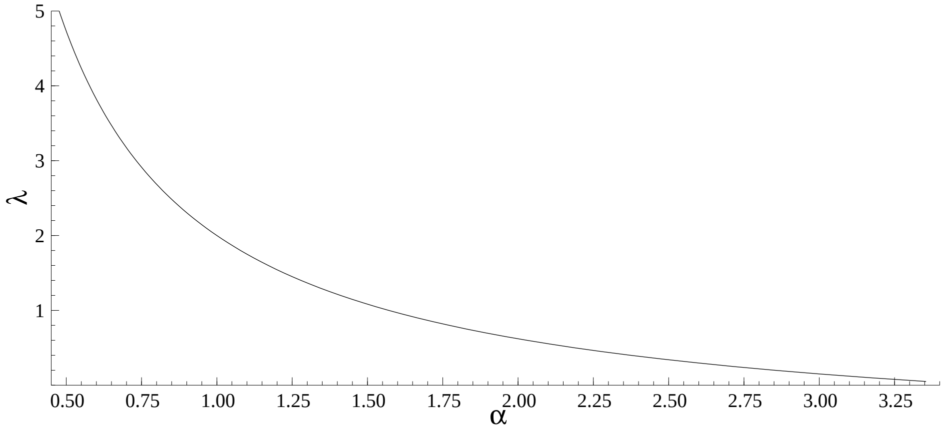

Figure 1: Correspondence between the ARCH parameter α and the tail index λ when z_t is assumed Gaussian.

ing asymptotic theory, but for our intended applications it is of equal importance that the lemma establishes minimal conditions under which there exists an initial distribution such that the process is stationary.

Lemma 1. *Assume that z_t has a density f with respect to the Lebesgue measure on $\mathbb{R}$, which is bounded away from zero on compact sets and furthermore that*

$$E[\log(\alpha_0 z_t^2)] < 0 \text{ and } |\rho_0| < 1$$

then the process $x_t = (y_{t-1}, \varepsilon_t)'$ generated by the AR-ARCH model, is geometrically ergodic. In particular there exists a stationary version and moreover if $E|g(x_t, ..., x_{t+k})| < \infty$, where expectation is taken with respect to the invariant distribution, the Law of Large Numbers given by

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} g(x_t, ..., x_{t+k}) \overset{a.s.}{=} E[g(x_t, ..., x_{t+k})], \tag{4}$$

holds irrespectively of the choice of initial distribution.

3 Limiting behavior of the OLS estimator

In this section the limiting behavior of the OLS estimator of the autoregressive parameter ρ is derived under two sets of moment assumptions on the innovation sequence. Initially the section reviews the well known result that the OLS estimator is asymptotically normal when the innovations have finite fourth order moment and states precise expressions for the parameters of the limiting distribution. In the second part of this section the limiting behavior of the OLS estimator when the innovations only have finite second order moment, is derived. To our knowledge this result has not previously been established in the literature. The final part of this section discusses some implications of the limiting results.

3.1 The normal case: Finite fourth order moment

Standard techniques combined with Lemma 1 give the following result regarding the OLS estimator of ρ when the innovations have finite fourth order moment.

Theorem 1. *In addition to the assumptions in Lemma 1 assume that*

(*i*) $E[\alpha_0^2 z_t^4] = \alpha_0^2 \kappa < 1$

(*ii*) *and z_t's distribution is symmetric.*

Then the OLS estimator of the autoregressive parameter ρ in the AR-ARCH model given by (1) - (3) is consistent with the following limiting distribution

$$\sqrt{T}(\hat{\rho}_{OLS} - \rho_0) = \frac{T^{-1/2} \sum_{t=1}^{T} y_{t-1} \varepsilon_t}{T^{-1} \sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{D} N(0, \Sigma), \tag{5}$$

where

$$\Sigma = (1 - \rho_0^2) + \frac{(\kappa - 1)(1 - \rho_0^2)^2 \alpha_0}{(1 - \kappa \alpha_0^2)(1 - \alpha_0 \rho_0^2)}.$$

Remark 1. Condition (i) implies that ε_t has finite fourth order moment under the invariant measure and that z_t has finite fourth order moment.

The proof can be found in the appendix. As the parameter values approaches the value for which the ARCH process no longer have finite fourth order moment

($\alpha_0^2\kappa = 1$) the asymptotic variance Σ converges toward infinity. One could therefore conjecture that in this case the limiting distribution would be a stable law with a slower than square root T rate of convergence. In the following section we will prove this conjecture.

3.2 The stable case: Non-finite fourth order moment

Next, we will analyze the effect of relaxing the fourth order moment condition of Theorem 1 to a second order condition. As conjectured this leads to both a non-standard limiting distribution as well as non-standard rate of convergence. Since the tools for manipulating stable laws are somewhat less evolved than similar tools for normal distributions it is necessary to assume stationarity of the process as geometric ergodicity does not suffice in the present version of the proof, which can be found in the appendix.

Theorem 2. *In addition to the assumptions in Lemma 1 assume that*

(*i*) *the initial values are distributed according to the stationary distribution,*

(*ii*) *the ARCH parameter α_0 is such that the ARCH process has finite second order moment, but non-finite fourth order moment; that is the tail index λ belongs to the interval* $]2,4]$*,*

(*iii*) *and z_t's distribution is symmetric*

then it holds that

$$T^{1-2/\lambda}(\hat{\rho}_{OLS} - \rho_0) \xrightarrow{D} S_0, \tag{6}$$

where S_0 is a $\lambda/2$ stable random variable, with the remaining parameters unknown.

Remark 2. The existence of a stationary distribution is guaranteed by Lemma 1.

3.3 Implications

As Theorem 1 is a standard result this section will only address the implications of Theorem 2, which has the most direct implications for the construction of confidence bands. When constructing confidence bands one needs to know both the asymptotical distribution (usual the normal distribution) including parameter estimates as well as the rate of convergence (usual square root T). However, if the true α_0 is such that the innovation sequence ε_t does not have finite fourth order moment the result implies that the rate of convergence will be non-standard and unknown and the parameters of the limiting distribution unknown as well. Since the rate of convergence in Theorem 2 is slower than the usual square root T the usually constructed confidence bands, based on normality and standard rate of convergence, might be way to narrow leading to erroneous conclusions. Unfortunately, since λ in unknown in practice and Theorem 2 does not include a precise specification of the remaining parameters of the limiting stable distribution, one cannot easily derive a corrected confidence band based on the theorem. In the next section we will address this problem by considering the heteroskedastic robust$t$- and Wald tests of White (1980).

4 Limiting behavior of the robust Wald test

In this section we will examine the behavior of the heteroskedastic robust t- and Wald tests of White (1980). The robust t-test statistic for the hypothesis $H_0 : \rho_0 = 0$ is given by

$$V_T = \sqrt{T}(\hat{\rho}_{OLS} - \rho_0)\big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 y_{t-1}^2\big)^{-1/2}\big(\frac{1}{T}\sum_{t=1}^{T}y_{t-1}^2\big),$$

and the robust Wald test statistic is V_T^2. Under the hypothesis V_T can be rewritten as

$$V_T = \big(\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}\big)\big(\sum_{t=1}^{T}\varepsilon_t^2\varepsilon_{t-1}^2\big)^{-1/2},$$

which is sometimes refereed to as a *self normalizing sum.* If the innovations ε_t have finite fourth order moment it is well known, see e.g. White (1980), that V_T converges to standard normal distribution under H_0 as T tends to infinity irrespectively of possible heteroscedasticity of the innovations. This result forms the basis for the usual robust t- and Wald test. However, if the the innovations do not have finite fourth order moment the limiting behavior of V_T has not been examined in the litterateur. From the theory for self normalizing sums of i.i.d. random variables, see Giné, Götze & Mason (1997), it is known that a self normalized sum of i.i.d. random variables converges to a normal distribution if and only if the numerator belongs to the domain of attraction for a normal distribution (loosely speaking this corresponds to requiring that the Lindeberg condition holds for the numerator), however, since the sequence $(\varepsilon_t)_{t=1}^T$ is not i.i.d. the results of Giné et al. (1997) are not applicable in our setup. Based on this result one would still conjecture that V_T does not have a Gaussian limiting distribution since Theorem 2 establishes that the numerator belongs to the domain of attraction of a stable law. Theorem 3 below formalizes this conjecture and the proof can be found in the appendix.

Theorem 3. *In addition to the assumptions in Lemma 1 assume that*

- *(i)* *the initial values are distributed according to the stationary distribution,*
- *(ii)* *the ARCH parameter α_0 is such that the ARCH process has finite second order moment, but non-finite fourth order moment; that is the tail index λ belongs to the interval $]2,4]$,*
- *(iii)* *and the distribution of z_t is symmetric*

then under H_0 it holds that

$$V_T \xrightarrow{D} \frac{S_1}{S_2^{1/2}} \tag{7}$$

where the vector $(S_1, S_2^{1/2})$ is jointly a $\lambda/2$ stable random variable, with the remaining parameters and dependence structure unknown.

Remark 3. A natural question is what happens when $\lambda \in]0,2]$, corresponding to $E[\varepsilon_t^2] = \infty$, but in this case Theorem 2 indicates that $\hat{\rho}_{OLS}$ is not even consistent rendering a test for a particular value fruitless.

Corollary 1. *Under the conditions of Theorem 3 the heteroskedastic robust Wald test given by V_T^2 converges in distribution to S_1^2/S_2 where the vector (S_1^2, S_2) is jointly a $\lambda/4$ stable random variable, with the remaining parameters and dependence structure unknown.*

In the following section the implications of Theorem 3 on the usual testing procedure will be discussed.

4.1 Implications for the standard testing procedure

Arguably the most important implication of Theorem 3 and Corollary 1 is that the normalization required to ensure a non-degenerate limiting distribution for both the robust t- and Wald tests does not depend on the fourth order moment being finite. However, since the limiting distribution is no longer Gaussian when the innovations have non-finite fourth order moment the critical values are not the usual ones.

In the usual Gaussian case, obtained when the innovations have finite fourth order moment, it can be directly established by utilizing the properties of the normal distribution that the limiting distribution for the robust t- and Wald test statistics are nuisance parameter free (indeed even in higher order models than the one considered in this paper). In contrast to this, it is not possible to verify that the stable limits in Theorem 3 and Corollary 1 do not depend on additional parameters besides the tail index, since a precise mathematical expression for parameter values and dependence structure is not available. However, in the first order AR-ARCH model considered in this paper the only remaining unknown parameter is the scale parameter ω_0 and the test statistic V_T is clearly invariant to the scale of the innovations.

Since the scale parameter ω_0 does not affect the limiting distributions in Theorem 3 and Corollary 1 the critical values for the hypothesis H_0 will only depend on the ARCH parameter α_0 and perhaps the exact distribution of the innovations z_t. Based on simulations Figure 2 Panel A-C illustrate how the critical values change as the tail index λ is decreased. Panel D shows the correspondence between the tail parameter and the ARCH parameter for the different choices of the distribution for z_t.

From Figure 2 it is evident that the critical values for the non-standard limiting distribution are higher than usual χ^2_1 critical values and increase as the tail index decreases (corresponding to increasing the ARCH parameter). In addition it is seen that the distribution of the underlying innovations z_t only affects the limiting distribution through the tail index. This implies that an acceptance of a hypothesis using the standard robust Wald test procedure remains valid even if the fourth order moment condition is not met. However, the size of the test

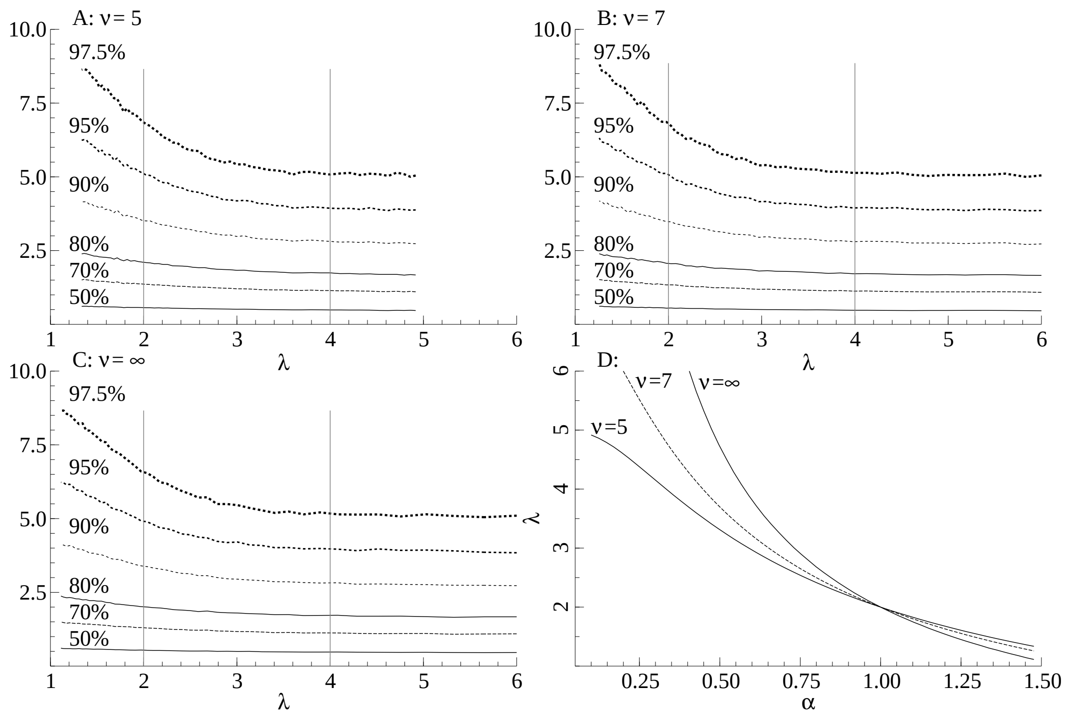

Figure 2: Quantiles for the limiting distribution for the robust Wald test statistic (S_1^2/S_2 from Corollary 1) computed by simulating from the AR-ARCH model with $\rho_0 = 0$, $\omega_0 = 1$ for a range of values for α_0 (to ease comparison the critical values are reported as functions of the tail index). In Panel A and B the innovations z_t follow a standardized student-t distribution with 5 and 7 degrees of freedom, respectively, while $z_t \sim N(0, 1)$ in Panel C. Each simulated path was 5,000 data points long and 100,000 Monte Carlo replications were conducted for each value of α_0. The two vertical lines corresponds to the values of the tail index where the process does no longer have finite fourth order and second order moment, respectively. Finally, Panel D shows the correspondence between the ARCH parameter and the tail index for the different distributions.

might be higher than the nominal size. Furthermore, Theorem 2 implies that the robust $t-$ and Wald tests are still consistent as long as the innovations have finite second order moment. Hence the analysis of this paper extends the usability of the robust $t-$ and Wald tests of White (1980).

5 Empirical illustration

In this section we will reexamine the evidence of linear predictability in the daily movements of interest rates. The data set consists of daily recordings of the 3-months US t-bill rate (r_t) covering the period from the 2nd of January 1990 to the 29th of February 2008, yielding a total of 4,544 observations, see Figure 3.

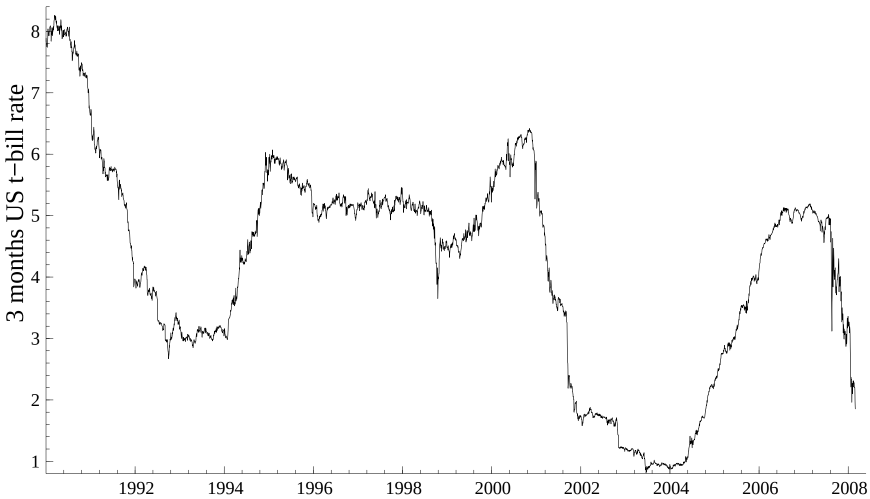

Figure 3: Daily recordings of the 3-months US t-bill rate.

To examine whether past interest rate movements are correlated with future interest rate movements we will test if the coefficient in the regression of $x_t = r_t - r_{t-1}$ on x_{t-1} is statistically significantly different from zero. As it is well documented that daily interest rates exhibit heteroscedasticity we will conduct both the usual Wald test as well as the robust Wald test of White (1980). Finally we will employ the full AR-ARCH model to estimate the magnitude of the ARCH effect by quasi maximum likelihood and thereby assess the potential size distortion of the robust Wald test caused by heavy tails.

	Wald test	Robust Wald test
Test statistic	30.27	3.44
p-value	0.000	0.065
Corrected p-value†	-	0.126

Table 1: Summary of test results for the hypothesis of a zero coefficient in the regression of x_t on x_{t-1}. The unrestricted OLS estimator is 0.081. †Computed using the estimated ARCH coefficient of 0.91 and Student-t degree of freedom of 2.45 from the full AR-ARCH model and the non-standard distribution from Corollary 1.

Based on the the test statistics and p-values presented in Table 1 it is evident that if the heteroscedasticity of the errors is ignored one would reject the hypothesis that the coefficient in the regression is zero. If the test is instead based on the robust Wald test statistic compared to the χ_1^2 distribution the hypothesis is accepted with a p-value of 0.065. Estimating the full AR-ARCH model with standardized Student-t innovations with ν degrees of freedom by quasi maximum likelihood provides the estimates $\hat{\alpha} = 0.91$ and $\hat{\nu} = 2.45$, corresponding to a tail index of 2.2. Hence the conditions for employing the robust Wald test of White (1980) are not met. However, by employing the non-standard limiting distribution from Corollary 1 the p-value increases to 0.126. Thus taking the magnitude of the ARCH effect into account reveals that the robust Wald test is somewhat size distorted, but as previously discussed this only strengthens the conclusion of no linear predictability.

It should be stressed that we do not suggest that practitioners do full quasi maximum likelihood estimation just to correct their robust Wald test inference. The purpose of this section is merely to illustrate the necessity of the extended usability of the robust tests and quantify the potential size distortion.

6 Conclusion

In this paper we have established that the usual OLS estimator of the autoregressive parameter in the AR-ARCH model has a non-standard limiting distribution with a non-standard rate of convergence if the innovation process is a realization of an ARCH(1) process with non-finite fourth order moment. Furthermore, we have established that the robust t- and Wald test statistics of White (1980) for the hypothesis $\rho_0 = 0$ have the correct normalization, but a non-standard limiting distribution when the innovations have non-finite fourth order moment. The critical values for the non-standard limiting distribution are higher than the usual N(0,1) and χ^2_1 critical values, respectively, which implies that an acceptance of the hypothesis using the standard robust t- or Wald tests remains valid even in the fourth order moment condition is not met. However, the size of the tests might be higher than the nominal size. Hence the analysis presented in this paper extends the usability of the robust t- and Wald tests of White (1980). In Figure 2 the critical values are summarized. Finally, a small empirical study shows how the extended usability of the robust test is required to establish that consecutive movements of interest rates are not correlated. In addition the empirical study quantifies the potential size distortion caused by the heavy tails of the innovations.

Appendix

Proof of Theorem 1. Note initially that since we have assumed finite fourth order moment and hence also finite second order moment, y_t has the stationary representation $y_t^* = \sum_{t=0}^{\infty} \rho^i \varepsilon_{t-i}$, which will be used when calculating expected values under the stationary distribution. The second order moments of the ARCH process and the volatility process are given by $E[\varepsilon_t^2] = E[h_t] = \frac{\omega_0}{1-\alpha_0}$. Next the fourth order moment can be derived from

$$E[\varepsilon_t^4] = E[z_t^4 h_t^2] = \kappa(\omega_0^2 + \alpha_0^2 E[\varepsilon_{t-1}^4] + 2\omega_0\alpha_0 E[\varepsilon_{t-1}^2]).$$

Since the expectation is taken with respect to the stationary distribution it holds that

$$E[\varepsilon_t^4] = \kappa \frac{\omega_0^2 + 2\omega_0^2 \frac{\alpha_0}{1-\alpha_0}}{1-\kappa\alpha_0^2} = \kappa \frac{\omega_0^2(1+\alpha_0)}{(1-\kappa\alpha_0^2)(1-\alpha_0)},$$

and $E[h_t^2] = \omega_0^2(1+\alpha_0)(1-\kappa\alpha_0^2)^{-1}(1-\alpha_0)^{-1}$. Utilizing the representation for h_t as a function of $z_t, ..., z_{t-k}$ and h_{t-k} from Nelson (1990) it holds that for some

$k \in \mathbb{N}_0$

$$\begin{aligned}
E[\varepsilon_{t-k}^2 \varepsilon_t^2] &= E\left[\varepsilon_{t-k}^2 z_t^2 \left(h_{t-k}\prod_{i=1}^{k}\alpha_0 z_{t-i}^2 + \omega_0\left(1+\sum_{k=1}^{k-1}\prod_{i=1}^{k}\alpha_0 z_{t-i}^2\right)\right)\right] \\
&= \kappa E[\alpha_0^k h_{t-k}^2] + \omega_0 E\left[\varepsilon_{t-k}^2 \frac{1-\alpha_0^k}{1-\alpha_0}\right] \\
&= \kappa\alpha_0^k\omega_0^2 \frac{1+\alpha_0}{(1-\kappa\alpha_0^2)(1-\alpha_0)} + \omega_0^2\frac{1-\alpha_0^k}{(1-\alpha_0)^2} \\
&= \frac{\omega_0^2}{(1-\alpha_0)^2} + \frac{\omega_0^2\alpha_0^k(\kappa-1)}{(1-\kappa\alpha_0^2)(1-\alpha_0)^2}.
\end{aligned}$$

Using the symmetry of z_t's distribution and the infinite representation of y_t yields

$$\begin{aligned}
E[y_{t-1}^2 h_t] &= E\left[\left(\sum_{i=0}^{\infty}\rho_0^i \varepsilon_{t-1-i}\right)^2 (\omega_0 + \alpha_0\varepsilon_{t-1}^2)\right] \\
&= \omega_0\sum_{i=0}^{\infty}\rho_0^{2i}E[\varepsilon_{t-i-1}^2] + \alpha_0\sum_{i=0}^{\infty}\rho_0^{2i}E[\varepsilon_{t-h}^2\varepsilon_t^2] \\
&= \frac{\omega_0^2}{(1-\alpha_0)(1-\rho_0^2)} + \frac{\alpha_0\omega_0^2}{(1-\alpha_0)^2(1-\rho_0^2)} + \frac{\omega_0^2\alpha_0^2(\kappa-1)}{(1-\kappa\alpha_0^2)(1-\alpha_0)^2}\sum_{i=0}^{\infty}\alpha_0^i\rho_0^{2i} \\
&= \frac{\omega_0^2}{(1-\alpha_0^2)(1-\rho_0^2)} + \frac{\omega_0(\kappa-1)}{(1-\kappa\alpha_0^2)(1-\alpha_0)^2(1-\alpha_0\rho_0^2)},
\end{aligned}$$

and

$$E[y_{t-1}^2] = E\left[\left(\sum_{i=0}^{\infty}\rho_0^i\varepsilon_{t-1-i}\right)^2\right] = \frac{\omega_0}{(1-\alpha_0)(1-\rho_0^2)}.$$

Next, define the filtration $\mathbb{F}_t = \sigma(\varepsilon_t, y_t, ...)$. In order to apply a standard CLT for martingale difference sequences (e.g. Brown (1971)) we first verify the Lindeberg condition

$$\frac{1}{T}\sum_{t=1}^{T} E[y_{t-1}^2\varepsilon_t^2 1_{\{|y_{t-1}\varepsilon_t|>\delta\sqrt{T}\}} \mid \mathbb{F}_t] \leq \frac{k}{\delta^\xi T^{1+\xi/2}}\sum_{t=1}^{T}(y_{t-1}\sqrt{h_t})^{2+\xi} \to 0,$$

where k is a positive constant and $\xi > 0$ is chosen such that $E[(y_{t-1}\sqrt{h_t})^{2+\xi}]$ is finite. The constant ξ exists because the inequality which ensures finite fourth order moment is a sharp inequality, see Lange et al. (2007) for details. Furthermore

$$\frac{1}{T}\sum_{t=1}^{T} E[y_{t-1}^2\varepsilon_t^2 \mid \mathbb{F}_{t-1}] = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 h_t \xrightarrow{P} E[y_{t-1}^2 h_t].$$

Hence

$$\sqrt{T}(\hat{\rho}_{OLS} - \rho_0) \xrightarrow{D} N(0, \Sigma) \quad \text{as} \quad T \to \infty,$$

where

$$\Sigma = \frac{(1-\alpha_0)^2(1-\rho_0^2)^2}{\omega_0^2} E[y_{t-1}^2 h_t] = (1-\rho_0^2) + \frac{(\kappa-1)(1-\rho_0^2)^2\alpha_0}{(1-\kappa\alpha_0^2)(1-\alpha_0\rho_0^2)}.$$

This completes the proof. □

The proof of the Theorem 2 rests to a large extent on the following lemma.

Lemma 2. *Under the assumptions of Theorem 2 all finite dimensional vectors* $\mathbf{y}_t(k) = (y_t, ..., y_{t+k})$ *have regularly varying tails as defined in Resnick (1987) with the same tail index* λ *as the ARCH process.*

The proof is inspired by the proofs of Lemma A.3.26 in Embrechts, Klüppelberg & Mikosch (1997) and Lemma 4.24 in Resnick (1987). However none of these results are directly applicable since the innovations are not independent.

Proof of Lemma 2. We begin by showing a tamer result, namely that y_t is regularly varying with tail index λ. Since regular variation is a property of the marginal distribution, the subscript t on y_t will be omitted. In addition due to symmetry of the distribution of y_t and ε_t all arguments will be given using the absolute value of both only.

Since the ARCH process has finite second order moment y has the representation $y = \sum_{i=0}^{\infty} \rho^i \varepsilon_{-i}$. Define $y^{(m)} = \sum_{i=0}^{m-1} \rho^i \varepsilon_{-i}$ for any $m \geq 1$. We will now show that the remainder $y - y^{(m)}$ has negligible influence on the tails of y for m sufficiently large. Observe that for any $\delta \in]0,1[$ and $x > 0$ it holds that,

$$P(|y| > x) \leq P(|y^{(m)}| > (1-\delta)x) + P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > \delta x) \tag{8}$$

and

$$P(|y| > x) \geq P(|y^{(m)}| > (1+\delta)x) - P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > \delta x) \tag{9}$$

In the following we show

$$\lim_{m\to\infty} \limsup_{x\to\infty} \frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} = 0. \tag{10}$$

Rewrite the numerator as

$$\begin{aligned}
&P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right) \\
&\quad = P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x, \bigvee_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right) \\
&\qquad + P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x, \bigvee_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| \leq x\right) \\
&\quad \leq P\left(\bigcup_{i=m}^{\infty} \left(|\rho|^i |\varepsilon_{-i}| > x\right)\right) \\
&\qquad + P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| 1_{\{|\rho|^i |\varepsilon_{-i}| \leq x\}} > x, \bigvee_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| \leq x\right) \\
&\quad \leq \sum_{i=m}^{\infty} P(|\varepsilon_{-i}| > x|\rho|^{-i}) + P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| 1_{\{|\rho|^i |\varepsilon_{-i}| \leq x\}} > x\right).
\end{aligned}$$

Hence by Markov's inequality it holds that

$$\begin{aligned}
&\frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \\
&\quad \leq \sum_{i=m}^{\infty} \frac{P(|\varepsilon_0| > x|\rho|^{-i})}{P(|\varepsilon_0| > x)} + x^{-1} \sum_{i=m}^{\infty} \frac{|\rho|^i E[|\varepsilon_0| 1_{\{|\varepsilon_0| \leq x|\rho|^{-i}\}}]}{P(|\varepsilon_0| > x)} \\
&\quad = I + II.
\end{aligned} \tag{11}$$

By Basrak, Davis & Mikosch (2002b) the random variable ε_0 has regular varying tails and by Proposition 0.8(ii) of Resnick (1987) it holds that for all $\tau > 0$ there exists a x_0 such that for all $x > x_0$

$$P(|\varepsilon_0| > x|\rho|^{-i})/P(|\varepsilon_0| > x) \leq (1+\tau)|\rho|^{i(\lambda-\tau)}.$$

For τ adequately small this bound is summable and hence by dominated convergence and the regular variation of ε_0 it holds that

$$\limsup_{x\to\infty} I \leq (1+\tau)\sum_{i=m}^{\infty}|\rho|^{i(\lambda-\tau)}.$$

In considering II, suppose temporarily that $0 < \lambda < 1$ (this will never be the case when $E[\varepsilon_0^2] < \infty$, but it is a necessary steep towards proving the full result). From an integration by parts it holds that

$$\frac{E[|\varepsilon_0|1_{\{|\varepsilon_0|\leq x\}}]}{xP(|\varepsilon_0| > x)} \leq \frac{\int_0^x P(|\varepsilon_0| > u)du}{xP(|\varepsilon_0| > x)}$$

and applying Karamata's Theorem (from e.g. Resnick (1987)) this converges to $(1-\lambda)^{-1}$ as x tends to infinity. Thus the function $x \mapsto E[|\varepsilon_0|1_{\{|\varepsilon_0|\leq x|\rho|^{-i}\}}]$ is regular varying with tail index $1-\lambda$ and applying again Proposition 0.8(ii) we have that for any $\tau > 0$, some constant k, and x sufficiently large it holds that

$$\begin{aligned}\frac{|\rho|^i E[|\varepsilon_0|1_{\{|\varepsilon_0|\leq x|\rho|^{-i}\}}]}{xP(|\varepsilon_0| > x)} &= |\rho|^i\left(\frac{E[|\varepsilon_0|1_{\{|\varepsilon_0|\leq x|\rho|^{-i}\}}]}{E[|\varepsilon_0|1_{\{|\varepsilon_0|\leq x\}}]}\right)\frac{E[|\varepsilon_0|1_{\{|\varepsilon_0|\leq x\}}]}{xP(|\varepsilon_0| > x)} \\ &\leq |\rho|^i(|\rho|^{-i})^{-(1-\lambda+\tau)}k = k|\rho|^{i(\lambda-\tau)},\end{aligned}$$

which is summable for τ adequately small. So we conclude

$$\limsup_{x\to\infty} II \leq k\sum_{i=m}^{\infty}|\rho|^{i(\lambda-\tau)}$$

and hence when $0 < \lambda < 1$ there exists constants $\tilde{\tau} \in]0,1[$ and $\tilde{k} > 0$ such that

$$\limsup_{x\to\infty}\frac{P\left(\sum_{i=m}^{\infty}|\rho|^i|\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \leq \tilde{k}\sum_{i=m}^{\infty}|\rho|^{i\tilde{\tau}} < \infty. \tag{12}$$

If $\lambda \geq 1$ we get a similar inequality by reducing to the case $0 < \lambda < 1$ as follows. Pick $\eta \in]\lambda, \lambda\tilde{\tau}^{-1}[$ and set $c = \sum_{i=m}^{\infty} |\rho|^i$ and $p_i = |\rho|^i/c$ then by Jensen's inequality (e.g. Feller (1971) p. 153) we get

$$\begin{aligned}\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}|\right)^{\eta} &= c^{\eta} \left(\sum_{i=m}^{\infty} p_i |\varepsilon_{-i}|\right)^{\eta} \\ &\leq c^{\eta} \sum_{i=m}^{\infty} p_i |\varepsilon_{-i}|^{\eta} = c^{\eta-1} \sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}|^{\eta}.\end{aligned}$$

Thus

$$\frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \leq \frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}|^{\eta} > c^{1-\eta} x^{\eta}\right)}{P(|\varepsilon_0|^{\eta} > x^{\eta})}.$$

By Bingham, Goldie & Teugels (1987) Proposition 1.5.7(i) the function $P(|\varepsilon_0|^{\eta} > x^{\eta})$ is regularly varying with tail index $\eta^{-1}\lambda \in]0, 1[$. Hence (12) gives

$$\limsup_{x\to\infty} \frac{P\left(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_{-i}| > x\right)}{P(|\varepsilon_0| > x)} \leq \tilde{k} \sum_{i=m}^{\infty} |\rho|^{i\lambda\eta^{-1}} c^{\lambda(1-\eta^{-1})} < \infty. \tag{13}$$

This proves (10). Combine (8) and (9) with the above to obtain the relations

$$\begin{aligned}&\liminf_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\ &\geq \liminf_{x\to\infty} \frac{P(|y^{(m)}| > (1+\delta)x)}{P(|\varepsilon_0| > x)} - \limsup_{x\to\infty} \frac{P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_0| > \delta x)}{P(|\varepsilon_0| > x)} \\ &\to \liminf_{x\to\infty} \frac{P(|y^{(m)}| > (1+\delta)x)}{P(|\varepsilon_0| > x)} \quad \text{as } m \to \infty\end{aligned}$$

and

$$\begin{aligned}&\limsup_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\ &\leq \limsup_{x\to\infty} \frac{P(|y^{(m)}| > (1-\delta)x)}{P(|\varepsilon_0| > x)} + \limsup_{x\to\infty} \frac{P(\sum_{i=m}^{\infty} |\rho|^i |\varepsilon_0| > \delta x)}{P(|\varepsilon_0| > x)} \\ &\to \limsup_{x\to\infty} \frac{P(|y^{(m)}| > (1+\delta)x)}{P(|\varepsilon_0| > x)} \quad \text{as } m \to \infty\end{aligned}$$

By Basrak et al. (2002b) equation (2.6) $y^{(m)}$ is regular varying with tail index λ and hence will

$$\lim_{x\to\infty} \frac{P(|y^{(m)}| > x)}{P(|\varepsilon_0| > x)} = c_m$$

for a sequence of constants c_m. Using the same type of arguments as for $y - y^{(m)}$ one can conclude that c_m tends to a finite limit c as m tends to infinity. Hence it can be concluded that

$$\begin{aligned}
(c-\delta)(1+\delta)^{-\lambda} &\leq \liminf_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\
&\leq \limsup_{x\to\infty} \frac{P(|y| > x)}{P(|\varepsilon_0| > x)} \\
&\leq (c+\delta)(1-\delta)^{-\lambda}.
\end{aligned}$$

Now by letting δ go towards zero it can be concluded that y is regular varying with index λ.

Finally we wish to extent this result to all vectors of the form $\mathbf{y}(k) = (y_0, ..., y_k)$. By Basrak, Davis & Mikosch (2002a) Theorem 1.1(ii) it suffices to show that all linear combinations $v \in \mathbb{R}^k \setminus \{0\}$ are regular varying. However, for all v

$$v'\mathbf{y}(k) = \sum_{i=0}^{\infty} c_i \varepsilon_{-i},$$

where the coefficients are absolutely summable and smaller than one in absolute value for i sufficiently large. Hence can regular variation of $\mathbf{y}(k)$ be verified by the same arguments as above. This completes the proof of Lemma 2 $\square$

Proof of Theorem 2. Define the empirical autocovariance and the empirical autocorrelation as

$$\begin{aligned}
\gamma_T(r) &= \frac{1}{T}\sum_{t=1}^{T} y_t y_{t+r}, \quad r = 0, 1 \\
\rho_T(1) &= \gamma_T(1)/\gamma_T(0).
\end{aligned}$$

These are clearly closely related to the OLS estimator of the autoregressive parameter. We will therefore in the following prove that $\gamma_T(r) - E[\gamma_T(r)]$ and $\rho_T(1) - E[\gamma_T(1)]/E[\gamma_T(0)]$ are both asymptotically stable with index $\lambda/2$.

Define $\mathbf{y}_t(k) = (y_t, ..., y_{t+k})'$ and let a_T be a sequence such that $TP(|y_t| > a_T) \to 1$ (one can choose a_T to be the $1-1/T$ quantile of the distribution function for $|y_t|$). The proof is structured as the proof of Theorem 2.10 in Basrak et al. (2002*b*) and we must therefore verify that

$(A.1)$ $\mathbf{y}_t(k)$ is regularly varying for all $k \geq 1$,

$(A.2)$ the mild mixing condition $\mathcal{A}(a_T)$ from Davis & Mikosch (1998) p. 2052,

$(A.3)$ condition (2.10) of Davis & Mikosch (1998), and

$(A.4)$ condition (3.3) of Davis & Mikosch (1998).

$(A.1)$ follows straight from Lemma 2. Furthermore Lemma 1 establishes that the Markov chain $(y_{t-1}, \varepsilon_t)'$ is geometrically ergodic, this implies in particular that the stationary version is strongly mixing (actually even β-mixing) with geometrically decreasing rate function. And since the condition $\mathcal{A}(a_T)$ is implied by strong mixing the verification of $(A.2)$ is complete.

The two remaining conditions require a bit more work. With $|\cdot|$ denoting the max norm condition (2.10) of Davis & Mikosch (1998) can be stated as

$$\lim_{m\to\infty} \limsup_{T\to\infty} P\big(\bigvee_{m \leq |t| \leq r_T} |\mathbf{y}_t(k)| > a_T x \;\big| |\mathbf{y}_0(k)| > a_T x \big) = 0, \;\; x > 0, \tag{14}$$

where r_T is an integer sequence such that $r_T \to \infty$ and $r_T/T \to 0$ as $T \to \infty$. By the definition of conditional probabilities, Markov's inequality, and the symmetry of the distributions it holds for $t > 0$ that

$$\begin{aligned} P(|y_t| > a_T x \mid |y_0| > a_T x) &\leq \frac{E[1_{\{|y_0|^2 > a_T^2 x^2\}} |y_t|^2]}{a_T^2 x^2 P(|y_0|^2 > a_T^2 x^2)} \\ &= \frac{E\left[1_{\{|y_0|^2 > a_T^2 x^2\}} (\rho^{2t} |y_0|^2 + \sum_{i=0}^{t-1} \rho^{2i} \varepsilon_{t-i}^2)\right]}{a_T^2 x^2 P(|y_0|^2 > a_T^2 x^2)} = I_{t,T}. \end{aligned}$$

The recursion of Nelson (1990) gives

$$E_0[\varepsilon_t^2] = \varepsilon_0 \alpha^t + \omega \frac{1-\alpha^t}{1-\alpha} \leq \varepsilon_0 \alpha^t + C_0,$$

for some positive constant C_0 independent of t. Direct calculations provide the relation

$$\sum_{i=0}^{t-1} \rho^{2i} \alpha^{t-i} = \begin{cases} \frac{\alpha(\rho^{2t}-\alpha^t)}{\rho^2-\alpha} & \text{if } \alpha \neq \rho^2 \\ t\alpha^t & \text{if } \alpha = \rho^2 \end{cases}.$$

Note that the sum converges to zero as t tends to infinity for all ρ, α smaller than one in absolute value. Introduce the auxiliary process $\tilde{y}_t = \sum_{i=0}^{\infty} |\rho|^i |\varepsilon_{t-i}|$, which is clearly positive. Inspecting the proof of Lemma 2 reveals that $\tilde{y}_t$ is regularly varying with tail index λ. In addition one has the relation $\tilde{y}_t \geq |y_t|$ for all t and $\tilde{y}_0 \geq |\varepsilon_0|$. Hence it holds that

$$\frac{E[1_{\{|y_0|>a_T x\}} \varepsilon_0^2]}{a_T^2 x^2 P(|y_0| > a_T x)} \leq \frac{E[1_{\{|\tilde{y}_0|^2 > a_T^2 x^2\}} \tilde{y}_0^2]}{a_T^2 x^2 P(|\tilde{y}_0^2| > a_T^2 x^2)} \frac{P(|\tilde{y}_0| > a_T x)}{P(|y_0| > a_T x)},$$

and by Karamata's Theorem (e.g. Resnick (1987) Proposition 0.6)

$$\limsup_{T\to\infty} \frac{E[1_{\{|y_0|>a_T x\}} \varepsilon_0^2]}{a_T^2 x^2 P(|y_0| > a_T x)} \leq \frac{C_1}{\lambda - 2},$$

for some constant C_1. Applying Karamata's Theorem again it can concluded that there exists T_0 such that for all $T > T_0$ it holds

$$\begin{aligned} I_{t,T} &\leq \frac{E\left[1_{\{|y_0|>a_T x\}} \left(\rho^{2t}|y_0|^2 + \varepsilon_0^2 \frac{\alpha(\rho^{2t}-\alpha^t)}{\rho^2-\alpha} + C_2\right)\right]}{a_T^2 x^2 P(|y_0| > a_T x)} \\ &\leq C_3 \rho^{2t} + C_4 \frac{\alpha(\rho^{2t} - \alpha^t)}{\rho^2 - \alpha} + \frac{C_2}{a_T^2 x^2} \\ &\leq C_5 a^t + \frac{C_2}{a_T^2 x^2}, \end{aligned}$$

for some positive constants $C_2, ..., C_5$ and $a \in]0, 1[$ all independent of t, since by assumption both ρ and α are smaller that one in absolute value. Note that the special case $\alpha = \rho^2$ can be treated using the same arguments. We are now ready to verify (14).

$$
\begin{aligned}
&\lim_{m\to\infty}\limsup_{T\to\infty} P\big(\bigvee_{m\le|t|\le r_T}|\mathbf{y}_t(k)|>a_Tx\big||\mathbf{y}_0(0)|>a_Tx\big)\\
&\le \lim_{m\to\infty}\limsup_{T\to\infty} 2(k+1)\sum_{t=m}^{r_T+k} P\left(|y_t|>a_Tx \mid |y_0|>a_Tx\right)\underbrace{\frac{P(|y_0|>a_Tx)}{P(|\mathbf{y}_0(k)|>a_Tx)}}_{\le 1}\\
&\le \lim_{m\to\infty} 2(k+1)\sum_{t=m}^{\infty} C_5 a^t + \lim_{m\to\infty}\limsup_{T\to\infty} 2(k+1)(r_T+k)/(a_T^2x^2)\\
&= 0,
\end{aligned}
$$

by choosing r_T such that $r_T/a_T^2 \to 0$. Note that negative values of t are dealt with by noting that due to stationarity the following relation holds for $t>0$

$$
\begin{aligned}
P(|y_{-t}|>a_Tx \mid |y_0|>a_Tx) &= P(|y_{-t}|>a_Tx, |y_0|>a_Tx)/P(|y_0|>a_Tx)\\
&= P(|y_{-t}|>a_Tx, |y_0|>a_Tx)/P(|y_{-t}|>a_Tx)\\
&= P(|y_t|>a_Tx \mid |y_0|>a_Tx)
\end{aligned}
$$

This completes the verification of $(A.3)$. Finally $(A.4)$ is considered. In the setup of the AR-ARCH model condition (3.3) of Davis & Mikosch (1998) reads

$$
\begin{aligned}
&\lim_{x\to 0}\limsup_{T\to\infty} P\Big(\big|a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_ty_{t+1}|\le a_T^2 x\}} - E\big[a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_ty_{t+1}|\le a_T^2 x\}}\big]\big|>\delta\Big)\\
&\quad = 0,
\end{aligned}
$$

for all $\delta>0$, which can also be found in Davis & Hsing (1995) p. 895. Markov's inequality and Kamarata's Theorem (the required regular variation of y_ty_{t+1} can be verified by the same arguments as for y_t) now give

$$\begin{aligned}
& P(|a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_t y_{t+1}|\le a_T^2 x\}} - E[a_T^{-2}\sum_{t=1}^{T} y_t y_{t+1} 1_{\{|y_t y_{t+1}|\le a_T^2 x\}}]| > \delta) \\
&\quad \le \frac{1}{\delta^2} a_T^{-4} \sum_{t=1}^{T} E[(y_t y_{t+1} 1_{\{|y_t y_{t+1}|\le a_T^2 x\}} - E[y_t y_{t+1} 1_{\{|y_t y_{t+1}|\le a_T^2 x\}}])^2] \\
&\quad \le \frac{4}{\delta^2} a_T^{-4} \sum_{t=1}^{T} E[y_t^2 y_{t+1}^2 1_{\{|y_t y_{t+1}|^2\le a_T^4 x^2\}}] \\
&\quad = \frac{4}{\delta^2} a_T^{-4} T E[y_0^2 y_1^2 1_{\{|y_0 y_1|^2\le a_T^4 x^2\}}] \\
&\quad \sim C_6 x^2 T P(|y_0 y_1|^2 > a_T^4 x^2) \text{ for large } T \\
&\quad \to C_7 x^2 \text{ as } T \to \infty \\
&\quad \to \text{ as } x \to 0.
\end{aligned}$$

Using the same arguments, it can be shown that $(A.4)$ also holds for the sequence y_t^2. This completes the verification of $(A.4)$. Due to $(A.1)$ - $(A.4)$ one can apply Theorem 3.5 of Davis & Mikosch (1998). Note that their condition (3.4) is not meet, but by inspecting the proof it becomes clear that $(A.4)$ suffices. Hence it holds that

$$\begin{aligned}
Ta_T^{-2}(\gamma_T(r) - E[\gamma_T(r)]) &\xrightarrow{D} W_r \quad \text{as } T \to \infty, \quad r = 0, 1 \\
Ta_T^{-2}(\rho_T(1) - E[\gamma_T(1)]/E[\gamma_T(0)]) &\xrightarrow{D} S_0 \quad \text{as } T \to \infty,
\end{aligned}$$

where W_0, W_1, and S_0 are $\lambda/2$-stable random variables. As the convergence result in Theorem 3.5 of Davis & Mikosch (1998) is based on an application of the continuous mapping theorem the stated convergence results hold jointly. Since a_T can be chosen to be the $1 - 1/T$ quantile of the distribution function of $|y_t|$ one gets that a_T can be chosen as $a_T = T^{1/\lambda}$. This implies that the normalizing sequence Ta_T^{-2} can be chosen as $T^{1-2/\lambda}$. Hence it holds that

$$T^{1-2/\lambda}(\hat{\rho}_{OLS} - \rho_0) = T^{1-2/\lambda}(\rho_T(1) - E[\gamma_T(1)]/E[\gamma_T(0)]) \xrightarrow{D} S_0.$$

This completes the proof. $\square$

Proof of Theorem 3. Since the process $(\varepsilon_t)_{t=1}^T$ is an ARCH(1) process it follows directly from Davis & Mikosch (1998) pp. 2069 - 2070 that

$$T^{-2/\lambda}\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1} \xrightarrow{D} S_1 \quad \text{and} \quad T^{-4/\lambda}\sum_{t=1}^{T}\varepsilon_t^2\varepsilon_{t-1}^2 \xrightarrow{D} S_2,$$

where S_1 is a $\lambda/2$ stable random variable and S_2 is a $\lambda/4$ stable random variable. Again the convergence hold jointly, but the remaining parameters and dependence structure are unknown. Under H_0 we can rewrite V_T as

$$V_T = \big(T^{-2/\lambda}\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}\big)\big(T^{-4/\lambda}\sum_{t=1}^{T}\varepsilon_t^2\varepsilon_{t-1}^2\big)^{-1/2},$$

and the continuous mapping theorem completes the proof. $\square$

On IGARCH and convergence of the QMLE for misspecified GARCH models

By Anders Tolver Jensen & Theis Lange

Department of Natural Sciences & Department of Mathematical Sciences
University of Copenhagen
E-mail: *tolver@life.ku.dk & lange@math.ku.dk*

Abstract: We address the IGARCH puzzle by which we understand the fact that a GARCH(1,1) model fitted by quasi maximum likelihood estimation to virtually any financial dataset exhibit the property that $\hat{\alpha} + \hat{\beta}$ is close to one. We prove that if data is generated by certain types of continuous time stochastic volatility models, but fitted to a GARCH(1,1) model one gets that $\hat{\alpha} + \hat{\beta}$ tends to one in probability as the sampling frequency is increased. Hence, the paper suggests that the IGARCH effect could be caused by misspecification. The result establishes that the stochastic sequence of QMLEs do indeed behave as the deterministic parameters considered in the literature on filtering based on misspecified ARCH models, see e.g. Nelson (1992). An included study of simulations and empirical high frequency data is found to be in very good accordance with the mathematical results.

Keywords: GARCH; Integrated GARCH; Misspecification; High frequency exchange rates.

1 Introduction

A complete characterization of the volatility of financial assets has long been one of the main goals of financial econometrics. Since the seminal papers of Engle (1982) and Bollerslev (1986) the class of generalized autoregressive heteroskedastic (GARCH) models has been a key tool when modeling time dependent volatility. Indeed the GARCH(1,1) model has become so widely used that it is often referred to as “the workhorse of the industry” (Lee & Hansen 1994).

Recall that given a sequence of returns $(y_t)_{t=0,\dots,T}$ the GARCH(1,1) model defines the conditional volatility as

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

for some non-negative parameters $\theta = (\omega, \alpha, \beta)'$. Quasi maximum likelihood estimation of GARCH(1,1) models on financial returns almost always indicates that $\hat{\alpha}$ is small, $\hat{\beta}$ is close to unity, and the sum of $\hat{\alpha}$ and $\hat{\beta}$ is very close to one and approaches one as the sample is increased, see e.g. Engle & Bollerslev (1986), Bollerslev & Engle (1993), Baillie, Bollerslev & Mikkelsen (1996), Ding & Granger (1996), Andersen & Bollerslev (1997), and Engle & Patton (2001). This feature seems to be present independently of the considered asset class or sampling frequency. Engle & Bollerslev (1986) proposed the integrated GARCH (IGARCH) model specifically to reflect this fact. Also in the recent litterateur on quasi maximum likelihood estimation in GARCH models it has been paramount to allow for $\alpha + \beta$ to be close to or even exceeding one, see e.g. Jensen & Rahbek (2004*a*) and Francq & Zakoïan (2004). IGARCH implies that the return series is not covariance stationary and multiperiod forecasts of volatility will trend upwards. Recently it has been suggested that either long memory, see e.g. Mikosch & Stărică (2004), or parameter changes, see e.g. Hillebrand (2005), in the data generating process can give the impression of IGARCH.

In a series of seminal papers Nelson (1992), Nelson & Foster (1994), and Nelson & Foster (1995) explore the consequences of applying ARCH type filters on discrete samples from continuous time stochastic volatility models. One important result demonstrates the existence of a deterministic sequence of parameters for the GARCH(1,1) model such that the difference between the GARCH conditional volatility estimates based on (1) and the true volatility converges to zero in probability as the sampling frequency is increased. This may explain the success of ARCH type models at recovering and forecasting volatility even though they are, no doubt, misspecified. The result, however, depends on the fact that the chosen parameters have the IGARCH property and do not depend on the data. Indeed Nelson & Foster (1994) stress the need for extending their results to cover filters based on quasi maximum likelihood estimators (QMLEs) from ARCH type models. In addition a number of papers (see, e.g. Drost & Nijman (1993), Drost & Werker (1996), and Francq & Zakoïan (2000)) explore the connection between continuous time stochastic volatility models and discrete time GARCH models. They establish a link between parameters of the two classes of models and consequently suggest that one estimates parameters in the continuous time model from the GARCH estimates. However, Wang (2002) warns against applying statistical inference based on a GARCH model to its continuous time counterpart as this may lead to erronerous conclusions. The concerns of Nelson & Foster (1995) and Wang (2002) illustrate the need to understand better the behaviour of QMLEs for misspecified GARCH models.

In this paper we propose a simple stochastic volatility type model that enables us to study the statistical properties of the QMLE based on the GARCH(1,1) model (1). We prove in Theorem 1 that as the sampling frequency is increased certain data generating processes will spuriously lead to the conclusion of IGARCH. The employed infill asymptotics has recently been much used in the literature on realized volatility, see e.g. Andersen, Bollerslev, Diebold & Labys (2003) and Barndorff-Nielsen & Shephard (2001).

The paper also provides a more intuitive explanation of the IGARCH puzzle by exposing similarities between the GARCH model and non-parametric estimation of a volatility process, see Stărică (2003) for a related study. The GARCH model provides a filter for computing the present volatility as, roughly speaking, a weighted average of past squared observations and a constant. Examination of the weights and the shape of the quasi likelihood function makes it plausible to believe that the performance of the filter is optimized when α and β sum to one.

Finally, since the theoretical results not only establish that the sum of the GARCH parameters will tend to one, but also indicate that they will do so at a polynomial rate, an illustration using high frequency exchange rates as well as simulated data is provided. The results are found to be in remarkably good accordance with the theoretical results and furthermore indicates that Theorem 1 is valid for other models than the ones covered by the present proof.

The rest of the paper is organized as follows. Section 2 presents the main result and explores connections between the GARCH(1,1) model and non-parametric estimation of volatility. Section 3 illustrates our results by both simulations and empirical data, while Section 4 concludes and presents ideas for future research. All technical lemmas are deferred to the Appendix.

2 Main Results

Based on a large class of volatility models this section initially provides a more heuristic explanation of the IGARCH puzzle by exposing similarities between the GARCH model and non-parametric estimation of a volatility process. In the second part of the section we present a mathematical setup where these heuristic arguments can be formalized and we state our main theorem.

2.1 An Intuitive Explanation of the IGARCH Puzzle

Essentially all volatility models for a sequence $(y_t)_{t=0,\ldots,T}$ can be captured by the formulation

$$y_t = \sqrt{f_t} \cdot z_t, \tag{2}$$

where z_t is a sequence of zero mean random variables with unit variance and $(f_t)_{t=0,\ldots,T}$ a sequence of stochastic volatilities such that z_t is independent of $(f_t, y_{t-1}, \ldots, y_0)$. Define $\sigma_t^2(\theta)$ to be the conditional variance process corresponding to the GARCH(1,1) model with parameters $\theta = (\omega, \alpha, \beta)'$

$$\begin{aligned} \sigma_t^2(\theta) &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta) \\ &= \omega \sum_{i=0}^{t-1} \beta^i + \alpha \sum_{i=0}^{t-1} \beta^i y_{t-1-i}^2 + \beta^t \sigma_0^2, \end{aligned} \tag{3}$$

with σ_0^2 a fixed constant. Consider the usual quasi log-likelihood function

$$l_T(\theta) = -\frac{1}{T} \sum_{t=1}^{T} (\log(\sigma_t^2(\theta)) + \frac{y_t^2}{\sigma_t^2(\theta)}) \tag{4}$$

and note that under the data generating process given by (2) the likelihood function may be rewritten as

$$l_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} (1 - z_t^2) \frac{f_t}{\sigma_t^2(\theta)} - \frac{1}{T} \sum_{t=1}^{T} (\log(\sigma_t^2(\theta)) + \frac{f_t}{\sigma_t^2(\theta)}). \tag{5}$$

Strictly speaking this is not a likelihood function, but just an objective function for the GARCH(1,1) model, but the terminology emphasizes the connection to the literature on estimation of GARCH models. Since the first term has zero mean (if finite) and the function $x \mapsto -\log(x) - a/x$ has a unique maximum at $x = a$, the decomposition (5) suggests that for a large class of data generating processes it is plausible that the likelihood function is optimized when the conditional variance process is close to the true unobserved volatility process f_t.

For large values of t the conditional variance process in (3) can be viewed as a kernel estimator of the unobserved volatility at time t with kernel weights $\alpha\beta^i, i = 0, \ldots, t-1$ on past observations $y_{t-1}^2, \ldots, y_0^2$ plus the constant $\frac{\omega}{1-\beta}$. In order for this to be an unbiased estimator of the non-constant volatility f on average over the entire sample one must have $\sum_{i=0}^{\infty} \alpha\beta^i = \frac{\alpha}{1-\beta} \approx 1$ and the constant $\frac{\omega}{1-\beta}$ small. Hence, when considering the conditional variance process, $\sigma_t^2(\theta)$, as a non-parametric estimator of the unobserved volatility one must have $\alpha + \beta \approx 1$ and ω small in order to avoid introducing a systematical bias. Clearly, the method above is not always the optimal way to match the conditional variance process, $\sigma_t^2(\theta)$, with the volatility process, f_t. For instance if the data generating process is in fact the GARCH(1,1) model one should choose θ to be the true parameter value and hence obtain $\sigma_t^2(\theta) = f_t$.

2.2 A Mathematical Explanation of the IGARCH Puzzle

In the following we introduce a mathematical framework allowing us to formalize the considerations above. Clearly, we cannot give unified mathematical proofs of our results covering all interesting stochastic volatility models. However, the framework below offers a compromise between flexibility of the model class and clarity of the formal mathematical arguments. Following Theorem 1 we discuss possible generalizations.

Let the continuous time process $(S_u)_{u \in [0,1]}$ be a solution to the stochastic differential equation

$$dS_u = \sqrt{f(u)}dW_u, \tag{6}$$

where W is a standard Brownian motion and f is a strictly positive continuous function on the unit interval. Consider a discrete sample $(r_t)_{t=1,\ldots,T}$ of returns given by

$$r_t = S_{t/T} - S_{(t-1)/T} \sim N(0, \int_{(t-1)/T}^{t/T} f(u)du).$$

For large T the distribution of the returns (scaled by $\sqrt{T}$) will resemble the distribution of a sequence $(y_t)_{t=1,\ldots,T}$ generated by

$$y_t = \sqrt{f(t/T)} \cdot z_t, \tag{7}$$

where z_t is an i.i.d. sequence of zero mean unit variance Gaussian random variables. By extending the model in (6) to allow for a stochastic volatility process f the formulation encompasses a number of applied stochastic volatility models. Thus we claim that results based on (7) will also be relevant for the continuous time model (6), see Remark 2. Indeed many of the results by D.B. Nelson concerning misspecified ARCH models, see e.g. Nelson (1992), are based on models of the type captured by (6) and also make use of rescaling increments of discrete samples from these models.

Consider the sequence of parameters $\theta_T = (0, T^{-d}, 1 - T^{-d})'$ and introduce the stochastic processes

$$h_T(u) = \sigma^2_{\lfloor Tu \rfloor}(\theta_T)$$

on $u \in [0,1]$, where $\sigma_t(\theta)$ is given by the GARCH(1,1) recursion (3). Here and throughout the paper $\lfloor x \rfloor$ denotes the integer part of x. Further, let $D([a,b])$ denote the space of càdlàg functions on the interval $[a,b]$.

Lemma 1. *If $E[z_t^8] < \infty$ then for any $d \in]1/2, 1[$ and $\gamma \in]0,1]$ the process $h_T \xrightarrow{P} f$ in the uniform norm on $D([\gamma, 1])$ as T tends to infinity.*

The lemma establishes that there exists a sequence of parameters such that the conditional variance process associated with the GARCH(1,1) model gets arbitrarily close to the unobserved volatility process when the sampling frequency is increased. The lemma is an analogue to Theorem 3.1 of Nelson (1992), however, our result is given for the uniform norm, but assuming a somewhat simpler data generating process.

Proof of Lemma 1. Introduce the notation $g_T(u) := \mathbb{E}[h_T(u)]$ for $u \in [0,1]$. For $\gamma, \eta > 0$

$$\begin{aligned} & \mathbb{P}(\sup_{u\in[\gamma,1]} |h_T(u) - f(u)| > \eta) \\ \leq \; & \mathbb{P}(\sup_{u\in[\gamma,1]} |h_T(u) - g_T(u)| > \eta/2) + \mathbb{P}(\sup_{u\in[\gamma,1]} |g_T(u) - f(u)| > \eta/2). \end{aligned}$$

By Lemma 3 in the Appendix the last term converges to zero as T tends to infinity. To handle the first term note that by Lemma 2 in the Appendix it holds that

$$\begin{aligned} & \mathbb{P}(\sup_{u\in[\gamma,1]} |h_T(u) - g_T(u)| > \eta/2) \\ = \; & \mathbb{P}(\max_{t=\lfloor T\gamma\rfloor -1,\dots,T} |h_T(t/T) - g_T(t/T)| > \eta/2) \\ \leq \; & \sum_{t=\lfloor T\gamma\rfloor -1}^{T} \mathbb{P}(|h_T(t/T) - g_T(t/T)| > \eta/2) \\ \leq \; & A\eta^{-4} T \alpha_T^2 \end{aligned}$$

which converges to zero as T tends to infinity since $\alpha_T = T^{-d}$ with $d > 1/2$. $\square$

Before stating our main theorem define the parameter set

$$\Theta = \{(\omega, \alpha, \beta)' \in \mathbb{R}^3 \mid 0 \leq \omega, 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1\} \tag{8}$$

and let $\hat{\theta}_T = (\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)' = \arg\max_{\theta\in\Theta} l_T(\theta)$ be the usual quasi maximum likelihood estimator based on (4).

Theorem 1. *Suppose that the data generating process is given by (7) where f is non-constant and $\mathbb{E}[z_t^8] < \infty$. Then the QMLE based on (4) satisfies that $(\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)' \xrightarrow{P} (0,0,1)'$ as T tends to infinity.*

Remark 1. The initial value σ_0^2 for the conditional volatility process $\sigma_t^2(\theta)$ does not need to be a constant. For instance Theorem 1 still holds if σ_0^2 is merely bounded in probability as T tends to infinity. This includes defining σ_0^2 as the unconditional variance of the full sample, which is implemented in many software packages.

Remark 2. When defining σ_0^2 as the unconditional variance of the full sample it is a simple consequence of the GARCH(1,1) recursion that any scaling of the observations only affects the estimate of the scale parameter ω. Hence, if the QMLE is based on the unscaled returns, r_t, from the continuous time model (6) Theorem 1 remains valid. When extending the model (6) to allow for a stochastic volatility the present proof of Theorem 1 requires that the volatility process be independent of the Brownian motion. In this case the result should be read as conditional on the sample path of the volatility process.

Remark 3. To facilitate the presentation we have assumed that the volatility process f is a continuous function. However, the proofs can be extended to cover a finite number of discontinuities at the price of a somewhat more cumbersome notation and Theorem 1 therefore remains valid.

Remark 4. The proof is given for the case of Gaussian innovations z_t, however, it can easily be adapted to most other distributions such as the t-distribution as long as the moment condition in Theorem 1 is met. Another generalization is to allow for some dependence in the sequence of innovations. For instance including an autoregressive structure on z_t would permit modeling leverage effects, but leads to considerably more complicated proofs.

Proof of Theorem 1. For $\omega_U > 0$ divide the full parameter space Θ defined in (8) into the compact subset

$$\Theta_{\omega_U} := \{\theta = (\alpha, \beta, \omega)' \in \Theta \mid \omega \leq \omega_U\}$$

and its complement $\Theta_{\omega_U}^c$. Let

$$V_\epsilon(0, 0, 1) = \{(\omega, \alpha, \beta)' \in \Theta \mid ||(\omega, \alpha, \beta)' - (0, 0, 1)'|| < \epsilon\}$$

and use Lemma 6 in the Appendix to construct a finite covering

$$\cup_{i=1}^k V(\theta_i) \supset \Theta_{\omega_U} \backslash V_\epsilon(0, 0, 1)$$

of the compact set $\Theta_{\omega_U} \backslash V_\epsilon(0, 0, 1)$ with open subsets of Θ and let $\gamma_{\theta_1}, \ldots, \gamma_{\theta_k} > 0$

be constants such that according to Lemma 6

$$\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^*\in V(\theta_i)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - \gamma_{\theta_i}) = 1$$

for $i = 1, \dots, k$. With $\gamma = \min(\gamma_{\theta_1}, \dots, \gamma_{\theta_k})$ we conclude that

$$\begin{aligned}
1 &\geq \mathbb{P}(\sup_{\theta\in\Theta\setminus V_\epsilon(0,0,1)} l_T(\theta) < -\int_0^1 \log(f(u))du - 1 - \gamma) \\
&\geq \mathbb{P}(\sup_{\theta\in\cup_{i=1}^k V(\theta_i)\cup\Theta^c_{\omega_U}} l_T(\theta) < -\int_0^1 \log(f(u))du - 1 - \gamma) \\
&\geq 1 - \sum_{i=1}^k \mathbb{P}(\sup_{\theta\in V(\theta_i)} l_T(\theta) \geq -\int_0^1 \log(f(u))du - 1 - \gamma) \qquad (9) \\
&- \mathbb{P}(\sup_{\theta\in\Theta^c_{\omega_U}} l_T(\theta) \geq -\int_0^1 \log(f(u)))du - 1 - \gamma) \qquad (10)
\end{aligned}$$

where by construction (9) converges to one as T tends to infinity. Further, as $\sigma_t^2(\theta) \geq \omega_U$ on $\Theta^c_{\omega_U}$ we get that

$$\sup_{\theta\in\Theta^c_{\omega_U}} l_T(\theta) = \sup_{\theta\in\Theta^c_{\omega_U}} -\frac{1}{T}\sum_{t=1}^T (\log(\sigma_t^2(\theta)) + \frac{y_t^2}{\sigma_t^2(\theta)}) \leq -\log(\omega_U)$$

hence the probability in (10) is zero if we choose ω_U large enough. By Lemma 4 in the Appendix it holds that $l_T(\theta_T) \xrightarrow{P} -\int_0^1 \log(f(u))du - 1$ and since $l_T(\hat{\theta}_T) \geq l_T(\theta_T)$ we conclude that for any $\epsilon > 0$

$$\lim_{T\to\infty} \mathbb{P}(\hat{\theta}_T \in V_\epsilon(0,0,1)) = 1.$$

□

3 Illustrations

The main result (Theorem 1) establishes that for certain data generating processes the quasi maximum likelihood estimators for the GARCH(1,1) model will converge to $(0,0,1)'$ as the sampling frequency increases. In this section we illustrate the convergence results and go a step further by examining the rate of convergence as well. Based on Lemma 1 one could conjecture that $\hat{\alpha}_T$ and $1 - \hat{\beta}_T$ are proportional to T^{-d} for some $d \in (0, 1)$. This assertion can be visualized by plotting $\log(\hat{\alpha}_T)$ and $\log(1 - \hat{\beta}_T)$ against $\log(T)$. If a linear relationship is found the parameter d can be estimated by ordinary least squares.

The first part of the study is based on high frequency recordings of the EUR-USD exchange rate. To increase the empirical relevance of the simulation part we use broadly applied continuous time models as data generating processes. However, formally these models do not satisfy the assumptions of Theorem 1. In this respect the simulation study actually demonstrates that the scope of the results might be extended to a wider class of models.

EUR-USD. Based on 30-minute recordings of the EUR-USD exchange rate spanning the period from the 2$^{\text{nd}}$ of February 1986 to the 30$^{\text{th}}$ of March 2007[1] log-returns are computed corresponding to 4 through 72 hour returns. This gives estimates $\hat{\theta}_T$ for T between 3.687 and 64.525.

Simulations. We consider three different simulation setups including the Heston model and the continuous GARCH model (obtained as the diffusion limit of a GARCH(1,1) model, see Nelson (1990)). The considered models can all be embedded in the formulation

$$dS_u = S_u V_u^{1/2} dW_{1u}, \quad dV_u = \kappa V_u^a (\mu - V_u) du + \sigma V_u^b dW_{2u},$$

where W_1 and W_2 are standard Brownian motions with a possibly non-zero correlation denoted by ρ. For ease of exposition we have omitted a drift term in the equation for dS_u. We will consider three configurations for the parameters a and b, corresponding to the Heston model, the continuous GARCH model, and the $3/2N$ model suggested in Christoffersen, Jacobs & Mimouni (2007). To make the simulations comparable to the empirical study we consider a fixed time span of 21 years. For the remaining parameters we choose the estimated values stated in Christoffersen et al. (2007), which are based on fitting the models to S&P-500 data. By this choice of time span and parameter values it is reasonable to com-

[1]Prior to January 1999 the series is generated from the DEM-USD exchange rate using a fixed exchange rate of 1.95583 DEM per EUR. Preceding the analysis the dataset has been cleaned as described in Andersen et al. (2003).

Name	a	b	κ	μ	σ	ρ
Heston	0	1/2	6.5200	0.0352	0.4601	-0.7710
Continuous GARCH	0	1	3.9248	0.0408	2.7790	-0.7876
$3/2N$	1	3/2	60.1040	0.0837	12.4989	-0.7591

Table 1: Parameter values used in the simulation study.

pare the empirical study and the simulation study directly. GARCH(1,1) models are fitted to log-returns based on a discrete sample from the S process[2]. Table 1 summarizes the parameters (per annum) for the included models. We will consider log-returns corresponding to weekly through 5 minute returns, which gives estimates $\hat{\theta}_T$ for T between 1,000 and 300,000.

Figure 1 reports the correspondence between the estimates of α and T for the four setups. The conjectured linear relationship between $\log(\hat{\alpha}_T)$ and $\log(T)$ is clearly present. The corresponding plots for $1 - \hat{\beta}_T$ have been omitted since they are indistinguishable from Figure 1. In particular we have verified the IGARCH property, i.e. that $(\hat{\alpha}_T, \hat{\beta}_T) \to (0, 1)$. The estimated values for d are in all cases found to be between 0.25 and 0.5, but explaining this phenomenon is left for future research.

The fact that none of the simulations satisfy the assumptions clearly indicates that Theorem 1 holds for a far larger class of models than those covered by the present version of our proof. This emphasizes that the IGARCH effect can be caused by the mathematical structure of a GARCH model alone and hence might not be a property of the true data generating mechanism. That the apparent polynomial convergence of the QMLEs is not only a property of the simulated series is illustrated by the striking similarities between plots based on simulated and real data.

4 Conclusion

In this paper we have established that when a GARCH(1,1) model is fitted to a discrete sample from a certain class of continuous time stochastic volatility models then the sum of the quasi maximum likelihood estimates of α and β will

[2]The continuous time process is simulated by a standard Euler scheme using 10^8 data points.

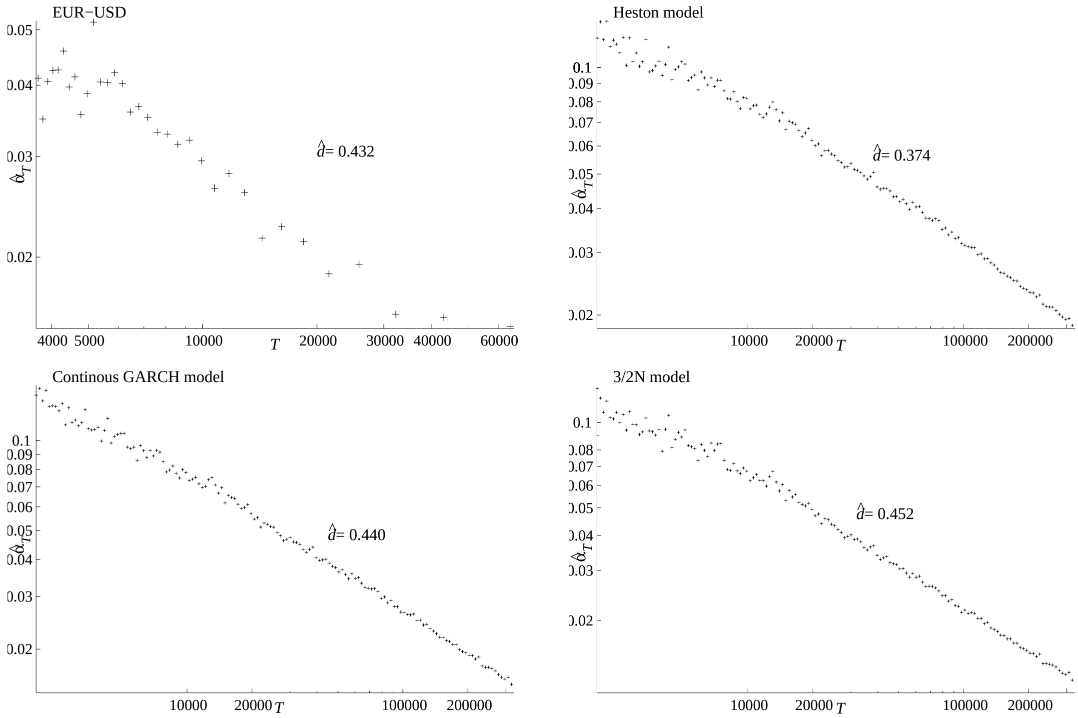

Figure 1: Correspondence between $\hat{\alpha}_T$ and T in log-scale for the four configurations. The estimate of d is obtained by regressing $\log(\hat{\alpha}_T)$ on $\log(T)$ and a constant.

converge to one in probability as the sampling frequency is increased. Our results therefore indicate that the IGARCH property often found in empirical work could be explained by misspecification.

The work of Nelson (1992) showed that it is possible to make the conditional variance process based on ARCH type models with deterministic parameters converge to the true unobserved volatility process. The parameters must here satisfy that $(\omega_T, \alpha_T, \beta_T) \to (0, 0, 1)$ as the number of sample points T tends to infinity. Our main result states that the same convergence holds for the *stochastic* sequence of quasi maximum likelihood estimators.

The simulations and the empirical study confirm the theoretical results and further suggest that: i) the assumptions of the main results may be weakened considerably and ii) that it may be possible to derive the exact rate of convergence of the estimators in specific mathematical frameworks. These questions are left for future research.

Appendix: Auxiliary lemmas

Lemma 2. *If* $\mathbb{E}[z_t^8] < \infty$ *there exists some* $A > 0$ *such that for any* $\eta > 0$

$$\sup_{u\in[0,1]} \mathbb{P}[|h_T(u) - g_T(u)| > \eta] \leq A\eta^{-4}\alpha_T^2.$$

Proof. It follows from Chebychev's inequality that

$$\begin{aligned}
& \mathbb{P}(|h_T(u) - g_T(u)| > \eta) \\
\leq\ & \eta^{-4}\mathbb{E}[|h_T(u) - g_T(u)|^4] \\
\leq\ & \eta^{-4}\mathbb{E}[(\alpha_T \sum_{t=0}^{\lfloor Tu\rfloor-1} \beta_T^t f(\tfrac{\lfloor Tu\rfloor-1-t}{T})(z^2_{\lfloor Tu\rfloor-1-t} - 1))^4] \\
\leq\ & \eta^{-4}||f||_\infty\alpha_T^4\{\sum_{t=0}^{\lfloor Tu\rfloor-1} \beta_T^{4t}\kappa_4 + 2\eta^{-4}\alpha_T^4 \sum_{t=1}^{\lfloor Tu\rfloor-1}\sum_{j=0}^{t-1}\beta_T^{2t+2j}\kappa_2^2\} \\
\leq\ & A_1\eta^{-4}\alpha_T^4(\sum_{t=0}^{\infty}\beta_T^{4t} + \sum_{t=1}^{\infty}\beta_T^{2t}\tfrac{1-\beta_T^{2t}}{1-\beta_T^2}),
\end{aligned}$$

where we make use of the fact that f is bounded and that $\kappa_1 = 0$ with $k_r := \mathbb{E}[(z_t^2 - 1)^r]$. Evaluating the geometric series above, using that $\alpha_T = 1 - \beta_T$, and that the last expression does not depend on u one arrives at an inequality of the form stated in the lemma. □

Lemma 3. *For any* $\gamma > 0$ *then* $\sup_{u\in[\gamma,1]} |g_T(u) - f(u)| \to 0$ *as* T *tends to infinity.*

Proof. For any sequence c_T and any $u \in [\gamma, 1]$ we get

$$\begin{aligned}
& |g_T(u) - f(u)| \\
=\ & |\beta_T^{\lfloor Tu\rfloor}\sigma_0^2 + \alpha_T \sum_{t=0}^{\lfloor Tu\rfloor-1}\beta_T^t(f(\tfrac{\lfloor Tu\rfloor-t-1}{T}) - f(u)) - \alpha_T\sum_{t=\lfloor Tu\rfloor}^{\infty}\beta_T^t f(u)| \\
\leq\ & \beta_T^{\lfloor Tu\rfloor}\sigma_0^2 + \alpha_T\sum_{t=0}^{c_T-1}\beta_T^t|f(\tfrac{\lfloor Tu\rfloor-t-1}{T}) - f(u)| + \alpha_T\sum_{t=c_T}^{\infty}\beta_T^t||f||_\infty \\
\leq\ & \beta_T^{\lfloor T\gamma\rfloor}\sigma_0^2 + \alpha_T\frac{1-\beta_T^{c_T}}{1-\beta_T}\sup_{v\in[u-\frac{c_T}{T},u]}|f(v) - f(u)| + \alpha_T\frac{\beta_T^{c_T}}{1-\beta_T}||f||_\infty.
\end{aligned}$$

If $c_T/T = o(1)$ the uniform continuity of f implies that the middle term can be made arbitrary small by choosing T adequately large and that the convergence is uniform over $u \in [\gamma, 1]$. To complete the proof note that

$$\log(\beta_T^{c_T}) = c_T \log(1 - T^{-d}) = -c_T T^{-d} \frac{\log(1 - T^{-d}) - \log(1)}{T^{-d}} \to -\infty$$

as T tends to infinity provided that we choose c_T so that c_T/T^d tends to infinity as T tends to infinity. $\square$

Lemma 4. *For $d > 1/2$ then*

$$l_T(\theta_T) \xrightarrow{P} -\int_0^1 \log(f(u))du - 1, \quad as \quad T \to \infty.$$

Proof of Lemma 4. Rewriting the expression for $l_T(\theta_T)$ yields

$$\begin{aligned} l_T(\theta_T) &= -\frac{1}{T}\sum_{t=1}^{T}(\log(\sigma_t^2(\theta_T)) + \frac{f(t/T)}{\sigma_t^2(\theta_T)}) & (11) \\ &- \frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta_T)}(z_t^2 - 1) & (12) \end{aligned}$$

By the law of large numbers for martingale difference sequences (12) $\xrightarrow{P} 0$. Formally, since $\mathbb{E}[z_t^2 - 1] = 0$ and $\sigma_t^2(\theta_T)$ is $\mathcal{F}_{t-1}-$measurable we get by applying Chebechev's inequality that

$$\begin{aligned} &\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta_T)}(z_t^2 - 1)| > \eta) \\ \leq\ & \frac{B_1}{T^2}\sum_{i=1}^{T}\sum_{j=i}^{T}\mathbb{E}[\mathbb{E}[\frac{(z_i^2 - 1)(z_j^2 - 1)}{\sigma_i^2(\theta_T)\sigma_j^2(\theta_T)} \mid \mathcal{F}_{j-1}]] \\ =\ & \frac{B_2}{T^2}\sum_{t=1}^{T}\mathbb{E}[\frac{1}{\sigma_t^4(\theta_T)}] \leq \frac{B_3}{T\alpha_T^2\beta_T^{2c_T}}\mathbb{E}[\frac{1}{(z_1^2 + \ldots + z_{c_T}^2)^2}], \end{aligned}$$

where c_T is a sequence of positive integers. For T sufficiently large (Mathai &

Provost (1992), p. 59)

$$\mathbb{E}[\frac{1}{(z_1^2 + \ldots + z_{c_T}^2})^2] \leq \frac{B_4}{c_T^2}$$

hence

$$0 \leq \limsup_{T\to\infty} \mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T} \frac{f(t/T)}{\sigma_t^2(\theta_T)}(z_t^2 - 1)| > \eta) \leq \limsup_{T\to\infty} \frac{B_5}{T\alpha_T^2 \beta_T^{2c_T} c_T^2}$$

and by choosing $c_T = \lfloor \alpha_T^{-1} \rfloor = \lfloor T^d \rfloor$ the right hand side is zero.

For any $\gamma > 0$ (11) may be written as

$$\begin{aligned}
- & \frac{1}{T}\sum_{t=1}^{\lfloor T\gamma \rfloor - 1} \log(\sigma_t^2(\theta_T)) - \frac{1}{T}\sum_{t=1}^{\lfloor T\gamma \rfloor - 1} \frac{f(t/T)}{\sigma_t^2(\theta_T)} \\
- & \int_\gamma^1 \log(h_T(u))du - \int_\gamma^1 \frac{f(u)}{h_T(u)}du + \sum_{t=\lfloor T\gamma \rfloor}^{T} \int_{(t-1)/T}^{t/T} \frac{f(u) - f(t/T)}{h_T(u)}du,
\end{aligned}$$

using that $h_T(u)$ is piecewise constant on intervals of the form $[(t-1)/T, t/T[$. We deduce from Theorem 1 and the continuous mapping theorem that

$$\begin{aligned}
\int_\gamma^1 \log(h_T(u))du &\xrightarrow{P} \int_\gamma^1 \log(f(u))du \\
\int_\gamma^1 \frac{f(u)}{h_T(u)}du &\xrightarrow{P} 1 - \gamma \\
\int_\gamma^1 \frac{1}{h_T(u)}du &\xrightarrow{P} \int_\gamma^1 \frac{1}{f(u)}du.
\end{aligned}$$

By the uniform continuity of f we conclude that

$$\sum_{t=\lfloor T\gamma \rfloor}^{T} \int_{(t-1)/T}^{t/T} \frac{f(u) - f(t/T)}{h_T(u)}du \xrightarrow{P} 0.$$

For $\eta > 0$ then

$$\begin{aligned}\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor -1}\frac{1}{\sigma_t^2(\theta_T)}| > \eta) &\leq \mathbb{P}(\max_{t=1,\ldots,\lfloor T\gamma\rfloor -1}\frac{1}{T}\frac{1}{\sigma_t^2(\theta_T)} > \frac{\eta}{\lfloor T\gamma\rfloor - 1})\\ &\leq \mathbb{P}(\min_{t=1,\ldots,\lfloor T\gamma\rfloor -1}\sigma_t^2(\theta_T) \leq \frac{\gamma}{B_6})\\ &\leq \sum_{t=1}^{\lfloor T\gamma\rfloor}\mathbb{P}(\sigma_t^2(\theta_T) \leq \frac{\gamma}{B_6}).\end{aligned}$$

Noting that $\mathbb{E}[\sigma_t^2(\theta_T)] \geq \min(\underline{f}, \sigma_0^2) \equiv \underline{\sigma}^2 > 0$ uniformly in t and T we find that for $\gamma > 0$ sufficiently small then

$$\begin{aligned}\mathbb{P}(\sigma_t^2(\theta) \leq \frac{\gamma}{B_6}) &\leq \mathbb{P}(|\sigma_t^2(\theta_T) - \mathbb{E}[\sigma_t^2(\theta_T)]| \geq \mathbb{E}[\sigma_t^2(\theta_T)] - \frac{\gamma}{B_6})\\ &\leq \mathbb{P}(|\sigma_t^2(\theta_T) - \mathbb{E}[\sigma_t^2(\theta_T)]| \geq \underline{\sigma}^2 - \frac{\gamma}{B_6})\end{aligned}$$

we get by applying Lemma 2 that

$$\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor -1}\frac{1}{\sigma_t^2(\theta_T)}| > \eta) \leq B_7\lfloor T\gamma\rfloor\alpha_T^2$$

which tends to zero as T tends to infinity. For $\eta > 0$ given we get

$$\begin{aligned}&\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor -1}\log(\sigma_t^2(\theta_T))| > \eta)\\ \leq\ &\mathbb{P}(\max_{t=1,\ldots,\lfloor T\gamma\rfloor -1}|\frac{1}{T}\log(\sigma_t^2(\theta_T))| > \frac{\eta}{\lfloor T\gamma\rfloor - 1})\\ \leq\ &\mathbb{P}(\max_{t=1,\ldots,\lfloor T\gamma\rfloor -1}\sigma_t^2(\theta_T) \geq \exp(B_8/\gamma)) + \mathbb{P}(\min_{t=1,\ldots,\lfloor T\gamma\rfloor -1}\sigma_t^2(\theta_T) \leq \exp(-B_8/\gamma))\\ \leq\ &\sum_{t=1}^{\lfloor T\gamma\rfloor -1}\mathbb{P}(\sigma_t^2(\theta_T) \geq \exp(B_8/\gamma)) + \sum_{t=1}^{\lfloor T\gamma\rfloor -1}\mathbb{P}(\sigma_t^2(\theta_T) \leq \exp(-B_8/\gamma)).\end{aligned}$$

From the previous argument we find that for $\gamma > 0$ sufficiently small

$$\begin{aligned}\mathbb{P}(\sigma_t^2(\theta_T) \geq \exp(B_8/\gamma)) &\leq \mathbb{P}(|\sigma_t^2(\theta_T) - \mathbb{E}[\sigma_t^2(\theta_T)]| \geq \exp(B_8/\gamma) - \overline{\sigma}^2) \\ \mathbb{P}(\sigma_t^2(\theta_T) \leq \exp(-B_8/\gamma)) &\leq \mathbb{P}(|\sigma_t^2(\theta_T) - \mathbb{E}[\sigma_t^2(\theta_T)]| \geq \underline{\sigma}^2 - \exp(-B_8/\gamma)),\end{aligned}$$

where $\overline{\sigma}^2 = \sigma_0^2 + \|f\|_\infty$. From Lemma 2 we get that

$$\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor-1} \log(\sigma_t^2(\theta_T))| > \eta) \leq B_9\lfloor T\gamma\rfloor\alpha_T^2$$

as T tends to infinity. □

Lemma 5. *For any $\theta \in \Theta$ it holds that if f is non-constant there exists a constant $c_\theta > 0$ such that*

$$\lim_{T\to\infty} \mathbb{P}(l_T(\theta) - \{-\int_0^1 \log(f(u))du - 1\} < -c_\theta) = 1.$$

Proof of Lemma 5. Assume initially that θ is such that $\alpha \neq 0$ and $\beta \neq 0, 1$ and rewrite the log-likelihood function as follows

$$\begin{aligned}&l_T(\theta) - \{-\int_0^1 \log(f(u))du - 1\} \\ =\ & \int_0^1 \log(f(u))du - \frac{1}{T}\sum_{t=1}^{T}\log(f(t/T)) - \frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta)}(z_t^2-1) \qquad (13)\\ +\ & \frac{1}{T}\sum_{t=1}^{T}\{\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}\}. \qquad (14)\end{aligned}$$

By the LLN for martingale differences (13) tends to zero in probability as T tends to infinity. Formally, since $\mathbb{E}[z_t^2 - 1] = 0$ and $\sigma_t^2(\theta)$ is measurable with respect to

$\mathcal{F}_{t-1} = \mathcal{F}(z_0, ..., z_{t-1})$ we get by applying Chebechev's inequality that

$$
\begin{aligned}
& \mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta)}(z_t^2-1)| > \eta) \\
\leq \quad & \frac{C_1}{T^2}\sum_{i=1}^{T}\sum_{j=i}^{T}\mathbb{E}[\mathbb{E}[\frac{(z_i^2-1)(z_j^2-1)}{\sigma_i^2(\theta)\sigma_j^2(\theta)} \mid \mathcal{F}_{j-1}]] \\
= \quad & \frac{C_2}{T^2}\sum_{t=1}^{T}\mathbb{E}[\frac{1}{\sigma_t^4(\theta)}] \leq \frac{C_3}{T}\mathbb{E}[\frac{1}{(\alpha(z_5^2+\beta z_4^2+\ldots+\beta^4 z_1^2))^2}]
\end{aligned}
$$

and the expectation on the right hand side is finite if $\alpha, \beta > 0$ c.f. Mathai & Provost (1992).

Next turn to the expression in (14) which we decompose into

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}(\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) - \mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)})]) && (15) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}(\mathbb{E}[\frac{f(t/T)}{\sigma_t^2(\theta)}] - \frac{f(t/T)}{\sigma_t^2(\theta)}) && (16) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta)-f(t/T)}{\sigma_t^2(\theta)}] && (17)
\end{aligned}
$$

Initially we will establish that (15) converges in probability to zero. For any $\eta > 0$ direct calculations yield

$$
\begin{aligned}
& \mathbb{P}(|T^{-1}\sum_{t=1}^{T}\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) - \mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)})]| > \eta) \\
= \quad & \mathbb{P}(|T^{-1}\sum_{t=1}^{T}\log(\sigma_t^2(\theta)) - \mathbb{E}[\log(\sigma_t^2(\theta))]| > \eta) \\
\leq \quad & \frac{2}{T^2\eta^2}\sum_{i=1}^{T}\sum_{j=i}^{T}|\mathrm{cov}(\log(\sigma_i^2(\theta)), \log(\sigma_j^2(\theta)))|. && (18)
\end{aligned}
$$

Utilizing the following inequalities

$$-\frac{1}{\sqrt{x}} \leq \log(x) \leq \sqrt{x}, \quad 0 \leq \log(1+x) \leq x,$$

which hold for all strictly positive x, it can be concluded that

$$\begin{aligned}
& |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(\sigma_j^2(\theta)))| \\
= \; & |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(\beta^{j-i}\sigma_i^2(\theta) + \underbrace{\omega\frac{1-\beta^{j-i}}{1-\beta} + \alpha\sum_{k=0}^{j-i-1}\beta^k y_{j-1-k}^2}_{:=Z(i,j)}))| \\
= \; & |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(Z(i,j)(1 + \frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)})))| \\
= \; & |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(1 + \frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)}))| \\
\leq \; & \sqrt{\mathbb{E}[(\log(\sigma_i^2(\theta)))^2]}\sqrt{\mathbb{E}[(\log(1 + \frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)}))^2]} \\
\leq \; & \sqrt{\mathbb{E}[(\frac{1}{\sqrt{\sigma_i^2(\theta)}} + \sqrt{\sigma_i^2(\theta)})^2]}\sqrt{\mathbb{E}[(\frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)})^2]} \\
\leq \; & \beta^{j-i}\sqrt{\mathbb{E}[(\sigma_i^2(\theta) + \frac{1}{\sigma_i^2(\theta)} + 2)]}\sqrt{\mathbb{E}[\sigma_i^4(\theta)]}\sqrt{\mathbb{E}[\frac{1}{Z(i,j)^2}]}.
\end{aligned}$$

For $j > i+1$ the right hand side can be bounded by $\beta^{j-1}C_4$, where the constant C_4 does not depend on either i nor j. In the derivations it is used repeatedly that $\sigma_i^2(\theta)$ is independent of $Z(i,j)$. Since $T^{-2}\sum_{i=1}^{T}\sum_{j=i}^{T}\beta^{j-i}$ tends to zero as T tends to infinity it can be concluded that (18) and hence also (15) tends to zero.

To show that (16) tends to zero in probability note that

$$\begin{aligned}
& |\mathrm{Cov}(\frac{f(i/T)}{\sigma_i^2(\theta)}, \frac{f(j/T)}{\sigma_j^2(\theta)})| \\
= \; & |f(\frac{i}{T})f(\frac{j}{T})(\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}\frac{1}{\beta^{j-i}\sigma_i^2(\theta)+Z(i,j)}] - \mathbb{E}[\frac{1}{\sigma_i^2(\theta)}]\mathbb{E}[\frac{1}{\beta^{j-i}\sigma_i^2(\theta)+Z(i,j)}])| \\
\leq \; & f(\frac{i}{T})f(\frac{j}{T})\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}] \mid \mathbb{E}[\frac{1}{Z(i,j)}] - \mathbb{E}[\frac{1}{\beta^{j-i}\sigma_i^2(\theta)+Z(i,j)}]| \\
\leq \; & f(\frac{i}{T})f(\frac{j}{T})\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}]\mathbb{E}[\frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)(\beta^{j-i}\sigma_i^2(\theta)+Z(i,j))}] \\
\leq \; & \beta^{j-i}f(\frac{i}{T})f(\frac{j}{T})\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}]\mathbb{E}[\sigma_i^2(\theta)]\mathbb{E}[\frac{1}{Z(i,j)^2}].
\end{aligned}$$

As before if $j > i+4$ the expression can be bounded by $\beta^{j-1}C_5$, where the constant C_5 does not depend on either i nor j. Hence it can be concluded that (16) tends to zero. Before turning towards (17) note that for any $\eta > 0$ it holds that

$$\begin{aligned}
\mathbb{P}(\sigma_t^2(\theta) \notin [\underline{f}-\eta, \|f\|_\infty + \eta]) \; & \geq \; \mathbb{P}(\sigma_t^2(\theta) > \|f\|_\infty + \eta) \\
& \geq \; \mathbb{P}(\alpha \underline{f} z_t^2 > \|f\|_\infty + \eta) = C_6 > 0.
\end{aligned}$$

Furthermore since the function $x \mapsto \log(a/x) + (x-a)/x$ has a unique maximum at a with the value 0 and the function f is strictly positive and bounded there exists a constant $C_7 > 0$ such that

$$\sup_{a \in [\underline{f}, \|f\|_\infty]} \; \sup_{x \in [0, a-\eta] \cup [a+\eta, \infty]} \log(a/x) + (x-a)/x < -C_7.$$

Finally it can be concluded that (17) can be bounded by

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}] \\
\leq \; & \frac{1}{T}\sum_{t=1}^{T} -C_7\mathbb{P}(\sigma_t^2(\theta) \notin [\underline{f} - \eta, \|f\|_\infty + \eta]) \\
\leq \; & \frac{1}{T}\sum_{t=1}^{T} -C_7C_6 = -C_7C_6 = c_\theta < 0,
\end{aligned}
$$

which verifies the claim of the lemma. For the special cases $\alpha = 0$ or $\beta = 0$ the lemma is trivially satisfied. If $\beta = 1$ the lemma follows from observing $\sigma_t^2(\theta)$ tends to infinity almost surely as t grows. □

Lemma 6. *For $\theta \in \Theta \backslash (0, 0, 1)$ there exists an open subset of Θ around θ denoted $V(\theta)$ and a constant $\gamma_\theta > 0$ such that*

$$\mathbb{P}(\sup_{\theta^* \in V(\theta)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - \gamma_\theta)$$

tends to one as T tends to infinity.

Proof of Lemma 6. We divide the proof into seven cases mainly because we have to be very careful when θ lies on the boundary of Θ.

1. $\theta = (\omega, \alpha, \beta)' \in (0, \infty) \times [0, 1] \times [0, 1)$
2. $\theta = (\omega, \alpha, \beta)' \in (0, \infty) \times (0, 1] \times \{1\}$
3. $\theta = (\omega, \alpha, \beta)' \in (0, \infty) \times \{0\} \times \{1\}$
4. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times (0, 1] \times \{0\}$
5. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times (0, 1] \times (0, 1)$
6. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times (0, 1] \times \{1\}$
7. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times \{0\} \times [0, 1)$

Case 1. Choose according to Lemma 5 a $c_\theta > 0$ such that

$$\lim_{T\to\infty} \mathbb{P}(l_T(\theta) - \{-\int_0^1 \log(f(u))du - 1\} \geq -c_\theta) = 0.$$

For $\epsilon > 0$ denote by

$$V_\epsilon(\theta) = \{\theta^* \in \Theta \mid ||\theta^* - \theta|| \leq \epsilon\}$$

and note that for T sufficiently large

$$\begin{aligned}
& \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - c_\theta/2) \\
= \; & 1 - \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta) \geq -\int_0^1 \log(f(u))du - 1 - c_\theta/2) \\
\geq \; & 1 - \mathbb{P}(l_T(\theta) \geq -\int_0^1 \log(f(u))du - 1 - c_\theta) - \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)| \geq c_\theta/2).
\end{aligned}$$

To complete the proof we only need to show that for some sufficiently small $\epsilon > 0$ then

$$\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)| \geq c_\theta/2) = 0. \tag{19}$$

Note that this is much weaker than proving that

$$\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)|$$

converges to zero in probability since the probability in (19) should not necessarily converge to zero for this particular ϵ if c_θ is replaced by an arbitrarily small positive number. We proceed by showing that there exists a constant, $D_1 > 0$, such that for any small $\epsilon > 0$ then

$$\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)|$$

can be bounded above by something that converges in probability to $D_1\epsilon$ as T tends to infinity. In particular, the conclusion given by (19) holds for $\epsilon > 0$ such that $D_1\epsilon < c_\theta/2$.

Trivially, for ϵ sufficiently small we get the inequalities

$$\begin{aligned}
\sup_{\theta^* \in V_\epsilon(\theta)} |\beta^t - \beta^{*t}| &\leq \epsilon t(\beta+\epsilon)^{t-1} \\
\sup_{\theta^* \in V_\epsilon(\theta)} |\alpha\beta^t - \alpha^*\beta^{*t}| &\leq \epsilon\alpha t(\beta+\epsilon)^{t-1} + \epsilon(\beta+\epsilon)^t \\
\sup_{\theta^* \in V_\epsilon(\theta)} |\omega\sum_{i=0}^{t-1}\beta^i - \omega^*\sum_{i=0}^{t-1}\beta^{*i}| &\leq \epsilon\frac{1}{1-\beta} + \epsilon(\omega+\epsilon)\sum_{i=0}^{\infty} i(\beta+\epsilon)^{i-1}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\sup_{\theta^* \in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq\ & D_1\epsilon + ||f||_\infty \epsilon \sum_{i=0}^{t-1} z_{t-1-i}^2 \underbrace{[\alpha i(\beta+\epsilon)^{i-1} + (\beta+\epsilon)^i]}_{:=c_i} + \epsilon t(\beta+\epsilon)^{t-1}\sigma_0^2 \quad (20)
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T} \sup_{\theta^* \in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq\ & D_1\epsilon + ||f||_\infty \epsilon \sum_{t=1}^{T}\sum_{i=0}^{t-1} z_{t-1-i}^2 c_i + \frac{1}{T}\sum_{t=1}^{T} t(\beta+\epsilon)^{t-1}\sigma_0^2\epsilon \\
\leq\ & D_2\epsilon + ||f||_\infty \epsilon \{\sum_{i=0}^{\infty} c_i\} \frac{1}{T}\sum_{t=0}^{T-1} z_t^2 \xrightarrow{P} D_3\epsilon
\end{aligned}$$

as T tends to infinity. As $\sigma_t^2(\theta^*)$ is bounded below by $\omega - \epsilon$ on $V_\epsilon(\theta)$ the derivations just above demonstrate that

$$\sup_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} |\log(\sigma_t^2(\theta)) - \log(\sigma_t^2(\theta^*))| \leq \sup_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\omega-\epsilon} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)|$$

is bounded above by something that converges in probability to $D_4\epsilon$ as T tends to infinity. Consider now the decomposition

$$
\begin{aligned}
& \sup_{\theta^* \in V_\epsilon(\theta)} |l_T(\theta) - l_T(\theta^*)| \\
\leq & \sup_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T} \sum_{t=1}^{T} |\log(\sigma_t^2(\theta)) - \log(\sigma_t^2(\theta^*))| \\
+ & ||f||_\infty \frac{1}{T} \sum_{t=1}^{T} z_t^2 \sup_{\theta^* \in V_\epsilon(\theta)} |\frac{1}{\sigma_t^2(\theta)} - \frac{1}{\sigma_t^2(\theta^*)}| \\
\leq & \sup_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\omega - \epsilon} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \qquad (21) \\
+ & \frac{||f||_\infty}{(\omega - \epsilon)^2} \frac{1}{T} \sum_{t=1}^{T} (z_t^2 - 1) \sup_{\theta^* \in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \qquad (22) \\
+ & \frac{||f||_\infty}{(\omega - \epsilon)^2} \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta^* \in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)|. \qquad (23)
\end{aligned}
$$

It follows by previous computations that (21) and (23) can be bounded above by variables converging in probability to constants of the form $D\epsilon$. The remaining term (22) is a martingale difference and by (20) we find that for $\epsilon > 0$ sufficiently small

$$
\begin{aligned}
0 \leq & \sup_{\theta^* \in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq & D_5\epsilon + D_6\epsilon \sum_{i=0}^{t-1} (z_{t-1-i}^2 - 1) c_i.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
& \mathbb{E}[(\frac{1}{T}\sum_{t=1}^{T}(z_t^2-1)\sup_{\theta^*\in V_\epsilon(\theta)}|\sigma_t^2(\theta)-\sigma_t^2(\theta^*)|)^2] \\
\leq\ & \kappa_2^2\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}[(\sup_{\theta^*\in V_\epsilon(\theta)}|\sigma_t^2(\theta)-\sigma_t^2(\theta^*)|)^2] \\
\leq\ & \frac{1}{T}D_5^2\epsilon^2+D_6^2\epsilon^2\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}[(z_1^2-1)^2]\sum_{i=0}^{t-1}c_i^2 \\
\leq\ & \frac{1}{T}D_5^2\epsilon^2+\frac{1}{T}D_6^2\epsilon^2\kappa_2\sum_{i=0}^{\infty}c_i^2
\end{aligned}
$$

verifying that (22) tends to zero in probability which is much stronger that what we need.

Case 2 and 6. Note initially that for ϵ adequately small

$$
\inf_{\theta^*\in V_\epsilon(\theta)}\sigma_t^2(\theta^*)\geq(\alpha-\epsilon)\sum_{i=0}^{t-1}(1-\epsilon)^i\underline{f}z_{t-1-i}^2\equiv\underline{\sigma}_t^2(\epsilon).
$$

Hence

$$
\sup_{\theta^*\in V_\epsilon(\theta)}l_T(\theta^*)=\sup_{\theta^*\in V_\epsilon(\theta)}-\frac{1}{T}\sum_{t=1}^{T}(\log(\sigma_t^2(\theta^*))+\frac{y_t^2}{\sigma_t^2(\theta^*)})\leq\frac{1}{T}\sum_{t=1}^{T}-\log(\underline{\sigma}_t^2(\epsilon)),
$$

which can be bounded by

$$
\begin{aligned}
& -\log(\alpha-\epsilon)-\log(\underline{f})-k\log(1-\epsilon)-\frac{1}{T}\sum_{t=1}^{T}\log(\sum_{i=0}^{t\wedge k-1}z_{t-i-1}^2) \\
\xrightarrow{P}\ & -\log(\alpha-\epsilon)-\log(\underline{f})-k\log(1-\epsilon)-\mathbb{E}[\log(U_k)] \qquad (24)
\end{aligned}
$$

where the convergence is due to the the law of large numbers and $U_k=z_1^2+\cdots+z_k^2$. Now choose $k\in\mathbb{N}$ and ϵ so small that (24) is strictly less then $\int_0^1\log(f(u))-1du$ as desired.

Case 3. Note initially that for ϵ adequately small

$$\inf_{\theta^* \in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \geq (\omega - \epsilon) \sum_{i=0}^{t-1} (1-\epsilon)^i \equiv \underline{\underline{\sigma}}_t^2(\epsilon).$$

Hence for suitably large T

$$\sup_{\theta^* \in V_\epsilon(\theta)} l_T(\theta^*) \leq \frac{1}{T} \sum_{t=1}^{T} - \log(\underline{\underline{\sigma}}_t^2(\epsilon)) \leq -\log(\omega - \epsilon) + \log(2) + \log(\epsilon),$$

and since the right hand side converges to minus infinity as ϵ tends to zero the desired result has been established.

Case 4. Note that for ϵ sufficiently small then $\inf_{\theta^* \in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \geq (\alpha - \epsilon) y_{t-1}^2$. In particular

$$\begin{aligned} l_T(\theta^*) &\leq -\frac{1}{T} \sum_{t=1}^{T} (\log((\alpha - \epsilon) y_{t-1}^2) + \frac{y_t^2}{\sigma_t^2(\theta^*)}) \\ &= -\log(\alpha - \epsilon) - \frac{1}{T} \sum_{t=1}^{T} \log(f(\frac{t-1}{T})) - \frac{1}{T} \sum_{t=1}^{T} \log(z_{t-1}^2) - \frac{1}{T} \sum_{t=1}^{T} \frac{y_t^2}{\sigma_t^2(\theta^*)}. \end{aligned}$$

Now, working on a probability space where we have a doubly infinite sequence, $(z_t)_{t \in \mathbb{Z}}$, of innovations we get that

$$\begin{aligned} \inf_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T} \sum_{t=1}^{T} \frac{y_t^2}{\sigma_t^2(\theta^*)} &\geq \frac{1}{T} \sum_{t=1}^{T} \frac{y_t^2}{\frac{\epsilon}{1-\epsilon} + (\alpha + \epsilon) \sum_{i=0}^{t-1} \epsilon^i y_{t-1-i}^2 + \epsilon^t \sigma_0^2} \\ &\geq D_7 \frac{1}{T} \sum_{t=1}^{T} \frac{z_t^2}{\epsilon + D_8 \sum_{i=0}^{t-1} \epsilon^i z_{t-1-i}^2} \\ &\geq D_7 \frac{1}{T} \sum_{t=1}^{T} \frac{z_t^2}{\epsilon + D_8 \sum_{i=0}^{\infty} \epsilon^i z_{t-1-i}^2}. \end{aligned}$$

By the ergodic theorem the right hand side converges in probability towards its mean, and since by Fatou's lemma

$$
\begin{aligned}
& \liminf_{\epsilon\to 0} \mathbb{E}[\frac{1}{T}\sum_{t=1}^{T}\frac{z_t^2}{\epsilon + D_8\sum_{i=0}^{\infty}\epsilon^i z_{t-1-i}^2}] \\
= & \liminf_{\epsilon\to 0} \mathbb{E}[\frac{z_t^2}{\epsilon + D_8\sum_{i=0}^{\infty}\epsilon^i z_{t-1-i}^2}] \\
\geq & \mathbb{E}[\liminf_{\epsilon\to 0} \frac{z_t^2}{\epsilon + D_8\sum_{i=0}^{\infty}\epsilon^i z_{t-1-i}^2}] = \mathbb{E}[\frac{z_t^2}{Dz_{t-1}^2}] = +\infty
\end{aligned}
$$

we conclude that for $\epsilon > 0$ sufficiently small

$$
\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) - \{-\int_0^1 \log(f(u))du - 1\} < -1) = 1.
$$

Case 5. Since for $\epsilon > 0$ sufficiently small

$$
\begin{aligned}
& \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta^*) - \sigma_t^2(\theta)| \\
\leq & \frac{\epsilon}{1-(\beta+\epsilon)} + \epsilon\sum_{i=0}^{t-1}(\beta+\epsilon)^i y_{t-1-i}^2 + \alpha\sum_{i=1}^{t-1} i\epsilon(\beta+\epsilon)^{i-1}y_{t-1-i}^2 + (\beta+\epsilon)^t\sigma_0^2
\end{aligned}
$$

and for any $k \in \mathbb{N}$

$$
\inf_{\theta^*\in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \geq (\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i y_{t-1-i}^2
$$

we deduce from previous arguments that

$$
\begin{aligned}
& \sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)| \\
\leq & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\inf_{\theta^*\in V_\epsilon(\theta)}\sigma_t^2(\theta^*)}\sup_{\theta^*\in V_\epsilon(\theta)}|\sigma_t^2(\theta^*) - \sigma_t^2(\theta)| \\
+ & \frac{1}{T}\sum_{t=1}^{T}\frac{y_t^2}{(\inf_{\theta^*\in V_\epsilon(\theta)}\sigma_t^2(\theta^*))^2}\sup_{\theta^*\in V_\epsilon(\theta)}|\sigma_t^2(\theta^*) - \sigma_t^2(\theta)|
\end{aligned}
$$

In particular, to demonstrate that $\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)|$ is bounded in probability by ϵD we only need to work with terms of the form

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\epsilon\sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2}{(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2} \tag{25}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\alpha\epsilon\sum_{i=1}^{t-1}i(\beta+\epsilon)^{i-1} z_{t-1-i}^2}{(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2} \tag{26}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\epsilon z_t^2\sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2}{[(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}]^2} \tag{27}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\alpha\epsilon z_t^2\sum_{i=1}^{t-1}i(\beta+\epsilon)^{i-1} z_{t-1-i}^2}{[(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}]^2}. \tag{28}$$

As in the proof of Case 4 introduce a doubly infinite sequence, $(z_t)_{t\in\mathbb{Z}}$, of innovations and note that for $\rho_1,\rho_2\in(0,1)$ then by the ergodic theorem

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=0}^{\infty}i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}\quad\stackrel{P}{\to}\quad\mathbb{E}[\frac{\sum_{i=0}^{\infty}i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]$$

where

$$\begin{aligned}
&\mathbb{E}[\frac{\sum_{i=0}^{\infty}i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]\\
=\ &\sum_{i=0}^{k}\mathbb{E}[\frac{i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]+\mathbb{E}[\frac{1}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]\mathbb{E}[\sum_{i=k+1}^{\infty}i\rho_1^i z_{t-1-i}^2]\\
\leq\ &\sum_{i=1}^{k}i(\rho_1/\rho_2)^i+\mathbb{E}[\frac{1}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]\sum_{i=k+1}^{\infty}i\rho_1^i
\end{aligned}$$

and the right hand side is finite for $k\geq 5$, c.f. Mathai & Provost (1992). This shows that asymptotically for T large then (25) and (26) may be bounded above in probability by ϵD. To show that (27) and (28) may be bounded in probability by ϵD note that

$$\frac{1}{T}\sum_{t=1}^{T}\frac{z_t^2\sum_{i=0}^{\infty} i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2} \xrightarrow{P} \mathbb{E}[\frac{z_t^2\sum_{i=0}^{\infty} i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}]$$

where

$$\begin{aligned}
& \mathbb{E}[\frac{z_t^2\sum_{i=0}^{\infty} i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}] \\
\leq & \sum_{i=1}^{k}\mathbb{E}[\frac{i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}] + \mathbb{E}[\frac{1}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}]\mathbb{E}[\sum_{i=k+1}^{\infty} i\rho_1^i z_{t-1-i}^2] \\
\leq & \sum_{i=1}^{k}\{\frac{1}{2}\mathbb{E}[(i\rho_1^i z_{t-1-i}^2)^2] + \frac{1}{2}\mathbb{E}[\frac{1}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^4}]\} \\
+ & \mathbb{E}[\frac{1}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}]\sum_{i=k+1}^{\infty} i\rho_1^i
\end{aligned}$$

with the right hand side finite for k large enough.

Case 7. For $\theta = (0, 0, \beta)', 0 \leq \beta < 1$ and $\epsilon > 0$ small enough we get that

$$\sup_{\theta^*\in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \leq \frac{1}{1-(\beta+\epsilon)}\epsilon + \epsilon||f||_\infty \sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2 + (\beta+\epsilon)^t\sigma_0^2 := \overline{\sigma}_t^2(\epsilon).$$

Using the inequality $-1/x \leq 2\log(x)$ we get that

$$\begin{aligned}
\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) &= \sup_{\theta^*\in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T}(-\log(\sigma_t^2(\theta^*)) - \frac{y_t^2}{\sigma_t^2(\theta^*)}) \\
&\leq \sup_{\theta\in V_\epsilon(\theta)} \sum_{t=1}^{T}(\log(\sigma_t^2(\theta^*)) - 2\log(y_t^2)) \\
&\leq \frac{1}{T}\sum_{t=1}^{T}(\log(\overline{\sigma}_t^2(\epsilon)) - 2\log(z_t^2) - 2\log(f(t/T))) \\
&\leq \log(\frac{1}{T}\sum_{t=1}^{T}\overline{\sigma}_t^2(\epsilon)) - \frac{2}{T}\sum_{t=1}^{T}\log(z_t^2) - \frac{2}{T}\sum_{t=1}^{T}\log(f(t/T)).
\end{aligned}$$

Clearly, the last two terms tend to a constant and since

$$\frac{1}{T}\sum_{t=1}^{T}\overline{\sigma}_t^2(\epsilon) \;\leq\; \frac{\epsilon}{1-(\beta+\epsilon)}+\frac{\epsilon}{1-(\beta+\epsilon)}\frac{||f||_\infty}{T}\sum_{t=1}^{T}z_t^2+\frac{1}{T}\frac{1}{1-(\beta+\epsilon)}\sigma_0^2$$

we conclude that for $\epsilon > 0$ small and a suitable $\gamma_\theta > 0$ then

$$\lim_{T\to\infty}\mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - \gamma_\theta) = 1.$$

□

References

Ait-Sahalia, Y. (1996), 'Testing continuous-time models of the spot interest rate', *Review of Financial Studies* **2**, 385–426.

Akram, Q.F. & R. Nymoen (2006), 'Econometric modelling of slack and tight labour markets', *Economic Modelling,* **23**, 579–596.

Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell, UK.

Andersen, T.G. & T. Bollerslev (1997), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**, 115–158.

Andersen, T.G., T. Bollerslev, F.X. Diebold & P. Labys (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Anderson, H.M. (1997), 'Transaction costs and nonlinear adjustment towards equilibrium in the US treasury bill market', *Oxford Bulletin of Economics and Statistics* **59**, 465–484.

Baillie, R.T., T. Bollerslev & H.O. Mikkelsen (1996), 'Fractional integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **74**, 3–30.

Balke, N.S. & T.B. Fomby (1997), 'Threshold cointegration', *International Economic Review* **38**, 627–645.

Barndorff-Nielsen, O.E. & N. Shephard (2001), 'Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion)', *Journal of the Royal Statistical Society, Series B* **63**, 167–241.

Basrak, B., R.A. Davis & T. Mikosch (2002*a*), 'A characterization of multivariate variation', *The Annals of Applied Probability* **12**(3), 908 – 920.

Basrak, B., R.A. Davis & T. Mikosch (2002*b*), 'Regular variation of GARCH processes', *Stochastic Processes and their Applications* **99**, 95–115.

Bauwens, L., S. Laurent & J.V.K. Rombouts (2006), 'Multivariate GARCH models: A survey', *Journal of Applied Econometrics* **21**, 79–109.

Bec, F., A. Guayb & E. Guerre (2008), 'Adaptive consistent unit-root tests based on autoregressive threshold model', *Journal of Econometrics* **142**, 94–133.

Bec, F. & A. Rahbek (2004), 'Vector equilibrium correction models with non-linear discontinuous adjustments', *The Econometrics Journal* **7**, 628–651.

Berkes, I., L. Horváth & P. Kokoszka (2003), 'GARCH processes: structure and estimation', *Bernoulli* **9**(2), 201–227.

Bingham, N.H., C.M. Goldie & J.L. Teugels (1987), *Regular Variation*, Vol. 27 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, UK.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model', *The Review of Economics and Statistics* **72**, 498–505.

Bollerslev, T. & R.F. Engle (1993), 'Common persistence in conditional variance', *Econometrica* **61**, 167–186.

Brown, B. M. (1971), 'Martingale central limit theorems', *Annals of Mathematical Statistics* **42**, 59–66.

Campbell, J. & R. Shiller (1987), 'Cointegration and tests of present value models', *Journal of Political Economy* **95**, 1062–1088.

Chan, K.S. & H. Tong (1985), 'On the use of deterministic Lyapunov function for the ergodicity of stochastic difference equations', *Advances in applied probability* **17**, 666–678.

Chan, N.H. & L. Peng (2005), 'Weighted least absolute deviations estimation for an AR(1) process with ARCH(1) errors', *Biometrika* **92**, 477–484.

Chow, Y. (1998), 'Regime switching and cointegration tests of the efficiency of futures markets', *Journal of Futures Markets* **18**, 871–901.

Christoffersen, P., K. Jacobs & K. Mimouni (2007), Models for S&P-500 dynamics: Evidence from realized volatility, daily returns, and option prices. Working paper McGill Univeristy.

Cline, D.B.H. (2007), 'Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models', *Advances in Applied Probability* **39**, 462–491.

Cline, D.B.H. & H. Pu (2004), 'Stability and the Lyapounov exponent of threshold AR-ARCH models', *Annals of Applied Probability* **14**, 1920–1949.

Corradia, V., N. R. Swanson & H. White (2000), 'Testing for stationarity-ergodicity and for comovements between nonlinear discrete time Markov processes', *Journal of Econometrics* **96**, 39–73.

Davies, R.B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **64**, 247–254.

Davies, R.B. (1987), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **74**, 33–43.

Davis, R.A. & T. Hsing (1995), 'Point process and partial sum convergence for weakly dependent random variables with infinite variance', *The Annals of Probability* **23**(2), 879–917.

Davis, R.A. & T. Mikosch (1998), 'The sample autocorrelations of heavy-tailed processes with applications to ARCH', *The Annals of Statistics* **26**, 2049–2080.

Ding, Z. & C.W.J. Granger (1996), 'Modeling volatility persistence of speculative returns: A new approach', *Journal of Econometrics* **73**, 185–215.

Doornik, J.A. (1998), *Object-Oriented Matrix Programming Using Ox 2.0*, Timberlake Consultants Press, London.

Drost, F.C. & B.J.M. Werker (1996), 'Closing the GARCH gap: Continuous time GARCH modeling', *Journal of Econometrics* **74**(1), 31–57.

Drost, F.C. & T.E. Nijman (1993), 'Temporal aggregation of GARCH processes', *Econometrica* **61**(4), 909–927.

Dufrenot, G. & V. Mignon (2002), *Recent Developments in Nonlinear Cointegration with Applications to Macroeconomics and Finance*, Kluwer Academic Publishers.

Dumas, B. (1992), 'Dynamic equilibrium and the real exchange rate in a spatially seperated world', *Review of Financial Studies* **5**, 153–180.

Embrechts, P., C. Klüppelberg & T. Mikosch (1997), *Modelling Extremal Events*, Springer, Berlin.

Enders, W. & P.L. Siklos (2001), 'Cointegration and threshold adjustment', *Journal of Business and Economic Statistics* **19**, 166–176.

Engle, R.F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**, 987–1007.

Engle, R.F. & A.J. Patton (2001), 'What good is a volatility model?', *Quantitative Finance* **1**, 237–245.

Engle, R.F. & T. Bollerslev (1986), 'Modelling the persistence of conditional variance', *Econometric Reviews* **5**, 1–50.

Escribano, A. (2004), 'Nonlinear error correction: The case of money demand in the United Kingdom', *Macroeconomic Dynamics* **8**, 76–116.

Evans, M.D. & K.K. Lewis (1994), 'Do stationary risk premia explain it all? Evidence from the term structure', *Journal of Monetary Economics* **33**, 285–318.

Feigin, P.D. & R.L. Tweedie (1985), 'Random coefficient autoregressive processes. a Markov chain analysis of stationarity and finiteness of moments', *Journal of Time Series Analysis* **6**, 1–14.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed., Wiley, New York.

Forbes, C. S. & P. Kofman (2000), Bayesian soft target zones. Working Paper 4/2000, Monash University, Australia.

Francq, C. & J. Zakoïan (2004), 'Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes', *Bernoulli* **10**(4), 605–637.

Francq, C. & J. Zakoïan (2000), 'Estimating weak GARCH representations', *Econometric Theory* **16**(5), 692–728.

Francq, C. & J. Zakoïan (2006), 'Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process', *Econometric Theory* **22**, 815–834.

Giné, E., F. Götze & D.M. Mason (1997), 'When is the Student t-statistic asymptotically standard normal?', *The Annals of Probability* **25**, 1514–1531.

Gourieroux, C. & C.Y. Robert (2006), 'Stochastic unit root models', *Econometric Theory* **22**, 1052–1090.

Hall, A.D., H.M. Anderson & C.W.J. Granger (1992), 'A cointegration analysis of treasury bill yields', *Review of Economics & Statistics* **74**, 116–126.

Hamilton, J.D. (2008), Macroeconomics and ARCH. Woring paper University of California, San Diego.

Hansen, B.E. (1996), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica* **64**, 413–430.

Hansen, B.E. (1997), 'Inference in TAR models', *Studies in Nonlinear Dynamics and Econometrics* **2**, 1–14.

Hendry, D.F. & N.R. Ericsson (1991), 'An econometric analysis of UK money demand in monetary trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz', *The American Economic Review* **81**, 8–38.

Hillebrand, E. (2005), 'Neglecting parameter changes in GARCH models', *Journal of Econometrics* **129**, 121–138.

Jensen, A.T. & T. Lange (2008), On IGARCH and convergence of the QMLE for misspecified GARCH models. Working paper Department of Mathematical Sciences, University of Copenhagen (4th paper of this thesis).

Jensen, S.T. & A. Rahbek (2004*a*), 'Asymptotic inference for nonstationary GARCH', *Econometric Theory* **20**, 1203–1226.

Jensen, S.T. & A. Rahbek (2004*b*), 'Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case', *Econometrica* **72**(2), 641–646.

Jensen, S.T. & A. Rahbek (2007), 'On the law of large numbers for (geometrically) ergodic Markov chains', *Econometric Theory* **23**, 761–766.

Johansen, Soren (1995), *Likelihood-Based inference in cointegrated vector autoregressive models*, Oxford University Press.

Johansen, Søren (2007), Cointegration. Overview and development, *in* 'Handbook of Financial Time Series', Springer.

Kapetanios, G., Y. Shin & A. Snell (2006), 'Testing for cointegration in nonlinear smooth transition error correction models', *Econometric Theory* **22**, 279–303.

Kristensen, D. (2005), Uniform erglodicity of a class of Markov chains with applications to time series models, Working paper, Columbia University.

Kristensen, D. & A. Rahbek (2005), 'Asymptotics of the QMLE for a class of ARCH(q) models', *Econometric Theory* **21**, 946–961.

Krolzig, H.M.L., M.L. Marcellino & G.E.L. Mizon (2002), 'A Markov-switching vector equilibrium correction model of the UK labour market', *Empirical Economics* **27**, 233–254.

Lange, T. (2006), Stable limits for the OLS estimator in the AR-ARCH model, Working paper, University of Copenhagen, Department of Applied Mathematics and Statistics.

Lange, T. (2008*a*), First and second order non-linear cointegration models. Working paper Department of Mathematical Sciences, University of Copenhagen (2nd paper of this thesis).

Lange, T. (2008*b*), Limiting behavior of the heteroskedastic robust Wald test when the underlying innovations have heavy tails. Working paper Department of Mathematical Sciences, University of Copenhagen (3rd paper of this thesis).

Lange, T. & A. Rahbek (2008), *Handbook of Financial Time Series*, Springer, chapter Regime Switching Time Series Models: A survey.

Lange, T., A. Rahbek & S.T. Jensen (2007), Estimation and asymptotic inference in AR-ARCH model. Working paper Department of Mathematical Sciences, University of Copenhagen (1st paper of this thesis).

Lanne, M. & P. Saikkonen (2003), 'Modeling the U.S. short-term interest rate by mixture autoregressive processes', *Journal of Financial Econometrics* **1**, 96–125.

Lee, S.W. & B. Hansen (1994), 'Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator', *Econometric Theory* **10**, 29–53.

Lehmann, E.L. (1999), *Elements of Large-Sample Theory*, Springer-Verlag.

Liebscher, E. (2005), 'Towards a unified approach for providing geometric ergodicity and mixing properties of nonlinear autoregressive processes', *Journal of Time Series Analysis* **26**, 669–689.

Ling, S. (2004), 'Estimation and testing stationarity for double-autoregressive models', *Journal of the Royal Statistical Society: Series B* **66**, 63–78.

Ling, S. (2007*a*), 'A double AR(p) model: Structure and estimation', *Statistica Sinica* **17**, 161–175.

Ling, S. (2007*b*), 'Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models', *Journal of Econometrics* **140**, 849–873.

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

Ling, S. & W. K. Li (1998), 'Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors', *The Annals of Statistics* **26**, 84–125.

Lu, Z. (1996), 'A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model', *Statistics & Probability Letters* **30**, 305–311.

Lumsdaine, R.L. (1996), 'Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models', *Econometrica* **64**, 575–596.

Mathai, A.M. & S.B. Provost (1992), *Quadratic Forms in Random Variables*, Marcel Dekker, New York.

Medeiros, M.C. & Á. Veiga (2004), Modeling multiple regimes in financial volatility with a flexible coefficient GARCH model, Technical report, Departamento de Economia, Pontifical Catholic University of Rio de Janeiro.

Meitz, M. & P. Saikkonen (2006), Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, Working paper, Stockholm School of Economics & University of Helsinki.

Meitz, M. & P. Saikkonen (2008), 'Stability of nonlinear AR-GARCH models', *Journal of Time Series Analysis* **29**, 453–475.

Meyn, S.P. & R.L. Tweedie (1993), *Markov chains and stochastic stability*, Communications and control engineering series, Springer-Verlag, London ltd., London.

Mikosch, T. & C. Stărică (2004), 'Non-stationarities in financial time series, the long run dependence and the IGARCH effect', *Review of Economics and Statistics* **86**, 378–390.

Nelson, D. B. (1990), 'Stationarity and persistence in the GARCH(1,1) model', *Econometric Theory* **6**, 318–334.

Nelson, D.B. (1992), 'Filtering and forecasting with misspecified ARCH models. I. Getting the right variance with the wrong model', *Journal of Econometrics* **52**, 61–90.

Nelson, D.B. & P. Foster (1994), 'Asymptotic filtering theory for univariate ARCH models', *Econometrica* **62**(1), 1–41.

Nelson, D.B. & P. Foster (1995), 'Filtering and forecasting with misspecified ARCH models. II. Making the right forecast with the wrong model', *Journal of Econometrics* **67**(2), 303–335.

Resnick, S.I. (1987), *Extreme Values, Regular Variation, and Point Processes*, Applied Probability, Springer.

Saikkonen, P. (2005), 'Stability results for nonlinear error correction models', *Journal of Econometrics* **127**, 69–81.

Saikkonen, P. (2008), 'Stability of regime switching error correction models under linear cointegration', *Econometric Theory* **24**, 294–318.

Sercu, P., R. Uppal & C. Van Hulle (1995), 'The exchange rate in the presence of transactions costs: Implications for tests of purchasing power parity', *Journal of Finance* **50**, 1309–1319.

Stock, J. & M. Watson (1988), 'Testing for common trends', *Journal of the American Statistical Association* **83**, 1097–1107.

Strumann, D. & T. Mikosch (2006), 'Quasi–maximum–likelihood estimation in heteroscedastic time series: A stochastic recurrence equations approach', *Annals of Statistics* **34**, 2449–2495.

Stărică, C. (2003), Is GARCH(1,1) as good a model as the accolades of the Nobel prize would imply? Department of Mathematical Statistics, Chalmers University of Technology.

Tjøstheim, D. (1990), 'Non-linear time series and Markov chains', *Advances in Applied Probability* **22**, 587–611.

Tzavalis, E. & M. Wickens (1998), 'A re-examination of the rational expectations hypothesis of the term structure: Reconciling the evidence from long-run and short-run tests', *International Journal of Finance & Economics* **3**, 229–239.

Wang, Y. (2002), 'Asymptotic nonequivalence of GARCH models and diffusions', *Annals of Statistics* **30**(3), 754–783.

Weiss, A. A. (1984), 'ARMA models with ARCH errors', *Journal of Time Series Analysis* **5**(2), 129–143.

Weiss, A. A. (1986), 'Asymptotic theory for ARCH models', *Econometric Theory* **2**, 107–131.

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**, 817–838.

Wong, C.S. & W.K. Li (2001), 'On a mixture autoregressive conditional heteroscedastic model', *Journal of the American Statistical Association* **96**, 982–995.